Characterization of O-Glycosylation and Phosphorylation on Nuclear Protein by Mass Spectrometry

Andrew William Dawdy
Greenville, Illinois

B.A., Chemistry, Taylor University, 2009

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
December, 2013

Thank you to Dr. Donald F. Hunt and Dr. Jeffrey Shabanowitz for allowing me to pursue my PhD in the Hunt Lab. I have learned an incredible amount during my doctoral education and research in this lab. My time here has well-equipped me for beginning the next chapter of my scientific career, and I don't take this opportunity that I have had for granted. I am very grateful, thank you.

Thank you to Drs. David S. Cafiso, Cameron Mura, Ralph O. Allen, and Daniel R. Foltz for serving on my graduate defense committee. I appreciate your time and contributions to my dissertation.

Thank you also to my many collaborators. I have enjoyed working with Drs. Tai-ping Sun, Kathy Wilson, Gerald Hart, Noel Lowndes, James Bieker, Peng Wu, and the outstanding graduate students and post-docs from these fine labs.

Much of my education and joy in the Hunt Lab has come from being surrounded by my intelligent and entertaining lab mates. Thank you to Dr. Dina Bai for enabling all our critical computing tools and for being an ever-positive force in the lab. Thank you to my senior lab mates, Drs. Jeremy Balsbaugh, Sushmit Maitra, Michelle English, Joseph Strukl, Aaron Bailey, Jenny Cottine, Patrick James, Joshua Nicklay, Jessica Chapman, Andrew Norris, Andrea DeSantis, and Erin Jeffrey. Thank you also to my peers and junior labmates, namely Drs. Weihan Wang, Peng Zhao, and Jenn Abelin, and Lissa Anderson, Lichao Zhang, Stacy Malaker, Amanda Wriston, Marshall Chaffee, Ben Barnhill, Scott Ugrin, Paisley Trantham, and Garrett Tanner. I learned so much from you, and I have a collection of fond memories with you all. Thank you especially to the "rowdy" side of the lab for all the laughs. These many memories will always bring a smile to

my face. Thank you also to the "quiet" side for providing a refuge away from the craziness. It has been a pleasure and honor to work alongside all of you. I wish you all well.

Thank you to all my educators in the Bond County Community Unit #2 school district, Taylor University, and the University of Virginia. The road to a PhD starts early, and it was paved in part by many great teachers and professors. I would like to particularly acknowledge and thank my high school chemistry teacher, Mr. Brian Grove, for creating a fun and educational atmosphere that sparked my first serious interest in chemistry.

Above all those previously named, thank you to my wonderful family. Mom and Dad, you are excellent parents, and you always gave me the love and support necessary for achieving my goals. One of those goals has now been realized in my attainment of a PhD, and I thank you for that. Thank you also to my siblings, Becky, Danny, and Sarah, for your support of "The Baby" over the years. I have so many fond memories growing up with you, and I look forward to making more memories with all of our families. Thank you to Pop and Gam Dawdy and Grandpa and Grandma Johnson for your love and prayers as I grew up, and for beginning this amazing family of which I am proud and privileged to be a part. Mom, Dad, Becky, Danny, Sarah, and my entire extended family, I am blessed to have you all as my family, and I love you very much.

During my first summer as a graduate student in the Hunt Lab, I was fortunate to join another wonderful family as well. Mitch, Pam, Garrett, and Ris, thank you for allowing me to marry Natalie and for adopting me into your family. You have been so welcoming and loving, and I look forward to beginning life as a Floridian with you. I love you all.

The person who I thank the most is my incredible wife, Natalie, the most amazing young woman on this earth. I believe our marriage is as close to perfect as humanly possible. I am

truly the most blessed man to have you as my wife. Thank you for being my best friend and companion. I have loved these first few years of marriage with you, and I am so looking forward to beginning this next chapter of our lives together. I love you as much as a man can love his wife. Thank you for all of your love and support.

Ultimately, I thank God for giving me a mind and body capable of studying His awesome world. May the skills I have learned always be used to glorify You, and may my ever-increasing knowledge of Your world bring me ever-increasingly closer to You.

Table of Contents

Table of Contents……iv

Table of Tables……vii

Table of Figures……viii

List of Abbreviations……x

Abstract……xv

Chapter 1: Introduction to the Dissertation

1. Introduction to the Dissertation……1

1.1 The Field of Proteomics……1

1.2 Protein Post-Translational Modification……2

1.3 Characterizing PTMs by Mass Spectrometry……4

1.4 References……14

Chapter 2: Characterization of O-GlcNAcylation, phosphorylation, and a novel O-linked nuclear PTM, O-hexose, on DELLA protein RGA

2.1 Introduction……16

2.1.1 Post-translational modification of proteins is ubiquitous and essential……16

2.1.2 Post-translational modification O-linked β-N-acetylglucosamine……17

O-GlcNAc transferase catalyzes the addition of O-GlcNAc to Ser and Thr……19

O-GlcNAcase catalyzes the removal of O-GlcNAc from Ser and Thr……20

O-GlcNAc donor molecule, UDP-GlcNAc, is produced by the Hexosamine Biosynthetic Pathway……22

O-GlcNAcylation occurs on wide variety of proteins involved in virtually all intracellular functions……23

Cross Talk between O-GlcNAcylation and phosphorylation……24

2.1.3 Evidence for O-GlcNAcylation in Plants……25

Two OGTs are found in plants: SPINDLY and SECRET AGENT……25

Evidence for O-GlcNAcylation of DELLA protein Repressor of ga1-3 (RGA)……26

2.2 Materials and Methods……30

2.2.1 Proteolytic Digestions and Sample Preparation for Mass Spectrometric Analysis……30

2.2.2 Subcleavage by CNBr……31

2.2.3 Derivatization of Carboxylic Acids with Histamine....31
2.2.4 HPLC Capillary Analytical and Pre-column Assembly....32
2.2.5 Sample loading and LC Methodology....32
2.2.6 Phosphorylated Peptide Enrichment by Immobilized Metal Affinity Chromatography 33
2.2.7 Mass Spectrometric Analysis....34

2.3 Results and Discussion....35
2.3.1 6xHis-3xFlag-tagged RGA Purification....35
2.3.2 *In Silico* RGA Digestion....37
LC-MS/MS Methodology....41
Sequencing Peptides using MS/MS analysis....54
MS/MS Sequencing by CAD....55
Site-mapping O-GlcNAcylation by ETD....56
2.3.4 His-Flag-RGA (+SEC) Sequence Coverage....59
Detection of the S/T-rich domain by alternative digests....62
Detection of the S/T-rich domain by alternative stationary phases....65
Detection of the S/T-rich domain by HILIC....65
Detection of the S/T-rich domain by Derivatization of Carboxylic Acids with Histamine....66
Detection of the S/T-rich domain by Lysine Insertion....67
2.3.5 Posttranslational Modification of RGA(+SEC)....69
Detection of RGA O-GlcNAcylation....69
Detection of RGA Phosphorylation....71
Detection of an unexpected and unprecedented glyco-PTM....71
Stoichiometry and Site-locations of RGA O-GlcNAcylation, Phosphorylation, and O-hexosylation....77
Enrichment of Phosphorylated Peptides by IMAC....83
An alternative peptide cleavage technique using CNBr....84
Enhancement of ETD by Derivatization of Carboxylic Acids with Histamine....92
Analysis of RGA(+SEC) PTMs....98
Site Locations of RGA PTMs....102
2.3.6 The effect of SEC on O-hexosylation....108
Summary and Future Analysis related to O-hexose....113
2.3.7 Phosphorylation of RGA by CK1....115

2.4 Conclusions....119

2.5 References....125

Chapter 3: Site-specific phosphorylation of the DNA damage response mediator rad9 by cyclin-dependent kinases regulates activation of checkpoint kinase 1

3.1 Introduction....131
3.1.1 DNA Damage Response....131

3.1.2 Rad9 ..... **132**
3.1.3 Checkpoint Kinase 1 (Chk1) Activation by Rad9 ..... **133**

3.2 Materials and Methods **134**
3.2.1 Molecular biology methods **134**
3.2.2 Rad9 purification **134**
3.2.3 Mass Spectrometric Analysis of Rad9-GFP-FLAG **135**

3.3 Results and Discussion **136**
3.3.1 Rad9-GFP-FLAG Purification **136**
***3.3.2 In silico* Rad9 Digestion** **137**
3.3.3 Analysis of Rad9 by LCMS **145**
3.3.4 Biological role of CAD phosphorylation in Chk1 activation **154**

3.4 Conclusion **160**

3.5 References **162**

Table of Tables

TABLE 2.1 – IN SILICO ASPN DIGESTION OF RGA 38
TABLE 2.2 – IN SILICO TRYPSIN DIGESTION OF RGA. 39
TABLE 2.3 – SITE-LOCALIZED PTMS ON RGA(+SEC) 78
TABLE 2.4 – COMPRENSIVE TABLE OF RGA(+SEC) PTM PEPTIDES 79
TABLE 2.5 – IN SILICO CNBR CLEAVAGE RGA BY CNBR 88
TABLE 2.6 - IN SILICO SUBCLEAVAGE OF APN-N DIGESTED RGA BY CNBR. 90
TABLE 2.7 – CONDENSED SUMMARY OF THE 13 MODIFIED RGA(+SEC) PEPTIDES 99
TABLE 2.8 – EXPANDED SUMMARY OF THE 13 MODIFIED RGA(+SEC) PEPTIDES 100
TABLE 2.9 – COMPREHENSIVE TABLE OF RGA(-SEC) PTM PEPTIDES 109
TABLE 2.10 – SUMMARY OF THE 4 MODIFIED RGA(-SEC) PEPTIDES 110
TABLE 2.11 – SITE-LOCALIZED PTM SITES ON RGA(-SEC). 111
TABLE 2.12 – COMPREHENSIVE TABLE OF RGA(+CK1) PTM PEPTIDES 116
TABLE 2.13 - SUMMARY OF THE 6 MODIFIED RGA(+CK1) PEPTIDES 117
TABLE 2.14 – SITE-LOCALIZED PTM SITES ON RGA(+CK1). 118
TABLE 2.15 - COMPREHENSIVE TABLE OF RGA(+CK1) PTM PEPTIDES 120
TABLE 3.1 – IN SILICO LYS-C DIGESTION OF RAD9 137
TABLE 3.2 – IN SILICO TRYPSIN DIGESTION OF RAD9. 141
TABLE 3.3 – COMPREHENSIVE TABLE OF RAD9 PTM PEPTIDES. 147
TABLE 3.4 – LITERATURE VS NOVEL PHOSPHORYLATION SITES. 150
TABLE 3.5 – SUMMARY OF RAD9 PTM SITES. 160

Table of Figures

FIGURE 1.1 – FIRST STABILITY REGION DEFINED BY MATHIEU EQUATION....7
FIGURE 1.2 – CAD FRAGMENTATION MECHANISM AND SPECTRUM. 10
FIGURE 1.3 – ETD FRAGMENTATION.MECHANISM AND SPECTRUM 11
FIGURE 2.1 – O-GLCNAC TRANSFERASE AND O-GLCNACASE RESPECTIVELY INSTALL AND REMOVE O-GLCNAC FROM SER AND THR RESIDUES....18
FIGURE 2.2 – OGT AND OGA ISOFORMS. 19
FIGURE 2.3 – OGA INHIBITORS....21
FIGURE 2.4 - THE HEXOSAMINE BIOSYNTHETIC PATHWAY PRODUCES O-GLCNAC DONOR, UDP-GLCNAC. 22
FIGURE 2.5 – O-GLCNACYLATION MODIFIEDS PROTEIN FROM ALL FUNCTIONAL CLASSES 23
FIGURE 2.6 - O-GLCNACYLATED PROTEINS ARE INVOLVED IN A WIDE VARIETY OF CELLULAR PROCESSES. 24
FIGURE 2.7 - ARABIDOPSIS SPY (TOP) AND HUMAN OGT (BOTTOM) ARE HIGHLY HOMOLOGOUS. 26
FIGURE 2.8 – A SCHEMATIC OF RGA 27
FIGURE 2.9 – DELLA PROTEIN RGA AND OGT PROTEIN SPY EACH BOTH HAVE ROLES IN NEGATIVELY REGULATING PLANT GROWTH. 28
FIGURE 2.10 - FIVE DELLA PROTEINS HAVE BEEN DISCOVERED IN ARABIDOPSIS. RGA IS A CENTRAL REGULATOR. 29
FIGURE 2.11 – GAS STIMULATE PLANT DEVELOPMENT BY INDUCING DEGRADATION OF THE REPRESSOR DELLA PROTEINS. 29
FIGURE 2.12 – RGA VISUALIZED BY COOMASSIE STAIN ON SDS-PAGE GEL. 36
FIGURE 2.13 – WESTERN BLOT (ANTI-FLAG-HRP) SHOWS CHANGES IN THE SDS-PAGE GEL SHIFT WHEN POTENTIAL SER/THR-RICH REGIONS OF RGA ARE MUTATED TO ALA, ELIMINATING THEIR ABILITY TO BE O-GLCNACYLATED. 37
FIGURE 2.14 – THEORETICAL ISOTOPIC DISTRIBUTION OF VASOACTIVE INTESTINAL PEPTIDE FRAGMENT, HSDAVFTDNYTR. 47
FIGURE 2.15 – EXPERIMENTAL ISOTOPIC DISTRIBUTION OF VASOACTIVE INTESTINAL PEPTIDE FRAGMENT, HSDAVFTDNYTR, USING LTQ-FT AT 50,000 FWHM RESOLUTION. 47
FIGURE 2.16 – XICS OF INTERNAL PEPTIDE STANDARDS VASOACTIVE INTESTINAL PEPTIDE (VASO) AND ANGIOTENSIN (ANGIO) USED TO CALCULATE RELATIVE QUANTITATION OF RGA PEPTIDES. 50
FIGURE 2.17 – ION SUPPRESSION OF O-GLCNACYLATED PEPTIDE. 53
FIGURE 2.18 – SCHEMATIC OF 4 SER/THR-RICH DOMAINS IN RGA 56
FIGURE 2.19 – CAD AND ETD SPECTRA OF MONO-O-GLCNACYLATED PEPTIDE DHHQFQGRLSNHGTSSSSSSIS. 57
FIGURE 2.20 –BASE PEAK CHROMATOGRAMS OF REPRESENTATIVE TRYPSIN AND ASPN RGA DIGESTS. 60
FIGURE 2.21 – RGA SEQUENCE COVERAGE FROM TRYPSIN AND ASPN....61
FIGURE 2.22– SEPARATION OF PEPTIDE SCSSPDSMVTSTSTGTQIGGVIGTTVTTTTTTTTAAGESTR BY FOUR DIFFERENT CHROMATOGRAPHIC STATIONARY PHASES, AND C18 AFTER HISTAMINE DERIVATIZATION OR TRYPSIN+ASPN DIGESTION. 64
FIGURE 2.23 – REACTION MECHANISM FOR DERIVATIZATION OF CARBOXYLIC ACIDS WITH HISTAMINE. .67
FIGURE 2.24 – WESTERN BLOT DEMONSTRATING CHARACTERISTIC RGA(+SEC) SHIFT WAS PRESERVED FOR RGA-G185_G186INSK (+SEC) SAMPLES. 68
FIGURE 2.25 – XICS OF THE PREVIOUSLY UNDETECTABLE RGA(+SEC) POLY-SER/THR REGION. 69
FIGURE 2.26 - XICS OF 1-6 O-GLCNACS DETECTED ON GVIGTTVTTTTTTTTAAGESTR. 70
FIGURE 2.27 – DETECTION OF A PREVIOUSLY UNDESCRIBED NUCLEAR PROTEIN PTM, O-HEXOSYLATION. 72

FIGURE 2.28 – XICS OF 1-5 UNEXPECTED Δ162 MASSES ON GVIGTTVTTTTTTTTAAGESTR DETERMINED TO BE O-HEXOSE PTMS. 73
FIGURE 2.29 – HEXOSE PTM SIGNATURE FRAGMENT IONS PRODUCED BY CAD, CHARACTERISTIC OF O-LINKAGE. 74
FIGURE 2.30 – IMAC STATIONARY PHASE 84
FIGURE 2.31 – ETD % FRAGMENTATION IS RELATIVELY INDEPENDENT OF CHARGE STATE AND PEPTIDE MASS, BUT IS HIGHLY DEPENDENT ON CHARGE DENSITY 86
FIGURE 2.32 – REACTION SCHEME FOR CLEAVAGE OF MET-X BONDS BY CNBR 87
FIGURE 2.33 – ETD CONSUMES ONE CHARGE, REQUIRING A MINIMUM OF 3 CHARGES FOR C AND Z· IONS TO BOTH BE DETECTED. 91
FIGURE 2.34 – DERIVATIZATION OF ASP AND THE C-TERMINUS OF UNMODIFIED VIPGNAIYQFPAIDSSSSSNNQNK WITH HISTAMINE 94
FIGURE 2.35 – REACTION SCHEME OF ASP RESIDUES FORMING STABLE SUCCINIMIDE SIDE-PRODUCTS....95
FIGURE 2.36 –LOCALIZATION OF O-GLNAC AND O-HEXOSE ON ASN-P+CNBR PEPTIDE LSELNPPPLPASSNGL DERIVATIZED WITH HISTAMINE 97
FIGURE 2.37 – HEAT MAP REPORTING ABUNDANCES OF THE 26 POST-TRANSLATIONALLY MODIFIED FORMS OF GVIGTTVTTTTTTTTAAGESTR. 107
FIGURE 2.38 – MIXED ETD SPECTRUM OF MONO-O-GLCNACYLATED GVIGTTVTTTTTTTTAAGESTR IN RGA(-SEC) 113
FIGURE 2.39 – SUM TOTAL OF O-GLCNAC, PHOSPHATE, AND O-HEXOSE IN RGA(+SEC,-SEC,&+CK1) FOR THE SIX MOST ABUNDANTLY MODIFIED RGA PEPTIDES. 121
FIGURE 2.40 - RGA SEQUENCES SUMMARIZING RESIDUES MODIFIED WITH COMBINATORIAL PTMS 122
FIGURE 3.1 – TANDEM PURIFIED RAD9-GFP-FLAG VISUALIZED BY SILVER STAINED SDS-PAGE GEL 137
FIGURE 3.2 – SEQUENCE COVERAGE OF TRYPSIN AND LYS-C-DIGESTED RAD9 FROM ASYNCHRONOUS, G1-, AND G2-ARRESTED CELLS. 145
FIGURE 3.3 – SCHEMATIC OF RAD9 PHOSPHORYLATION SITES DETECTED 154
FIGURE 3.4 – (A) SCHEMATIC OF RAD9 CAD HIGHLIGHTING CDC28 CONSENSUS SITES 1-9 MUTATED TO ALA IN $RAD9^{CDK1-9A}$ PROTEIN. (B) CDC28/CLB2 PRODUCES THE HIGHEST LEVEL OF IN VITRO PHOSPHORYLATION OF RAD9 CAD^{WT}. (C) DNA-DAMAGE-DEPENDENT PHOSPHORYLATION OF CHK1 IS REDUCED OR ELIMINATED IN $RAD9^{CDK1-9A}$, $RAD9^{CAD\Delta}$, RAD9Δ CELLS. (D&E) – CHK1 ACTIVATION IS PRINCIPALLY DEPENDENT ON CDK CONSENSUS SITES 1-9 IN G2/M CELLS. 155
FIGURE 3.5 – CHK1 ACTIVITY WAS DEPENDENT ON CDC28 UPON DNA DAMAGE SEPARATELY INDUCED BY BLEOCIN OR 4-NQO. 157
FIGURE 3.6 – (A) RAD9/CHK1 INTERACTION WAS DEPENDENT ON CDK1-9 BASED ON YEAST TWO-HYBRID ASSAY IN VIVO. (B) RAD9/CHK1 INTERACTION WAS DEPENDENT ON CDC28 IN G2/M AND WAS CDC28-INDEPENDENT IN G1. 158
FIGURE 3.7 – (A) YEAST-TWO HYBRID ASSAY DEMONSTRATED THAT T125 AND T143 MOST CONTRIBUTED TO CAD/CHK1 INTERACTION OF CDK1-9 SITES. (B) – PEPTIDES CONTAINING PT143 MOST EFFECTIVELY PULLED DOWN CHK1 WHILE PT125 SHOWED AN INHIBITORY EFFECT. 159
FIGURE 3.8 – MODEL OF CHK1 ACTIVATION IN RESPONSE TO DNA DAMAGE. 161

Abbreviations

°C	Degrees Celsius
·	Radical species
A	Solvent A, 100 mM acetic acid in water
AcOH	Acetic Acid
AGC	Automated gain control
Ala, A	Alanine
Angio	Angiotensin I
Ambic	Ammonium bicarbonate
Arg, R	Arginine
Asn, N	Asparagine
Asp, D	Aspartic acid
ATM	Ataxia telangiectasia, mutated
ATP	adenosine triphosphate
ATR	ATM and Rad3-related
B	Solvent B, 100 mM acetic acid in 70% acetonitrile
c	Centi (10^{-2})
C	Carboxyl terminus of a polypeptide
C18	Octadecylsilane
CAD	Collison-activated dissociation or Chk1 activating domain
CDK	Cyclin-dependent kinase
Chk1	Checkpoint kinase 1
Chk2	Checkpoint kinase 2
CI	Chemical ionization
CNBr	Cyanogen bromide
Cys, C	Cysteine
Da	Dalton

DC	Direct current
DDR	DNA Damage Response
DNA	deoxyribonucleic acid
DTT	Dithiothreitol
ECD	Electron-capture dissociation
EDC	1-ethyl-3-(3-dimethylaminopropyl)carbodiimide
EDTA	Ethylenediaminetetraacetic acid
EI	Electron ionization
EM	Electron multiplier
ER	endoplasmic reticulum
ESI	Electrospray ionization
ETD	Electron-transfer dissociation
f	Femto (10^{-15})
FAB	Fast atom bombardment
FeCl3	Iron (III) chloride
FT-ICR	Fourier transform ion cyclotron resonance
FWHM	Full width half max
g	Gram
G1	Gap 1 phase
G2	Gap 2 phase
Gal	Galactose
Gln, Q	Glutamine
Glu, E	Glutamic acid
Gly, G	Glycine
gS, gT	O-GlcNAcylated residue
hS, hT	O-hexosylated residue
h, hr	Hour
HBP	Hexosamine biosynthetic pathway

HCl	Hydrochloric acid
HILIC	Hydrophilic interaction chromatography
His, H	Histidine
HPLC	High-performance liquid chromatography
ICR	Ion cyclotron resonance
i.d.	Inner diameter
IAA	Iodoacetamide
Ile, I	Isoleucine
IMAC	Immobilized metal affinity chromatography
IP	Immunoprecipitation
IT	Ion trap
k	Kilo (10^3)
Kasil	Potassium silicate solution
L	Liter
LC	Liquid chromatography
Leu, L	Leucine
LTQ	Linear quadrupole ion trap
Lys, K	Lysine
µ	Micro (10^{-6})
m	Milli (10^{-3}) or meter
M	Molar (mol/L) or Mega (10^6) or Mitosis
MALDI	Matrix-assisted laser desorption ionization
m/z	Mass-to-charge ratio
M+nH	Molecular ion with n charge(s)
MALDI	Matrix-assisted laser desorption ionization
MeCN	Acetonitrile
MeOH	Methanol
Met, M	Methionine

min	Minute
mol	Mole (6.022×10^{23})
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
MS^1	Full mass spectrum
MS^2	Tandem mass spectrum
MW	Molecular weight
n	Nano (10^{-9})
N	Amino terminus of a polypeptide
NCBI	National Center for Biotechnology Information
NHS	N-hydroxysuccinimide
Ni^{2+}- NTA	Nickel (II) - nitriloacetic acid
NL	Normalized level ion counts
o.d.	Outer diameter
OGA	O-GlcNAcase
O-GlcNAc	O-linked β-N-acetylglucosamine
OGT	O-GlcNAc transferase
O-hexose	O-linked hexose
OMSSA	Open Mass Spectrometry Search Algorithm
p	Pico (10^{-12})
PAGE	Polyacrylamide gel electrophoresis
Phe, F	Phenylalanine
ppm	Part-per-million
Pro, P	Proline
pS, pT, pY	Phosphorylated residue
PTFE	Polytetrafluoroethylene
PTM	Post-translational modification
QIT	Quadrupole ion trap

Rad9	Prototypical DNA Damage Response Mediator protein
RGA	‘Repressor of ga1-3’ protein
RF	Radio frequency
RNA	Ribonucleic acid
RP	Reverse phase
s	Second(s)
SDS	Sodium dodecyl sulfate
SEC	SECRET AGENT
Ser, S	Serine or Synthesis phase
SPY	SPINDLY
SUMO	Small ubiquitin-like modifier
TFA	Triflouroacetic acid
Thr, T	Threonine
Trp, W	Tryptophan
Tyr, Y	Tyrosine
UDP	Uridine diphosphate
Val, V	Valine
Vaso	Vasoactive intestinal peptide
X	Any amino acid
XIC	Extracted ion chromatogram
WB	Western blot
WT	Wild-type

Abstract

Proteins are biomolecules that represent the ultimate expression of DNA, as they perform the majority of structural and functional roles in organisms. Proteins accomplish this through their interactions with other proteins and biomolecules, dictated by their molecular confirmation and location. A major regulator of protein activity is covalent post-translational modification (PTM). Characterizing protein PTMs is critical for developing a complete understanding of the proteome, the entire expressed protein complement of DNA. Over 400 PTMs have been identified, and liquid chromatography mass spectrometry (LCMS) has emerged as the most powerful technique to directly characterize protein PTMs. This dissertation reports the implementation of electron transfer dissociation (ETD) and collision-activated dissociation (CAD) in tandem with high-resolution MS to characterize O-linked glycosylation and phosphorylation PTMs on two nuclear proteins. Phosphorylation is a ubiquitous nucleocytoplasmic PTM that is found on Ser, Thr, and Tyr residues of thousands of metazoan proteins. It is involved in the regulation of a wide variety of cellular processes and is perhaps the mostly highly-studied PTM to date. Just 30 years ago, it was largely believed that nucleocytoplasmic protein glycosylation was non-existent. We now know that the nucleocytoplasmic monosaccharide, O-linked β–N-acetylglucosamine, or O-GlcNAc, modifies Ser and Thr residues of metazoan proteins from nearly every functional class, and it has a complex interplay with phosphorylation. O-GlcNAcylation has historically been challenging to study, but improvements in mass spectrometric techniques, such as the development of ETD that preserves labile O-linked PTMs like O-GlcNAc upon fragmentation, have significantly increased the ability to characterize this PTM. It is becoming increasingly evident that O-GlcNAcylation

also regulates numerous cellular processes including transcription, translation, protein trafficking, and protein degradation.

Firstly, we report the comprehensive characterization of O-GlcNAcylation on *Arabidopsis* protein 'repressor of ga1-3', or RGA, a negative regulator of plant growth and development. This marks the first direct and comprehensive characterization of O-GlcNAcylation on a plant protein. In addition, numerous phosphorylation sites were discovered. Most notably and surprisingly, a previously unreported O-linked hexose monosaccharide PTM was discovered on RGA, representing only the second nucleocytoplasmic glyco-PTM reported to date.

Secondly, we report novel phosphorylation of *S. cerevisiae* DNA damage response mediator protein Rad9. Phosphorylation was detected on all 20 putative Cyclin-dependent kinase (CDK) motifs, including 7 previously unreported. All 9 CDK sites located in the Checkpoint kinase 1 (Chk1) activating domain (CAD) were phosphorylated and sites T125 and T143 were found to be especially important in facilitating Rad9-Chk1 interaction and Chk1 activation.

Chapter 1: Introduction to the Dissertation

1.1 The Field of Proteomics

Proteomics is an analogous term to genomics, and was created to describe the study of the third and final core component in the central dogma of molecular biology, protein [1, 2]. The first two molecular components, DNA and RNA, are comprised of nucleic acids. An organism's DNA contains all its hereditary information and collectively forms its genome. This genetic map may be transiently expressed by transcription of a DNA sequence to form the second molecular component of the central dogma, RNA. RNA, comprised of types ribosomal (rRNA), transfer (tRNA), messenger (mRNA), and regulatory non-coding (ncRNA), functions as the expressed form of genetic information by providing molecular machinery and the DNA-derived template necessary for protein synthesis.

The primary structure of a protein consists of a linear molecular chain, or polypeptide, built from combinations of 20 fundamental amino acids. Polypeptides range in size from <10 kDa to >1 MDa and typically fold into tertiary structures dictated directly by their primary sequence of amino acids. Two of the most common protein secondary structural motifs are α-helices and β-sheets. The tertiary structure of the protein is formed by the collective interactions that facilitate folding into its final three-dimensional form. These interactions include hydrogen and disulfide bonding, and ionic and van der Waals interactions amongst different regions of the protein. A protein generally adopts a tertiary conformation that minimizes free energy and may be comprised of independently folding secondary structures, folding between secondary structures and/or undefined domains, and unfolded regions. Lastly,

quaternary protein complexes may be created by non-covalent bonding of multiple polypeptide chains together that function as a single unit (e.g. hemoglobin) [3].

In this way, amino acid sequences dictate higher order structure, and this ultimately dictates the interactions a protein will have with other molecules. Proteins perform most of the structural and functional roles in the cell and are the ultimate molecular product of expressed genes. The proteome may be defined as "the entire complement of proteins expressed in a specific state of an organism or a cell population" [4, 5]. The ultimate goal of proteomics is the detailed characterization of the proteome.

There are 20,000-25,000 protein-coding genes in the human genome [6], yet the proteome is estimated to be 2-3 orders of magnitude larger than this due to the numerous modifications that proteins may undergo. The two main routes to accomplish this are alternative gene splicing at the mRNA level, and enzymatic covalent post-translational modification (PTM) of proteins. The latter of these is the subject of this dissertation [7].

1.2 Protein Post-Translational Modification

Post-translational modifications add to the complexity of the proteome and are often difficult to predict and detect, because they directly change the primary chemical structures of proteins beyond what is directly dictated by the DNA sequence. This influences protein structure, thereby influencing molecular interactions and protein function. PTMs occur in prokaryotes, but are more diverse in type and higher in occurrence in eukaryotes. Approximately 5% of higher eukaryotes genomes can be dedicated to enzymes that perform protein post-translational modification [7]. Over 400 types of PTMs have been discovered [8] with five most common types being phosphorylation, glycosylation, acylation, alkylation, and

oxidation [7]. A complete understanding of the proteome therefore requires the characterization of all the PTMs that alter it. The combinatorial power of multiple PTMs quickly and vastly increases the diversity of proteins' molecular states in a manner that allows intricate biological tasks to be performed on a developmental and physiological timescale [3]. This dissertation focuses on the study of two PTMs critical for the regulation of nuclear and cytoplasmic proteins: phosphorylation and O-linked β-N-acetylglucosamine (O-GlcNAc), a type of O-linked glycosylation [9, 10].

PTMs have historically been studied using a variety of biochemical techniques. Radioactive substrates have been used to perform *in vitro* PTM reactions assays, although these can be hazardous and relatively insensitive. Western blot analysis is also used to detect PTMs by raising antibodies against them. This can be done with monoclonal antibodies raised against a specific PTM site, or by generating pan-specific antibodies to recognize a type of PTM without regard to sequence. The first allows one to probe a specific site, but prior knowledge or an indication of the site must exist and generating specific antibodies is time intensive and expensive. The second approach allows one to assay a larger population of proteins for a particular PTM, but generating pan-specific antibodies for small molecule modifications, like many of the most common PTMs, can be difficult to do with high affinity and independently of the adjacent sequences. Both antibody methods suffer from the potential for ambiguity in results due to the inevitable lack of specificity for 100% of modified sites and potential infidelity of the antibody. They are also not reliable when attempting to determine a library of combinatorial PTMs on one or more proteins. In another more recent approach, peptide and protein arrays have been used to identify protein substrates *in vitro* that are targeted by enzymes specific for a particular PTM. This approach still requires *in vivo* confirmation and is

also not useful for the determination of multiple PTM sites [11]. Mass spectrometry has emerged as the premier method for primary protein sequence and PTM identification, because it offers the ability to directly detect multiple PTMs in complex samples on the femtomole – attomole level. Its continuing advancements are enabling researchers to study PTMs more effectively.

1.3 Characterizing PTMs by Mass Spectrometry

Major breakthroughs in the last 30 years have enabled mass spectrometry to be applied to the study of large biomolecules like peptides and proteins. Reducing it to its most basic components, a mass spectrometer consists of an ionization source, a mass analyzer, and a detector.

In the past, the most commonly used ionization methods were 'hard ionization' techniques like electron ionization (EI) and chemical ionization (CI), whereby samples may be ionized by an electron beam or an ionized gas, respectively [12]. EI is particularly energetic and can generally not be applied to thermally labile compounds or else they will completely decompose. CI is gentler and more controllable but still generates ion fragments upon ionization. Both techniques are limited to the analysis of volatile, low-molecular weight compounds.

A major breakthrough for the mass spectrometric analysis of biomolecules came with the development of the 'soft ionization' technique, fast-atom bombardment (FAB) in 1981 [13]. For the first time, peptides and other biomolecules up to 4000 Da could be routinely analyzed. FAB utilizes a beam of high-energy inert gas atoms to strike analyte dissolved in a liquid matrix to produce a secondary beam consisting of ionized analyte molecules. Two other ionization techniques were developed a few years later that further enhanced mass spectrometry of

biomolecules. Matrix-assisted laser desorption/ionization (MALDI), developed simultaneously by Karas and Hillenkamp [14], and Tanaka and co-workers [15], is similar to FAB but utilizes a laser pulsed onto an improved analyte-matrix. MALDI improved upon the sensitivity and upper mass range of FAB as <100 kDa biomolecules could now be ionized [16].

Electro-spray ionization (ESI), first introduced by Dole in 1968 [17] and successfully coupled to mass spectrometry by Fenn and coworkers in 1984 [18], arguably signifies the most major breakthrough in the analysis of biomolecules by mass spectrometry to date. This technique produces gas-phase ions from electrically charged liquid droplets by 'electrospraying' an analyte-containing solution at high voltage (~2 kV) and atmospheric pressure. ESI made two major advancements in bioanalytical mass spectrometry: (1) it offered compatibility with large, nonvolatile, thermally-labile biomolecules >1 MDa by imparting multiple charges to them while being simultaneously converted to the gas phase, and (2) it allowed continuous flow that has enabled researchers to couple ESI with separation techniques like liquid chromatography, enhancing the ability to analyze complex mixtures. Coupling high performance liquid chromatography (HPLC) with mass spectrometry is commonly referred to as LCMS, and represents the most popular mass spectrometric method for analyzing peptides and proteins today. The two major ionization models of ESI are the charged residue model (CRM) proposed by Dole [17, 19], and the ion evaporation model (IEM) proposed by Iribarne and Thomson [20]. The CRM model reasons that as charged liquid droplets become smaller by solvent evaporation, the charged particles in the droplet become closer and closer until the droplet reached its 'Rayleigh limit.' At this point, the coulombic repulsion forces exceed the surface tension of the droplet and it breaks apart into smaller droplets. This process continues until the eventual formation of 'ultimate droplets' that contain only one solute molecules and excess charge. As the remaining

solvent evaporates, the solute molecule retains some of all of the charge to become a gas-phase analyte ion. According to IEM, electrospray droplets do not undergo fission at the 'Rayleigh limit' but rather the charge density at the surface becomes sufficient to force one or more charged solute ions into the ambient gas. This process repeats to continually 'bud' gas-phase analyte ions from the charged droplets. Numerous investigations of the ESI mechanism have led to conclusions that support both models, so it is plausible that a combination of both mechanisms occur in practice [21].

Ion trap mass analyzers like the quadrupole linear ion trap (LTQ) work well in conjunction with LCMS. The multiply-charged biomolecules produced by ESI generally fall well within the mass-to-charge (m/*z*) range (0-4000) of the LTQ. For example, a typical 40kDa protein containing ~350 residues may be expected to contain an average of 1 charge per 10 residues. The monoisotopic m/*z* of this protein would be <1200. The LTQ is an improvement upon the quadrupole ion trap (QIT), which like the LTQ, operated on the principle of trapping a population of ions and subsequently scanning them out according to their m/z to acquire a mass spectrum. The LTQ effectively stretched the trapping area in the z plane which increased ion space-charging limits before reductions in resolution and mass accuracy, isolation efficiency, activation efficiency, and storage capacity would occur. The LTQ is a sensitive mass analyzer that can acquire a mass spectrum of ions at femtomole-attomole levels with 1 unit m/*z* resolution on a chromatographic timescale (<100 ms) [22]. A typical quadrupole consists of two pairs of opposing metal rods. In the LTQ, 12 hyperbolic poles form three distinct sections of the trap. Ions are trapped and analyzed according to quadrupolar theory using a combination of DC and RF potentials. The motion of ions through the LTQ is described by the Mathieu equation [23],

the solution of which contains the dimensionless parameters *a* and *q* that together describe the stability of an ion in the LTQ (**Equation 1.1 & 1.2**).

(**Equation 1.1**) $a = (\frac{8zU}{m\omega^2 r_0^2})$

(**Equation 1.2**) $q = (\frac{4zV}{m\omega^2 r_0^2})$

Parameters *a* and *q* are directly proportional to the applied DC (*U*) and RF (*V*) potentials, respectively, where *z* is the particular charge, *m* is the particular mass, ω is the angular

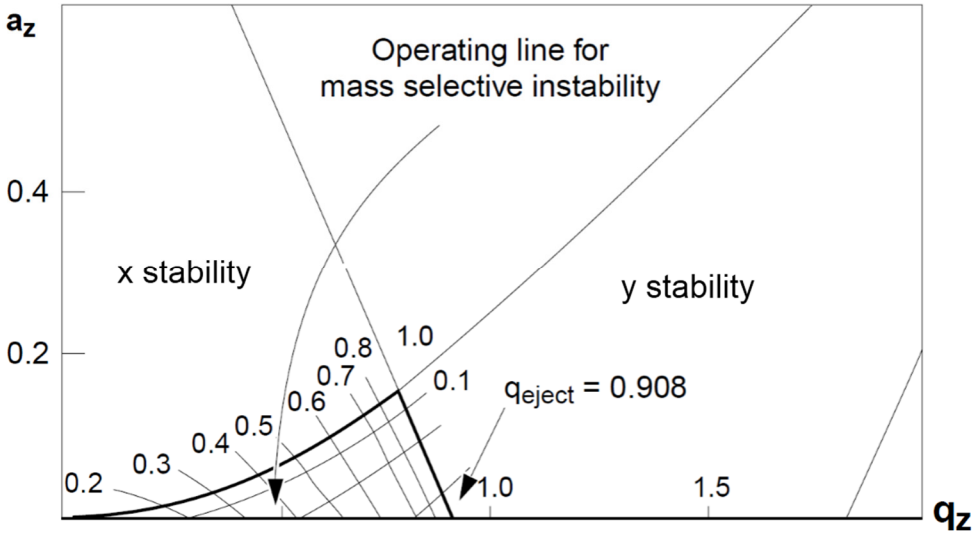


Figure 1.1 – The first stability region closest to the origin in which parameters a and q result in x and y stability in the LTQ.

frequency, and r_0 is the radius between rods. The first stability region described by these parameters (**Figure 1.1**) is the most operational useful [23]. RF potentials applied to the rods keep the ions in stable trajectories in the x and y planes. The ions are axially (z plane) trapped by DC potentials applied on each of the three sections of the trap and/or the lenses on the front and back. Mass spectra are acquired using mass selective instability whereby RF amplitude applied to the rods is ramped upward, causing progressively higher m/*z* values to adopt an unstable *q* value along the q_z axis and become ejected in the x plane. The ions exit through narrow slits on

each x-axis rod toward conversion dynodes where they are detected using electron multipliers that can achieve gains in excess of 10^5, bolstering the sensitivity of the LTQ-based analysis [23].

The speed and sensitivity of the LTQ make it ideal for quickly collecting a large number of spectra for many ions, even if they are in femtomole to attomole abundance. However, this speed can only be practically performed at 1 m/*z* unit resolution which reduces mass accuracy and is not adequate to resolve isotopes. For this reason, LTQs are routinely placed in tandem with high-resolution mass analyzers like the FT-ICR and Orbitrap that may be routinely operated at resolutions >100,000 Full Width Half Max (FWHM) with mass accuracies of <5 parts-per-million (ppm).

The cylindrical ICR cell is situated inside a large magnetic field and a population of ions is injected and then trapped by front and rear trapping electrodes. The ions then receive a broadband excitation 'chirp' that sends them into a circular orbit perpendicularly to the magnetic field. The ions orbit at a cyclotron frequency that is proportional to their m/*z* values. The image current of the orbiting ions is detected by two detector plates, simultaneously yielding amplitude and frequency data for all ions after translation by Fourier transform. [23-25].

More recently, the Orbitrap has become the most popular high resolution mass analyzer for peptide and protein analyses as it does not require an expensive and high-maintenance superconducting magnetic to obtain high resolution mass spectra. Instead, ions are trapped around a central electrode using a purely electrostatic field. Ions are injected perpendicularly to and off-axis from the central electrode through a small inlet in the outer barrel-like electrode. They are electrostatically 'squeezed' closer to the central electrode by increasing the field strength and each m/*z* ion packet will begin to rotate around the central electrode and also oscillate along the central electrode axis without any further stimulation needed. The frequency

of axial oscillation is inversely proportional to the square root of m/*z* and is independent of the spatial distribution and energy of the ions. Analogous to detection in the ICR cell, the ions' image current that contains amplitude and axial oscillation frequency information may be detected and converted into mass spectra by Fourier transform [23, 26-28]. Taking advantage of the strengths of each mass analyzer in the LTQ-FT-ICR or LTQ-Orbitrap, it is common practice to acquire MS spectra of intact peptides or proteins in the high resolution analyzers and acquire MS/MS spectra for sequencing purposes in the LTQ.

Soft ionization through ESI enables the mass spectrometric analysis of proteins and peptides, but it does not offer any sequencing data due to a lack of fragmentation. Therefore, fragmentation techniques have been developed to randomly cleave peptides along their backbone to generate ladders of fragment ions differing in mass by each successive residue. In this way, the sequence of a peptide may be read in either direction by measuring the difference in mass between each successive ion fragment.

Collision-activated dissociation (CAD) is a common form of fragmentation in which the translational energy from multiple low-energy collisions between precursor ions and an inert bath gas (e.g. helium) is converted into vibrational energy imparted to the precursor ions. Random protonation of the amide nitrogens along the peptide backbone creates potential cleavage points. Once the energy from combinatorial collisions overcomes the necessary activation barrier, peptide bonds will be randomly cleaved to generate a series of complementary b and y ions, containing the N-terminal and C-terminal fragments of the peptide, respectively [29] (**Figure 1.2A&B**).

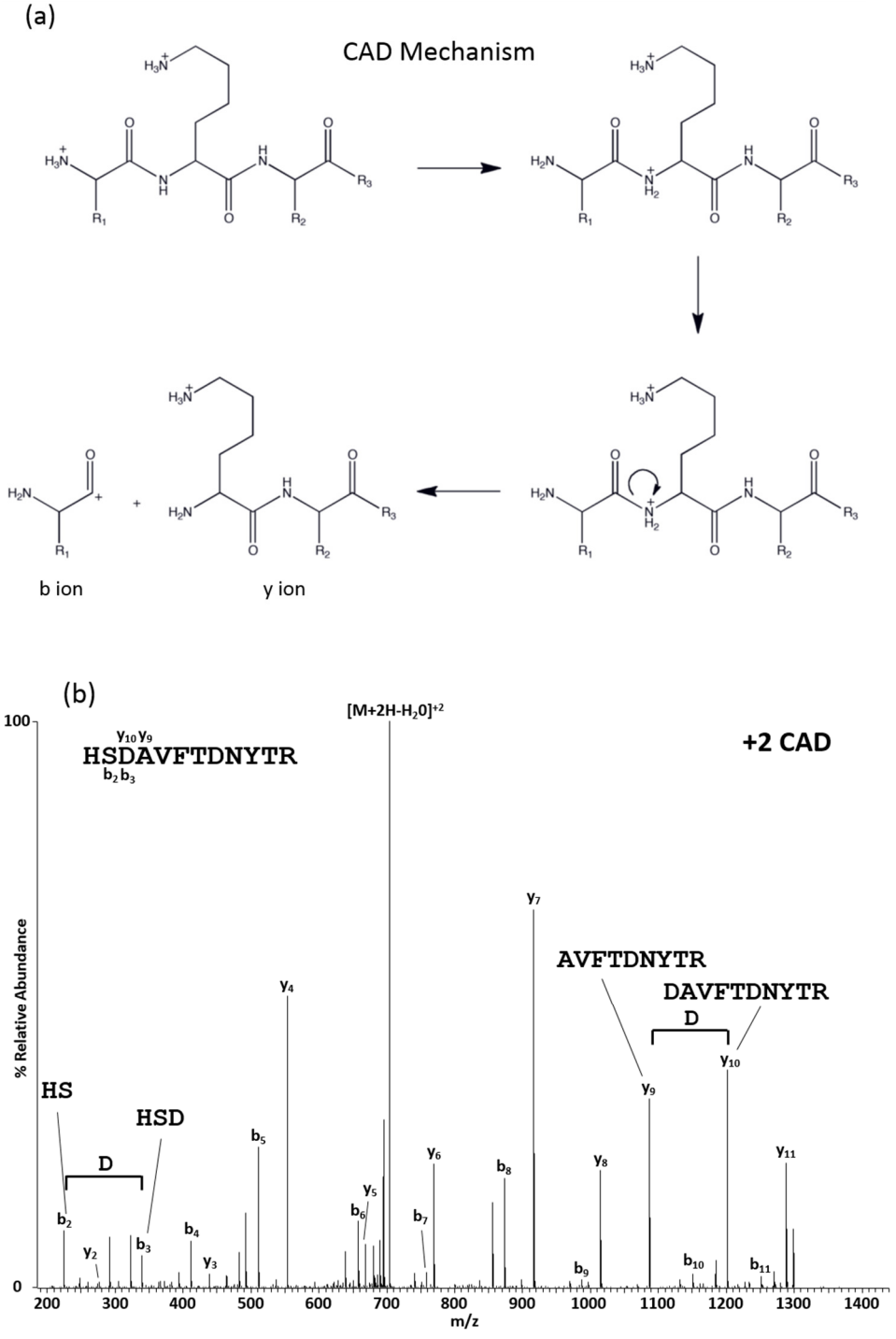
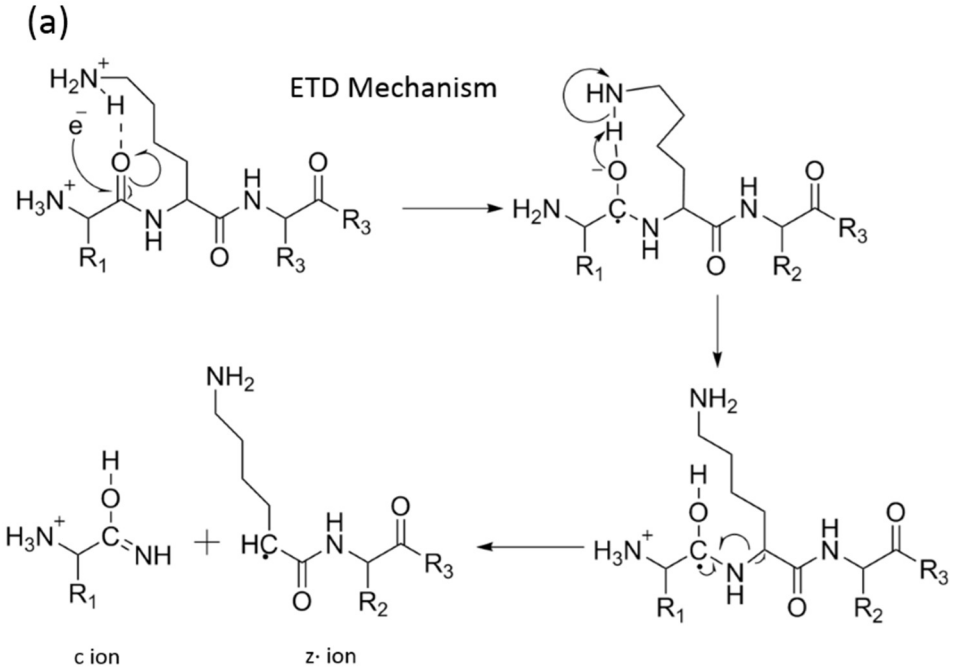


Figure 1.2 – *(a) The fragmentation mechanism of CAD, and (b) a CAD MS/MS spectrum of peptide HSDAVFTDNYTR.*

Unfortunately, ions with many charges and/or labile PTMs disrupt evenly distributed

protonation and the cleavage of peptide bonds, making CAD much less useful for long peptides and those that contain phosphorylation and O-glycosylation.

To circumvent this problem, Hunt and coworkers developed electron transfer dissociation (ETD) [30], an ion/ion analogue of electron-capture dissociation (ECD) [31] which unlike ECD, can be performed in the LTQ. Fragmentation is induced after a radical anion reagent (e.g. azulene) transfers an electron to a random peptide carbonyl group, initiating a cascade based on free radical chemistry that induces fragmentation of N-C_α bonds along the peptide backbone, generating a series of c and z ions (**Figure 1.3A&B**). ETD is complementary to CAD, because it provides the most optimal structural information for peptide ions of $z \geq 3$, and CAD is optimal for $z=2$ ions. Also, the ETD fragmentation mechanism is driven by the free radical electron vs.



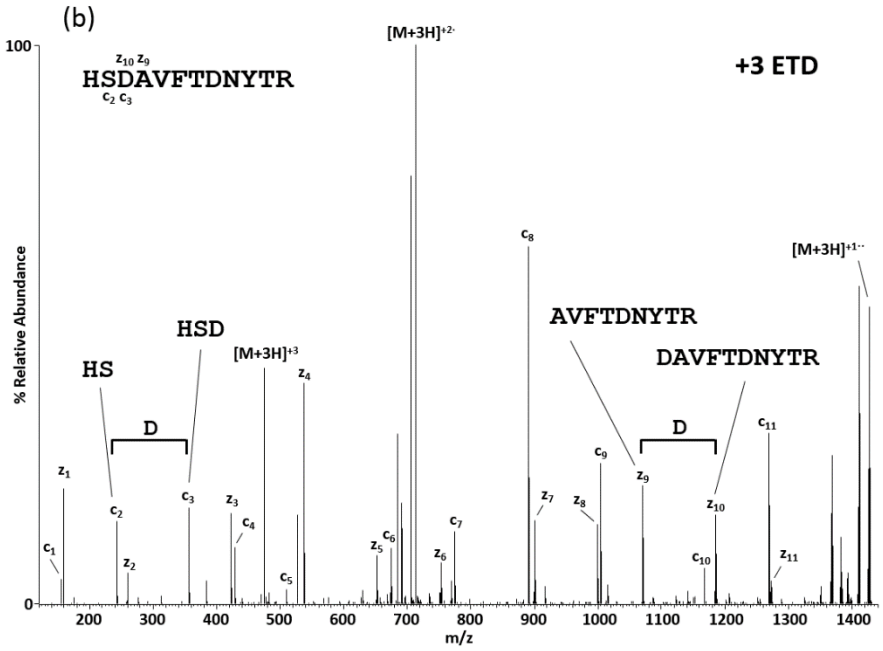


Figure 1.3 – *(a) The fragmentation mechanism of ETD, and (b) an ETD MS/MS spectrum of peptide HSDAVFTDNYTR.*

collisions with a bath gas, so in contrast to CAD, labile PTMs are preserved. Consequently, ETD is an essential MS/MS tool for detecting PTMs like phosphorylation and O-GlcNAcylation.

The following chapters describe the implementation of nanoflow micro-ESI LCMS using LTQ-FT-ICR and LTQ-Orbitrap mass spectrometers to characterize numerous novel PTMs on two nuclear proteins. Chapter 2 reports the post-translational modification of *Arabidopsis* protein repressor of *ga1-3* (RGA), a negative regulator of plant growth, by O-GlcNAc, phosphate, and a

previously unreported O-linked hexose. Chapter 3 reports on the phosphorylation of *Saccharomyces cerevisiae* DNA repair protein Rad9.

1.4 References

1. CRICK, F. H. On protein synthesis. Symp. Soc. Exp. Biol. 12, 138-163 (1958).

2. Crick, F. Central dogma of molecular biology. Nature 227, 561-563 (1970).

3. http://www.ncbi.nlm.nih.gov/books/NBK26830/.

4. Cox, J. & Mann, M. Quantitative, high-resolution proteomics for data-driven systems biology. Annu. Rev. Biochem. 80, 273-299 (2011).

5. Wilkins, M. R. et al. From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis. Biotechnology (N. Y) 14, 61-65 (1996).

6. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. Nature 431, 931-945 (2004).

7. Walsh, C. T., Garneau-Tsodikova, S. & Gatto, G. J.,Jr. Protein posttranslational modifications: the chemistry of proteome diversifications. Angew. Chem. Int. Ed Engl. 44, 7342-7372 (2005).

8. Creasy, D. M. & Cottrell, J. S. Unimod: Protein modifications for mass spectrometry. Proteomics 4, 1534-1536 (2004).

9. Farley, A. R. & Link, A. J. Identification and quantification of protein posttranslational modifications. Methods Enzymol. 463, 725-763 (2009).

10. Hart, G. W., Slawson, C., Ramirez-Correa, G. & Lagerlof, O. Cross talk between O-GlcNAcylation and phosphorylation: roles in signaling, transcription, and chronic disease. Annu. Rev. Biochem. 80, 825-858 (2011).

11. Zhao, Y. & Jensen, O. N. Modification-specific proteomics: strategies for characterization of post-translational modifications using enrichment techniques. Proteomics 9, 4632-4641 (2009).

12. Ashcroft, A. E. & Barnett, N. W. in Ionization Methods in Organic Mass Spectrometry X001-X004 (The Royal Society of Chemistry, 1997).

13. Barber, M., Bordoli, R. S., Sedgwick, R. D. & Tyler, A. N. Fast atom bombardment of solids as an ion source in mass spectrometry. Nature 293, 270-275 (1981).

14. Karas, M. & Hillenkamp, F. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. Anal. Chem. 60, 2299-2301 (1988).

15. Tanaka, K. et al. Protein and polymer analyses up to m/z 100 000 by laser ionization time-of-flight mass spectrometry. Rapid Communications in Mass Spectrometry 2, 151-153 (1988).

16. Dass, C. in Fundamentals of Contemporary Mass Spectrometry 15-65 (John Wiley & Sons, Inc., 2007; 2006).

17. Dole, M. et al. Molecular Beams of Macroions. J. Chem. Phys. 49, 2240-2249 (1968).

18. Whitehouse, C. M., Dreyer, R. N., Yamashita, M. & Fenn, J. B. Electrospray interface for liquid chromatographs and mass spectrometers. Anal. Chem. 57, 675-679 (1985).

19. Mack, L. L., Kralik, P., Rheude, A. & Dole, M. Molecular Beams of Macroions. II. J. Chem. Phys. 52, 4977-4986 (1970).

20. Iribarne, J. V. & Thomson, B. A. On the evaporation of small ions from charged droplets. J. Chem. Phys. 64, 2287-2294 (1976).

21. Nguyen, S. & Fenn, J. B. Gas-phase ions of solute species from charged droplets of solutions. Proc. Natl. Acad. Sci. U. S. A. 104, 1111-1117 (2007).

22. Schwartz, J. C., Senko, M. W. & Syka, J. E. P. A two-dimensional quadrupole ion trap mass spectrometer. J. Am. Soc. Mass Spectrom. 13, 659-669 (2002).

23. Dass, C. in Fundamentals of Contemporary Mass Spectrometry 67-117 (John Wiley & Sons, Inc., 2007; 2006).

24. Gross, M. L. & Rempel, D. L. Fourier transform mass spectrometry. Science 226, 261-268 (1984).

25. Holliman, C. L., Rempel, D. L. & Gross, M. L. Detection of high mass-to-charge ions by Fourier transform mass spectrometry. Mass Spectrom. Rev. 13, 105-132 (1994).

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

27. Hardman, M. & Makarov, A. A. Interfacing the orbitrap mass analyzer to an electrospray ion source. Anal. Chem. 75, 1699-1705 (2003).

28. Hu, Q. et al. The Orbitrap: a new mass spectrometer. J. Mass Spectrom. 40, 430-443 (2005).

29. Hunt, D. F., Zhu, N. Z. & Shabanowitz, J. Oligopeptide sequence analysis by collision-activated dissociation of multiply charged ions. Rapid Commun. Mass Spectrom. 3, 122-124 (1989).

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

31. Zubarev, R. A., Kelleher, N. L. & McLafferty, F. W. Electron capture dissociation of multiply charged protein cations. A nonergodic process. J. Am. Chem. Soc. 120, 3265-3266 (1998).

Chapter 2: Characterization of O-GlcNAcylation, phosphorylation, and a novel O-linked nuclear PTM, O-hexose, on DELLA protein RGA

2.1 Introduction

2.1.1 Post-translational modification of proteins is ubiquitous and essential

Post-translational modifications of proteins are essential for the cell to perform many biological functions. More than 400 PTMs have been discovered [1], and this number could continue to increase as detection methods improve. A few of the most commonly studied PTMs include phosphorylation, methylation, acetylation, ubiquitination, sumoylation, and glycosylation. These and other PTMs regulate cellular function by covalent addition to proteins, the main biological components of cellular structure and function, thereby altering the proteins' physical and chemical properties. In this way, PTMs can influence protein activity, cellular location, and stability [2].

The reversible phosphorylation of nuclear and cytoplasmic proteins was among the earliest PTMs discovered and it is one of the most ubiquitous PTMs known. There are ~500 human kinases, the proteins responsible for catalyzing protein phosphorylation from an ATP donor molecule, and nearly 150 human phosphatases, which catalyze hydrolytic phosphate removal [3]. It is estimated that 2-3% of all eukaryotic genes code for protein kinases [4] and that 1/3 of the proteome is phosphorylated [5]. Mammalian kinases modify nucleocytoplasmic proteins' Ser, Thr, and Tyr residues with phosphate to regulate most types of cellular function and their response to multiple environmental factors. This mediates cellular processes like proliferation, apoptosis, migration, RNA transcription, and protein translation. Consequently, dysregulation of protein kinases and phosphates is associated with many pathologies including cancer, autoimmune diseases, metabolic disorders, and infections [6]. The importance of

phosphorylation makes it a primary focus for biochemical and proteomic studies. Yet, an analogous nucleocytoplasmic PTM, O-GlcNAc, was discovered less than 30 years ago and reveals an additional post-translational mechanism by which nuclear and cytoplasmic proteins are regulated in the cell.

2.1.2 Post-translational modification O-linked β-N-acetylglucosamine

O-linked β-N-acetylglucosamine, or O-GlcNAc, is a reversible monosaccharide PTM of nuclear and cytoplasmic proteins that forms an O-linkage to Ser and Thr, cycling on and off on a timescale similar to phosphorylation [7]. O-GlcNAcylation was discovered in 1984 when Hart and coworkers were probing for terminal GlcNAc moieties on the surface of human lymphocytes using [^{3}H]-galactose labeling. A portion of the radiolabel was covalently linked to intracellular proteins, and the signal could be removed by alkali-induced β-elimination, indicating O-linkage and revealing the existence of the nucleocytoplasmic PTM, O-GlcNAc [8]. Previously, it was widely believed that protein glycosylation was limited to the exterior of the cell and the luminal face of intracellular compartments. O-GlcNAcylation is a ubiquitous protein PTM in mammalian cells, with over 3000 O-GlcNAcylated proteins detected to date. Elucidating the scope and role of this glyco-PTM on nuclear and cytoplasmic proteins has been the focus of many biochemical and proteomic studies for the last 25 years.

O-GlcNAcylation differs from classic forms of protein glycosylation in several ways: (1) it modifies nuclear and cytoplasmic proteins rather than traditional glycosylation found on proteins localized to the extracellular membrane or matrix, lumen of organelles, or Golgi/ER, (2) it exists as a monosaccharide and is not elongated, contrary to the complex oligosaccharides lining the extracellular membrane, (3) O-GlcNAcylation is a dynamic modification, as opposed to static like most complex extracellular glycans, (4) O-GlcNAc forms an O-linkage to the oxygen

atom on Ser and Thr but GlcNAc was previously known to form a direct bond to proteins only by an N-linkage to the amide nitrogen on Asp, and (5) a consensus sequence for O-GlcNAc has not been determined while N-linked GlcNAcylation follows the consensus Asp-X-Ser/Thr (where X is any amino acid except Pro and Asp) [9]. In these ways, O-GlcNAc much more closely resembles phosphorylation. Both are dynamic nucleocytoplasmic protein PTMs, O-linked to Ser and Thr residues, usually substoichiometrically abundant, and are regulated by separate enzymes responsible for their addition and removal **(Figure 2.1)**.

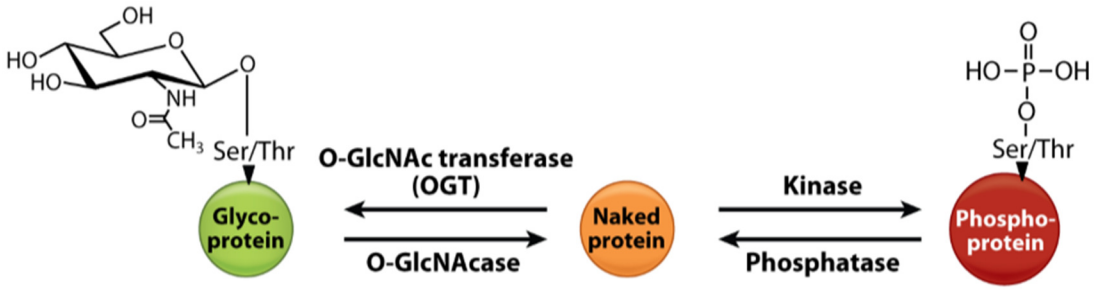


Figure 2.1 – *O-GlcNAc transferase and O-GlcNAcase respectively install and remove O-GlcNAc from Ser and Thr residues, analogous to the addition and removal of phosphate from these residues by kinases and phosphatases Figure reproduced from [7].*

However, in contrast to the many kinases and phosphatases responsible for phosphorylation addition and removal, just two metazoan genes encode the enzymes that cycle O-GlcNAcylation on and off Ser and Thr residues. These two respective enzymes are O-GlcNAc transferase (OGT)

and O-GlcNAcase (OGA), and are highly conserved in virtually all metazoans studied, from *C. elegans* to humans [10] **(Figure 2.2)**.

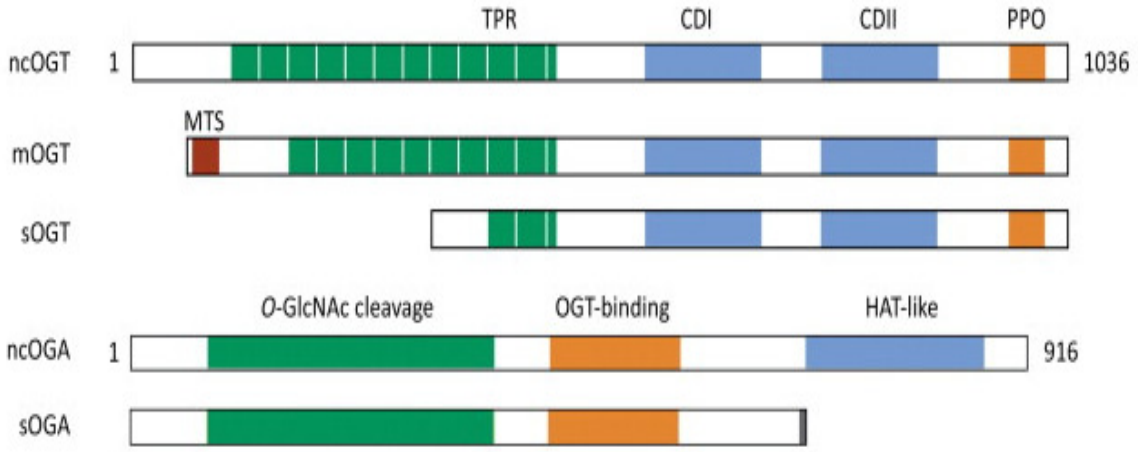


Figure 2.2 - *O-GlcNAc transferase (OGT) has three known human isoforms and O-GlcNAcase (OGA) has two, formed by alternative splicing and multiple start codons. The longest and shortest isoforms (ncOGT and sOGT) are found in the nucleus and the cytoplasm, and a third isoform (mOGT) targets the inner mitochondrial membrane form by an N-terminal mitochondrial targeting sequence (MTS). Each OGT isoform has a unique number of N-terminal tetratricopeptide repeats (TPR), and a common C-terminal region containing catalytic domains I and II (CDI and II) in addition to a phosphoinositide-binding domain (PPO). The nucleocytplasmic and main form of OGA bears an N-terminal O-GlcNAc cleavage domain, an OGT-binding domain, and a C-terminal histone acetyltransferase (HAT)-like domain. A shorter HAT-like domain-deficient splice variant has been detected in the nucleus. Figure reproduced from [10].*

O-GlcNAc transferase catalyzes the addition of O-GlcNAc to Ser and Thr

OGT is essential in mammals. Loss of the *OGT* gene is embryonic lethal in mouse stem cells. OGT is ubiquitously expressed in all mammalian tissue, but is most abundant in the brain and pancreas. The single human OGT gene is located on the X chromosome. A full length 110-kDa OGT isoform (nucleocytoplasmic, ncOGT) and shorter 78-kDa isoform (sOGT), formed by a combination of alternative RNA splicing and start codons, both localize to the nucleus and cytoplasm, although they are more heavily localized in the nucleus [11]. They have been reported to form a heterotrimer composed of two ncOGT subunits and one sOGT subunit in rat liver, but most tissues including brain and pancreas, contain only the full length ncOGT isoform. A third form, mOGT, has been detected localizing to the inner membrane of mitochondria, and its

function is less characterized [12]. One of the major characteristics of OGT is the multiple 34 residue tetratricopeptide repeats (TPRs) with the loosely conserved sequence, WLGYAFAP. Crystal structures demonstrate that each pair of TPRs form anti-parallel α-helices that together form a super-helical structure important for protein-protein interaction [13, 14]. Using its TPRs to form many different protein complexes is thought to provide the major means by which OGT is able to recognize and modify Ser and Thr residues on a wide variety of sequences. Several studies have collectively identified several OGT-interacting proteins. These include OGA which removes O-GlcNAc, Trak1 that recruits OGT to RNA polymerase II, and the arginine methyltransferase CARM1 [15]. Overexpression of OGT alters the acetylation and methylation patterns of histones [7].

O-GlcNAcase catalyzes the removal of O-GlcNAc from Ser and Thr

OGA, also known as Hexaminidase C, was identified on human chromosome 10 after the discovery of OGT and is responsible for the removal of O-GlcNAc from Ser and Thr residues [16]. OGA, like OGT, is also expressed by a single highly conserved gene among mammalian species, and homologs also exist in *C. elegans*, and *Drosophila.* OGA too localizes to the nucleus and cytoplasm, but contrary to OGT, the canonical full length isoform (ncOGA) is more concentrated in the cytoplasm. It contains the N-terminal catalytic domain, an OGT binding domain, and a C-terminal histone-acetyltransferase (HAT)-like domain [16]. The HAT-like domain is reported to demonstrate acetyltransferase activity *in vitro* on both all four core histones both free or DNA-bound in the context of oligonucleosome arrays [17]. This activity was only observed when the O-GlcNAcase was expressed in mammalian as opposed to bacterial cells. A shorter isoform (sOGA) which lacks the C-terminus and HAT-like domain has been reported to localize to the nucleus [18]. The sOGA isoform displays activity to O-GlcNAcylated proteins, but at a lesser rate vs. ncOGA

(400-fold less efficient) [19]. Both OGA isoforms are subject to effective inhibition by small molecule inhibiters like PUGNAc, NAG-thiazoline, and NButGT (**Figure 2.3**), implying similar catalytic reaction mechanisms. Much research is presently conducted toward the goal of

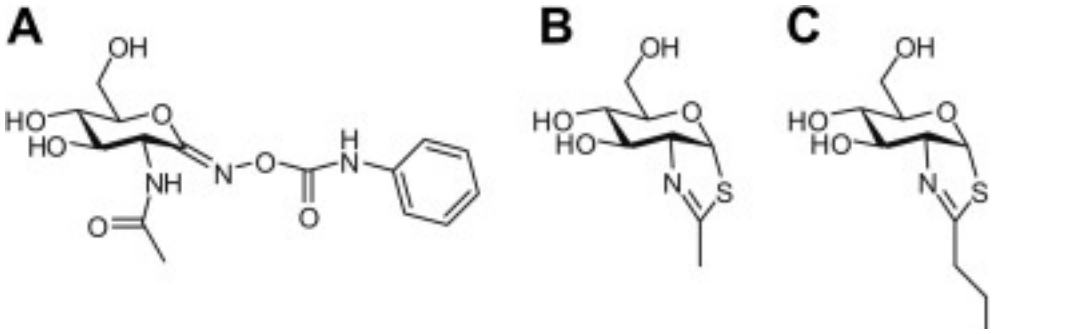


Figure 2.3 – *Three OGA inhibitors, (A) PUGNAc, (B) NAG-thiazole, and (C) NButGT that effectively inhibited the ncOGA and sOGA isoforms. Figure reproduced from [19].*

producing very effective and specific OGA inhibitors for the potential therapeutic benefit of controlling cellular OGA levels.

O-GlcNAc donor molecule, UDP-GlcNAc, is produced by the Hexosamine Biosynthetic Pathway

OGT has a relatively high binding affinity (Km = 545 nM) for O-GlcNAc's donor molecule, UDP-GlcNAc, which is typically present in the cell at 0.1-1 mM. This molecule is produced by the Hexosamine Biosynthetic Pathway (**Figure 2.4**), a process tied to nearly every major metabolism pathway in the cell. OGT activity is dependent on UDP-levels from the low nM range to >50 mM [20]. Accordingly, UDP-GlcNAc and therefore O-GlcNAcylation levels are sensitive to many factors such as insulin, nutrients, and cellular stress. For this reason, O-GlcNAcylation is believed to act as a "nutrient sensor", modulating cellular signaling and transcription regulatory pathways in response to these factors [21-24].

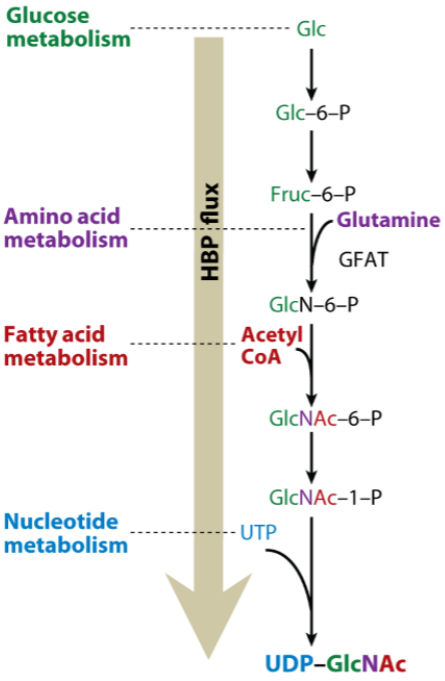


Figure 2.4 - All major forms of metabolism feed the Hexosamine Biosynthetic Pathway which produces the O-GlcNAc donor molecule, UDP-GlcNAc. Reproduced from 7.

O-GlcNAcylation occurs on wide variety of proteins involved in virtually all intracellular functions

Nucleocytoplasmic proteins belonging to virtually all functional classes (**Figure 2.5**) have been found to be O-GlcNAcylated, regulating numerous processes that include transcription,

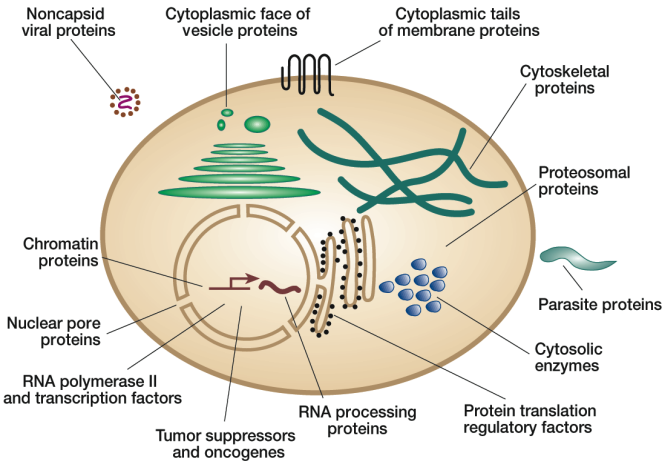


Figure 2.5 – *Proteins from virtually every functional class are O-GlcNAcylated within the cell. Figure reproduced from [20].*

translation, protein trafficking, and protein degradation (**Figure 2.6**). Dysregulation of O-GlcNAcylated proteins and altered cellular levels of O-GlcNAc and its regulating enzymes has negative implications for several pathologies: cancer, diabetes, neurodegenerative disease, and cardiovascular disease [25]. Transcription factors are amongst the most commonly O-GlcNAcylated proteins identified. RNA polymerase II and most of its basal transcription factors are O-GlcNAcylated, in addition to other transcription factors including Sp1, estrogen receptors, STAT5, NF-κB, p53, YY1, Elf-1, c-Myc, Rb, PDX-1, and CREB [20]. The list of identified O-GlcNAcylated proteins also includes many nuclear pore proteins, RNA-binding proteins, cytoskeletal proteins, and enzymes including kinases, phosphatases, and OGT itself.

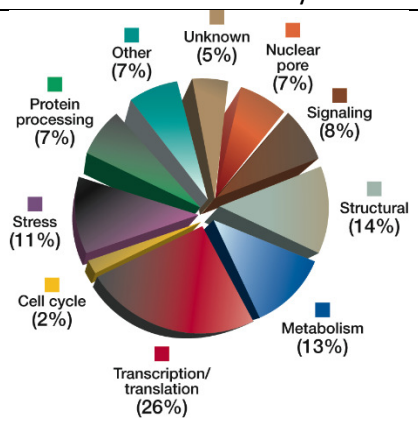


Figure 2.6 - *The largest category of O-GlcNAcylated proteins are involved in transcriptional regulation, followed by those involved in cytoskeletal regulation and structure, energy metabolism, cellular stress, and signaling. Figure reproduced from* [20].

Cross Talk between O-GlcNAcylation and phosphorylation

Due to their many similarities, the relationship between O-GlcNAcylation and phosphorylation is of particular interest and is one of the most studied aspects of the glyco-PTM. There are several mechanisms by which the two PTMs exhibit regulation over each other. As mentioned, OGT modifies numerous kinases and phosphates. Also, OGT and OGA may be tyrosine- and serine-phosphorylated, respectively, in their catalytic domains [26]. Notably, tyrosine phosphorylation by the insulin receptor has been shown to activate mammalian OGT, initiating a process by which OGT down-regulates insulin signaling [27]. OGT may also be activated by Src kinase and calcium calmodulin kinase IV (CAMKIV) amongst others [28]. Inhibition of kinase GSK3β increased O-GlcNAcylation on many cytoskeletal and heat shock proteins and decreased O-GlcNAc levels on many transcription factors and RNA-binding proteins [29]. Likewise, when O-GlcNAcase is inhibited, overall O-GlcNAc levels increase three-fold, and nearly all 700 phosphorylation sites determined to be actively cycling in the cell demonstrated either increased or decreased when compared to (empty vector?) controls [30]. A large-scale enrichment of phosphorylated and O-GlcNAcylated proteins from M-phase HeLa cells following a two-fold

overexpression of OGT during revealed over 400 O-GlcNAc modified peptides, comparable to the number of phosphorylated peptides present. [31]. OGT overexpression reduced the majority of phosphorylation sites, but a portion increased relative to basal levels. Ser and Thr residues on many cytoskeletal proteins were reciprocally modified with either PTM on the same site, while most transcription factors showed reciprocal occupancy of both PTMs on proximal sites.

2.1.3 Evidence for O-GlcNAcylation in Plants

Two OGTs are found in plants: SPINDLY and SECRET AGENT

Despite the increasing amount of knowledge being gained regarding metazoan O-GlcNAcylation, the body of evidence for this PTM in other species, including plants, is much less developed. Still, two OGTs are now known to exist in plants, primarily through the study of *Arabidopsis thaliana* [32]. The first to be discovered, SPINDLY (SPY), was originally identified as a gene locus important in gibberellin (GA) signaling [33]. GA's are plant growth hormones that regulate many aspects of plant development: germination, growth, flowering, and seed development [34]. Experiments with mutant *spy* alleles demonstrated that SPY negatively regulates GA signaling and plant development and also plays a role in circadian and light signaling [33]. When human OGT was discovered, researchers recognized the high homology of hOGT with SPY (**Figure 2.7**) and identified SPY as an O-GlcNAc transferase. A separate *Arabidopsis* OGT was later discovered and named SECRET AGENT [32]. The catalytic domain of SEC

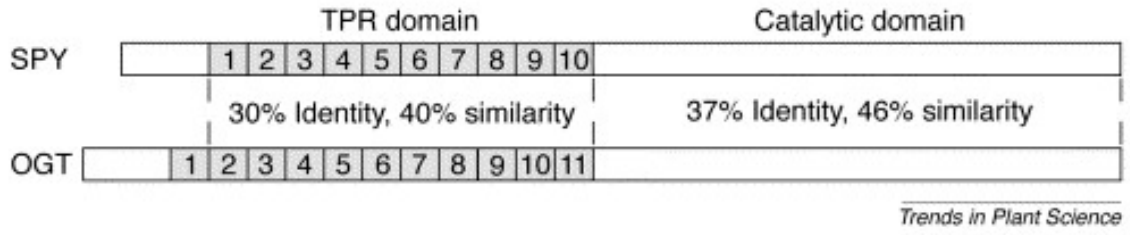


Figure 2.7 - *Arabidopsis SPY (top) and human OGT (bottom) have high homology in both their TPR and catalytic domains. Reproduced from* [35].

is even more similar (>50% similarity) to animal OGT than SPY. In addition, SEC's TPR domain is contiguous like that of animal OGT while SPY has amino acid insertions between the second and fifth TPRs. Direct evidence for O-GlcNAcylation by these OGTs is limited. Both appear to have the ability to auto-O-GlcNAcylate themselves, like animal OGT [36]. Direct mass spectral data demonstrates that several sites of the Plum Pox Virus capsid protein are O-GlcNAcylated, enhancing viral infection [37, 38]. SEC has been shown to function in germination, rosette leaf growth/production rate, and inflorescence stem height, but evidence for its role in GA signaling is weak. However, the two OGTs appear to be at least partially functionally redundant. The loss of either was still viable though *spy sec* double knockout mutants were lethal in early development, indicating they are essential for gametogenesis and embryogenesis, similar to animal OGT [32, 39]. It is possible that SPY has a greater ability to compensate for the loss of SEC than vice versa, making the effects of *sec* phenotypes more difficult to study.

Evidence for O-GlcNAcylation of DELLA protein Repressor of ga1-3 (RGA)

The repressor of *ga1-3* (RGA) is one of five *Arabidopsis* growth regulator proteins, characterized in part by their conserved N-terminal DELLA (named for the sequence Asp-Glu-Leu-Leu-Ala) and C-terminal GRAS (named after the first three isolated proteins GAI, RGA, and

SCR) domains [40, 41] (**Figure 2.8**).

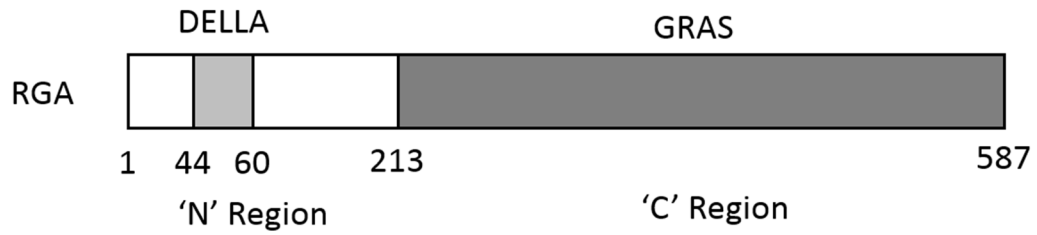


Figure 2.8 – *A schematic of RGA is pictured with the DELLA domain (light gray) in the N-terminal region and the GRAS domain (dark gray) in the C-terminal. All five Arabidopsis DELLA orthologs are highly conserved.*

The *ga1-3 Arabidopsis* mutant is the most studied mutant of the GA1 gene, whose loss of function abolishes GA biosynthesis ability [40]. Sun and coworkers performed screens of *Arabidopsis* proteins to find repressors of the mutant *ga1-3* phenotype. Loss of function in RGA was found to repress the *ga1-3* dwarf phenotype, partially restoring its ability to flower, and elongate its stem [42]. This demonstrated a similar phenotype to *spy* knockout mutants, already known to suppress development. Interestingly, *rga spy* double knockout mutants resulted in a virtually complete rescue of the *ga1-3* dwarf phenotype (**Figure 2.9**). This could imply that OGT

Figure 2.9 – *35-day-old mutant ga1-3 Arabidopsis plants lacking GA biosynthetic components necessary for proper growth (far left) were subject to additional mutation in the rga and/or spy locus (middle three). Wildtype (Ler) is displayed on the far right. DELLA protein RGA and OGT protein SPY each have a role in negatively regulating growth. Remarkably, rga spy double knockout mutants virtually restore ga1-3 to the Ler phenotype. Figure reproduced from* [42].

directly regulates RGA. Sun and coworkers also proposed that the two proteins each regulate separate but closely related branches of the GA signaling network with some overlapping functions [42]. Compared with the other DELLA proteins found in *Arabidopsis*, RGA appears to have a central role in developmental regulation, controlling germination, flowering, and stem elongation (**Figure 2.10**) [43].

DELLA proteins do not have a canonical DNA binding motif, indicating that they regulate transcription through interactions with other transcription factors [44]. GA signaling is shown to promote plant growth by degradation of DELLAs. The mechanism by which this occurs has been

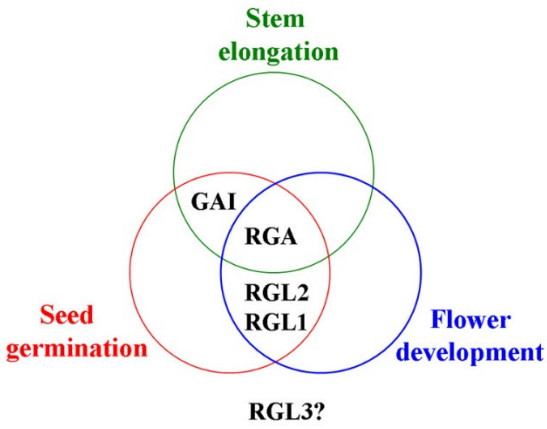


Figure 2.10 - *Five DELLA proteins have been discovered in Arabidopsis. RGA plays a central role as a negative growth regulator. Figure reproduced from [43].*

recently studied and elucidated. Upon GA binding to the GA receptor Gibberellin Insensitive Dwarf 1 (GID1), the affinity of the DELLA domain for GID1 also increases, causing it to bind GID1 as well.

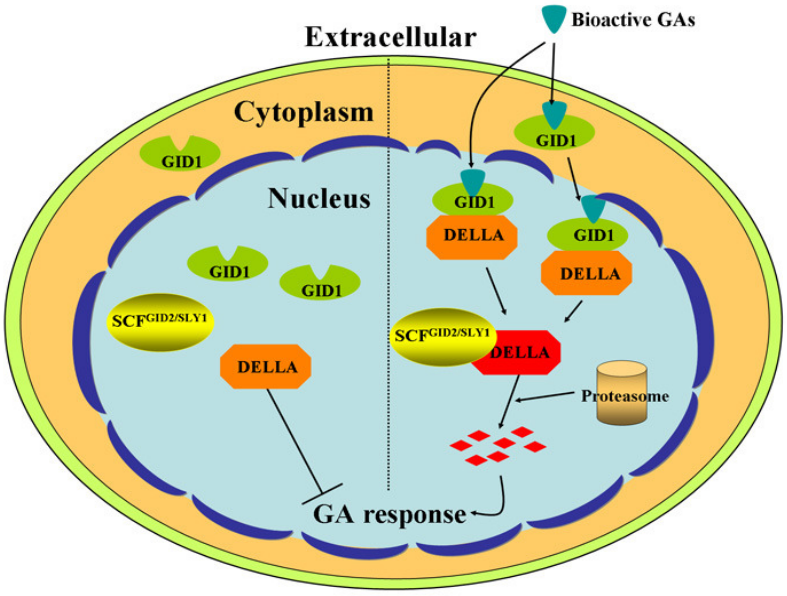


Figure 2.11 – *GAs are reported to stimulate plant development by inducing degradation of the repressor DELLA proteins. Figure reproduced from [43].*

This recruits E3 ubiquitin ligase to mark DELLA for degradation through the Skp, Cullin, F-box containing (SCF) complex, using F-box protein SLEEPY1 (SLY1) (**Figure 2.11**).

No direct evidence exists for the post-translation modification of RGA. Neither has any *Arabidopsis* or any other plant protein been directly characterized for O-GlcNAcylation by SPY or SEC. To enable a more complete understanding of the mechanisms by which RGA and OGT act to regulate plant growth and GA signaling down-regulates RGA, this study aims to fully characterize RGA for O-GlcNAcylation and other PTMs by mass spectrometry.

2.2 Materials and Methods

2.2.1 Proteolytic Digestions and Sample Preparation for Mass Spectrometric Analysis

Seven separate preparations of 6xHis-3xFLAG-tagged *Arabidopsis* RGA overexpressed in tobacco were tandem-purified (10-50 pmol each) and reduced on anti-FLAG M2 agarose beads (Sigma) (1 hour) with dithiothreitol (DTT; Sigma) at a molar ratio of 20: 1 (DTT: cysteine), carbamidomethylated (45 min) with iodoacetamide (IAA; Sigma) at a molar ratio of 3: 1 (IAA: DTT), digested with trypsin (Promega), endoproteinase LysC (Roche), or endoproteinase AspN (Roche) (7 hours) at a molar ratio of 20: 1 (RGA: protease), then quenched with glacial acetic acid to pH 3-4 for storage at -35°C. In four preparations, *Arabidopsis* O-GlcNAc transferase SEC was co-overexpressed with WT *Arabidopsis* RGA. In a fifth preparation, *Arabidopsis* SEC was co-overexpressed with *Arabidopsis* RGA containing a lysine insertion between G185 and G186 (G185_G186insK). In a sixth preparation, *Arabidopsis* kinase CK1 was co-overexpressed with *Arabidopsis* RGA-G185_G186insK. In a seventh preparation, *Arabidopsis*RGA-G185_G186insK was overexpressed in the absence of SEC or CK1. Reduction, alkylation, and digestion were carried out at room temperature in volumes ≤100 uL 100 mM ammonium bicarbonate pH 7.9.

2.2.2 Subcleavage by CNBr

A subset of AspN-digested WT RGA was cleaved C-terminally to Met using cyanogen bromide (CNBr). The reaction was performed as described by [1] with slight adjustments. Peptides were dissolved in 20uL 70% TFA in water (*v*/*v*), one small crystal <1mg of CNBr was added to the solution (a large excess over total Met), and the reaction proceeded 16 hours at room temperature. Following cleavage, the peptides were dried down and resuspended in 0.1% acetic acid.

2.2.3 Derivatization of Carboxylic Acids with Histamine

Peptides' carboxylic acid groups were derivatized with histamine as described in [45]. Peptides were first purified from proteases and desalted by loading (<1 uL/min) onto a 360 µm outer diameter (o.d.) × 75 µm inner diameter (i.d.) fused silica (Polymicro Technologies, LLC) column packed with 6 cm of C18 reverse-phase resin (5–20 µm diameter, 120 Å pore size, YMC Co., Ltd.) behind a 1-2mm Kasil® 1624 (PQ Corporation) frit. The column was rinsed with 0.1M acetic acid for 20 min then peptides were gradient eluted into a clean Eppendorf tube for 70 min using an LC gradient of 0-80% Solvent B in 60 min, hold 80% Solvent B for 30 min, 80-100% Solvent B in 10 min, 100%-0% Solvent B in 5 min, 100% Solvent A for 20 min (Solvent A: 0.1M acetic acid in water and Solvent B: 70% acetonitrile in 0.1M acetic acid in water). Peptides were dried down and resuspended in 20 uL 1 M histamine in 1 M pyridine-HCl and 5uL 0.1 M EDC in 1 M pyridine-HCl. The reaction was sonicated for 2 hours at room temperature then left to proceed without sonication for 14 hours. Following reaction, peptides were loaded directly onto the pre-column.

2.2.4 HPLC Capillary Analytical and Pre-column Assembly

Pre-columns were constructed from 8 cm of 360 µm o.d. × 75 µm i.d. fused silica and packed with 4 cm of C18 reverse-phase resin (5–20 µm diameter, 120 Å pore size, YMC Co., Ltd.) behind a 1-2mm Kasil® 1624 (PQ Corporation) frit. Analytical columns were constructed from 10 cm of 360 µm o.d. × 50 µm i.d. fused silica and packed with 7 cm of C18 resin (5 µm diameter, 120 Å pore size, YMC Co., Ltd.) behind a Lithisil® (PQ Corporation) frit, prior to being equipped with a laser-pulled electrospray emitter tip as previously described [3]. Columns packed with C4/C8 (3 µm diameter, 120 Å pore size, YMC Co., Ltd.), Zirchrome®-CARB, and POROS® R2 were prepared in the same manner. Tens of picomoles of standard peptide Angiotensin I (Sigma) were loaded and eluted from columns to optimize their chromatographic resolution and sensitivity by occupying non-specific binding regions before use on RGA samples.

2.2.5 Sample loading and LC Methodology

Prior to mass spectrometric analysis, each digest was dried down and resuspended in 0.1% acetic acid (pH 3) in water (*v*/*v*). Typically, a 0.5-1 pmol fraction of sample + 100fmol of internal standard peptides Angiotensin I (Sigma) and Vasoactive Intestinal Peptide fragment 1-12 (Sigma) in a volume ≤ 10uL was pressure loaded onto a pre-column at a flow rate of <1µL/min followed by a 15 min. desalting rinse with 0.1M acetic acid at several µL/min. The pre-column was butt-connected to the analytical column with a 2 cm Teflon sleeve (0.060 in o.d. x 0.012 in i.d.; Zeus Industrial Products). Trypsin-, AspN-, or AspN/CNBr-generated peptides were gradient eluted and electrosprayed into the mass spectrometer at a split-flow-generated rate of 60 nL/min by an 1100 series binary LC pump (Agilent Technologies) using a linear LC gradient of 0-60% Solvent B in 60 min, 60-100% Solvent B in 4 min, hold 100% Solvent B for 4

min, 100%-0% Solvent B in 4 min, 100% Solvent A for 20 min (Solvent A: 0.1M acetic acid in water and Solvent B: 70% acetonitrile in 0.1M acetic acid in water).

2.2.6 Phosphorylated Peptide Enrichment by Immobilized Metal Affinity Chromatography

A 5 pmol portion each of AspN- and trypsin-digested RGA-WT(+SEC) were combined and enriched for phosphorylation by Immobilized Metal Affinity Chromatography (IMAC). In addition, 5 pmol of AspN- and trypsin-digested RGA-G185_G186insK (+CK1) were separately enrichment by IMAC. Prior to enrichment, carboxylic acids were blocked by esterification with methanol to prevent them from preferentially binding to the IMAC column. Peptides were dissolved in 80 uL of a 6.25: 1 acetyl chloride: anhydrous methanol mixture that had been stirred for 5 min, the peptides and reaction mixture were allowed to stand 1 hr, dried down, and the process was repeated. Esterified samples were stored at -35°C for <24 hrs until enrichment by IMAC.

An 8cm 360 µm o.d. × 100µm i.d. fused silica column equipped with a 1 mm Kasil® 1624 frit was packed with 5 cm of POROS® MC 20 beads (20 uM diameter; Applied Biosystems™). The IMAC column was washed for 10 min with the following at 20 µL/min in this order: water, 50mM EDTA (pH 9), and water. The column was activated 3x with 100mM $FeCl_3$ for 10 min at 20 µL/min. Between each round of activation, the $FeCl_3$ was allowed to sit for 3 min on column without flow. Excess $FeCl_3$ was removed by washing with 25 µL 0.01% acetic acid at <1 µL/min.

Peptide samples were reconstituted in a 1: 1: 1 mixture of 0.01% acetic acid: acetonitrile: methanol, then loaded onto the IMAC column at <1 µL/min. The 1: 1: 1 mix was added to the empty sample tube again and the loading process repeated. The IMAC column was washed with 20 µL of 1: 1: 1 mix, followed by 20 µL of 0.01% acetic acid at <1 µL/min. The IMAC

column was connected to a pre-column, constructed as previously described, and enriched phosphorylated peptides were eluted and captured on the pre-column using 15 µL 250 mM ascorbic acid at 1 µL/min. Combined columns were washed with 6 µL 0.1% acetic acid, the pre-column was removed, and the wash was repeated on the pre-column only at 20 µL/min. The pre-column was connected to an analytical column, 100 fmol of Angiotensin and Vasoactive Intestinal Peptide fragment 1-12 were loaded onto the columns in 0.1% acetic acid, and LCMS analysis was conducted as described.

2.2.7 Mass Spectrometric Analysis

Mass analysis was completed with a modified ETD-enabled high-resolution LTQ-FT or LTQ-Orbitrap mass spectrometer (Thermo Scientific) using an acquisition method consisting of one high resolution MS^1 scan {resolving power of 50,000 (FT Ultra, Thermo Scientific) or 60,000 (Orbitrap, Thermo Scientific) at m/z 400} acquired in the FT-ICR or Orbitrap followed by 8-10 collision-activated dissociation (CAD) and/or electron transfer dissociation (ETD) MS^2 scans acquired with the LTQ operating in either the data-dependent or -independent mode. Data-dependent parameters included a repeat count of 2-3, repeat duration of 20-30 s, and exclusion list duration of 20-30 s. MS^2 parameters for ETD scans included 30-50 ms reaction times, 3 m/z precursor isolation window, charge state rejection "on" for +1 (and +2 for ETD only analyses) and unassigned charge state precursor ions, 5×10^5 FTMS (Orbitrap) or 1×10^6 (FT Ultra) full automated gain control target, 1×10^4ITMS^n(CAD and CAD/ETD toggle) or 6×10^4 automated gain control target (ETD), and 2×10^5 reagent target with azulene as the electron transfer reagent. ETD was employed for all PTM site identifications on peptides bearing ≥2 modifiable residues. CAD and ETD date-independent targeting modes were employed for site-mapping or obtaining signature ions for select modified peptides when no data-dependent MS^2 spectra were

acquired. Database analysis was performed using The Open Mass Spectrometry Search Algorithm (OMSSA), version 2.1.1., to search (parameters: ±0.01 Da precursor mass tolerance, ±0.35 Da fragment ion mass tolerance) against the RGA sequence (N-terminally His-FLAG-tagged RGA Uniprot Accession **Q9SLH3**). Either "Trypsin", "AspN", or "No enzyme" were used as protease options, with up to 3 missed cleavages allowed when applicable. Specified variable modifications were carbamidomethylation of Cys, oxidation of Met, O-GlcNAcylation of Ser & Thr, Phosphorylation of Ser, Thr, & Tyr, and O-hexosylation of Ser & Thr residues. Search parameters also included variable modifications of Met to homoserine and homoserine lactone after CNBr cleavage, and acid amidation with histamine on Asp, Glu, and C-terminal residues when applicable. Signals corresponding to charge-reduced species were removed from the spectra by OMSSA before the database was searched. While OMSSA searches were used as a guide, all PTM site identifications were validated by manual interpretation of raw data.

2.3 Results and Discussion

2.3.1 6xHis-3xFlag-tagged RGA Purification

Tobacco plants were used as an expression system to generate highly O-GlcNAcylated RGA protein amenable to MS analysis. Tobacco plants were transfected by Agrobacterium to over-express *Arabidopsis Thaliana* 6x His- and 3x FLAG-tagged protein RGA and *Arabidopsis Thaliana* O-GlcNAc transferase proteins SEC or SPY. These two N-terminal affinity tags allowed for a tandem purification method that firstly took advantage of the affinity of histidine residues for Ni^{2+}-NTA resin [46], then secondly utilized an antibody-antigen interaction using the 22 amino acid 3xFLAG tag, a variation of the original 8 amino acid FLAG sequence [47]. This two-dimensional purification minimized non-RGA protein contamination while yielding pmols of RGA bound to anti-FLAG antibody on agarose beads, amenable for on-beads proteolytic digestion and analysis

by MS. RGA co-expressed with SEC was characterized by a gel shift of 2-3 kDa, implying it was heavily O-GlcNAcylated (**Figure 2.12**). When select poly-Ser/Thr regions were mutated to alanine, the shift was nearly absent as visualized by Western Blot (**Figure 2.13**). RGA co-expressed with SEC was analyzed by mass spectrometry to fully characterize putative RGA(+SEC) O-GlcNAcylation and any other post-translational modification.

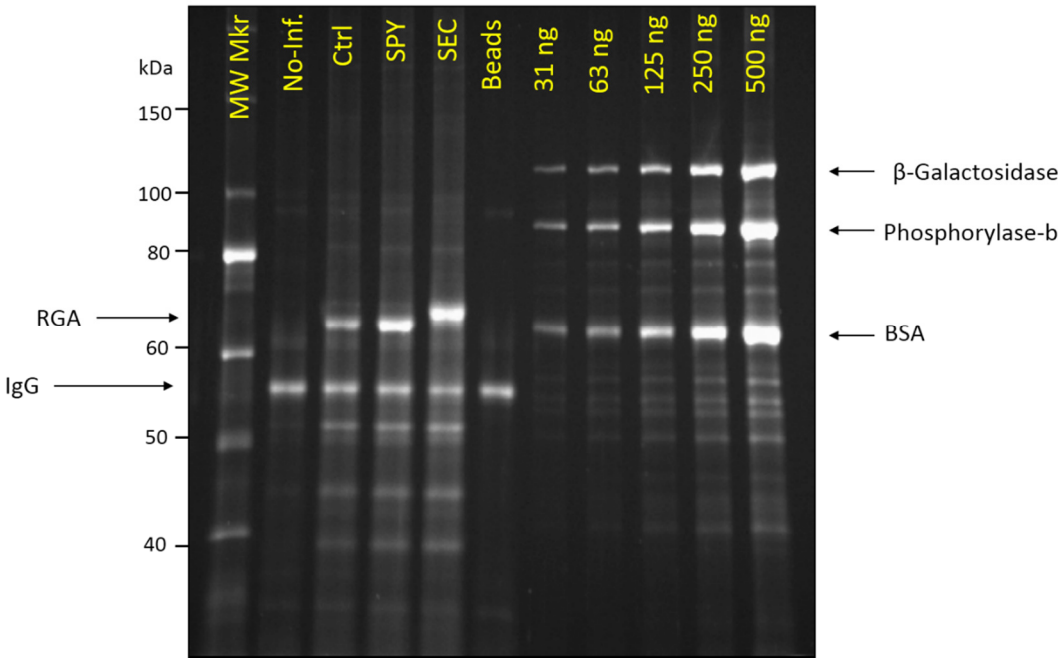
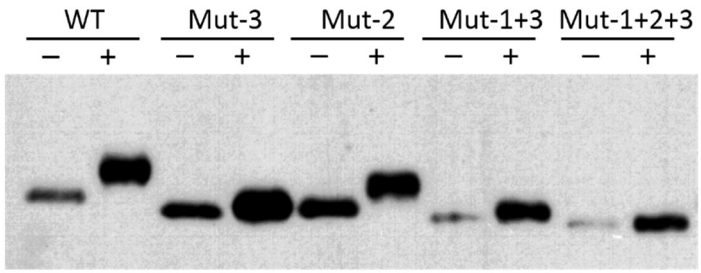


Figure 2.12 – *RGA visualized by Coomassie Brilliant Blue stain on a 6% acrylamide SDS-PAGE gel. MW Mkr – Molecular weight marker, No-Inf. – Non-transfected Tobacco leaf extract tandem purified, Ctrl – RGA expressed without OGT, SPY – RGA coexpressed with SPY, SEC – RGA coexpressed with SEC, Beads – Anti-FLAG agarose beads control, 31-500ng – three standard proteins used for abundance curve.*



Mut 1: LSNHGTSSSSSSISK to
LSNHGAAAAAAAIAK

Mut 2: PAIDSSSSSNNQNKRLKSCSSPDSMVTSTSTGTQ to
PAIDAAAAANNQNKRLKSCAAPDAMVAAAAAGTQ

Mut 3: PASSNGLDP. . . . IGGVIGTTVTTTTTTTTAAGEST to
PAAANGLDP. . . . IGGVIGAAVAAAAAAAAAAGEST

Figure 2.13 – *Western Blot (anti-FLAG-HRP) shows changes in the SDS-PAGE gel shift when potential Ser/Thr-rich regions of RGA are mutated to Ala, eliminating their ability to be O-GlcNAcylated.*

2.3.2 *In Silico* RGA Digestion

Protein mass spectrometrists generally regard trypsin as their protease of choice, because it efficiently cleaves peptides C-terminal to Lys and Arg residues with high fidelity [48]. This produces peptides well-suited for (i) separation by octadecyl carbon chain (C18) reverse-phase chromatography due to their ideal size and charge state upon ESI (typically z = +2 and <15 residues) and (ii) acquiescence for CAD MS/MS whereby protonated amide nitrogens facilitate the collision-activated cleavage of peptide bonds to generate a series containing every possible b and y fragment ion. Furthermore, the presence of one charge originating from the N-terminal amine and another from the C-terminal Lys or Arg residue respectively allows for the detection of each resulting singly-charged (z = +1) b and y ion. In this way, a peptide's amino acid sequence along with any non-labile post-translational modifications can be read forwards and backwards by determining the mass differences between each successive fragment. Still, relying

solely upon this trypsin digest and CAD approach has two major disadvantages: (i) it is uncommon to achieve detection of a protein's entire sequence using one protease, and the predominately doubly-charged (z = +2) peptides typically generated from a tryptic digest are not amenable to ETD, the fragmentation method of choice for site-mapping labile O-linked PTMs like O-GlcNAc [49].

In silico digestions of RGA were performed to determine candidate peptides that would allow for the detection of all Ser/Thr residues and produce quality fragmentation by ETD. Two separate theoretical digests using Endoproteinase AspN and trypsin met this goal by generating peptides of ~10-40 amino acids in length that were predicted to result in the detection of all Ser/Thr residues and largely hold z ≥ +3 (**Tables 2.1 & 2.2**).

Mass	Position	Peptide	# HKR	# ST	Comments
433.2	1-3	(DP)MKR	2	0	?
2483.2	4-26	DHHQFQGRLSNHGTSSSSSS ISK	5	9	+
1267.6	27-36	DKMMMVKKEE	3	0	+
549.2	37-42	DGGGNM	0	0	?
133.0	43-43	D	0	0	-
3953.0	44-78	DELLAVLGYKVRSSEMAEVA LKLEQLETMMSNVQE	3	4	+
812.4	79-86	DGLSHLAT	1	2	+
1722.8	87-100	DTVHYNPSELYSWL	1	3	+
1965.0	101-119	DNMLSELNPPPLPASSNGL	0	3	+
1527.7	120-134	DPVLPSPEICGFPAS	0	2	+
296.1	135-136	DY	0	0	-
1758.0	137-152	DLKVIPGNAIYQFPAI	1	0	+
2024.9	153-171	DSSSSSNNQNKRLKSCSSP	3	8	+
4003.0	172-212	DSMVTSTSTGTQIGGVIGTT VTTTTTTTTAAGESTRSVIL V	1	20	?

7721.0	213-283	DSQENGVRLVHALMACAEAI QQNNLTLAEALVKQIGCLAV SQAGAMRKVATYFAEALARR IYRLSPPQNQI	8	5	-
573.2	284-288	DHCLS	1	1	?
4495.3	289-326	DTLQMHFYETCPYLKFAHFT ANQAILEAFEGKKRVHVI	7	3	+
3934.0	327-363	DFSMNQGLQWPALMQALALR EGGPPTFRLTGIGPPAP	2	3	?
334.1	364-366	DNS	0	1	-
3565.8	367-398	DHLHEVGCKLAQLAEAIHVE FEYRGFVANSLA	5	1	+
246.1	399-400	DL	0	0	-
1117.5	401-410	DASMLELRPS	1	2	+
5750.2	411-463	DTEAVAVNSVFELHKLLGRP GGIEKVLGVVKQIKPVIFTV VEQESNHNGPVFL	7	4	?
1777.8	464-477	DRFTESLHYYSTLF	2	4	+
1044.5	478-487	DSLEGVPNSQ	0	1	+
2395.2	488-509	DKVMSEVYLGKQICNLVACE GP	2	1	+
8579.2	510-587	DRVERHETLSQWGNRFGSSG LAPAHLGSNAFKQASMLLSV FNSGQGYRVEESNGCLMLGW HTRPLITTSAWKLSTAAY	10	15	-

Table 2.1 – In silico AspN digestion of RGA. Residues belonging to the N-terminal 3xFLAG-tag are indicated in parentheses. Peptide masses (Da) and position numbers are listed along with the number of basic (# HKR) and GlcNAc-modifiable (# ST) residues for each theoretical peptide. The comments section refers to peptides predicted to be detected (+) or not detected (-), with (?) indicating uncertainty of detection.

Mass	Position	Peptide	# HKR	# ST	Comments
277.1	1-2	(TDP)MK	1	0	-
174.1	3	R	1	0	-
1023.5	4-11	DHHQFQGR	3	0	?
1477.7	12-26	LSNHGTSSSSSSISK	2	9	+
261.1	27-28	DK	1	0	-
638.3	29-33	MMMVK	1	0	?
146.1	34-34	K	1	0	-
2023.9	35-53	EEDGGGNMDDELLAVLGYK	1	0	+
273.2	54-55	VR	1	0	-
1063.5	56-65	SSEMAEVALK	1	2	+

8140.8	66-139	LEQLETMMSNVQEDGLSHLA TDTVHYNPSELYSWLDNMLS ELNPPPLPASSNGLDPVLPS PEICGFPASDYDLK	3	12	-
2550.2	140-163	VIPGNAIYQFPAIDSSSSSN NQNK	1	5	+
174.1	164-164	R	1	0	-
259.2	165-166	LK	1	0	-
3952.8	167-207	SCSSPDSMVTSTSTGTQIGG VIGTTVTTTTTTTTAAGEST R	1	22	?
1414.7	208-220	SVILVDSQENGVR	1	2	+
2663.4	221-245	LVHALMACAEAIQQNNLTLA EALVK	2	1	+
1403.7	246-259	QIGCLAVSQAGAMR	1	1	+
146.1	260-260	K	1	0	-
1210.6	261-271	VATYFAEALAR	1	1	+
174.1	272-272	R	1	0	-
450.3	273-275	IYR	1	0	-
3320.5	276-303	LSPPQNQIDHCLSDTLQMHF YETCPYLK	3	4	+
1892.9	304-320	FAHFTANQAILEAFEGK	2	1	+
146.1	321-321	K	1	0	-
174.1	322-322	R	1	0	-
2737.4	323-346	VHVIDFSMNQGLQWPALMQA LALR	2	1	+
859.4	347-354	EGGPPTFR	1	1	+
2156.0	355-375	LTGIGPPAPDNSDHLHEVGC K	3	1	+
1787.9	376-390	LAQLAEAIHVEFEYR	2	0	+
3744.9	391-425	GFVANSLADLDASMLELRPS DTEAVAVNSVFELHK	3	5	+
1038.6	426-435	LLGRPGGIEK	2	0	+
613.4	436-441	VLGVVK	1	0	?
2765.5	442-465	QIKPVIFTVVEQESNHNGPV FLDR	3	2	+
2776.3	466-489	FTESLHYYSTLFDSLEGVPN SQDK	2	6	+
1024.5	490-498	VMSEVYLGK	1	1	+
1416.6	499-511	QICNLVACEGPDR	1	0	+
402.2	512-514	VER	1	0	-
1226.6	515-524	HETLSQWGNR	2	2	+
1659.8	525-541	FGSSGLAPAHLGSNAFK	2	3	+
1756.9	542-557	QASMLLSVFNSGQGYR	1	3	+

2728.3	558-581	VEESNGCLMLGWHTRPLITT SAWK	3	5	+
624.3	582-587	LSTAAY	0	2	?

Table 2.2 – In silico trypsin digestion of RGA. Residues belonging to the N-terminal 3xFLAG-tag are indicated in parentheses. Peptide masses (Da) and position numbers are listed along with the number of basic (# HKR) and GlcNAc-modifiable (# ST) residues for each theoretical peptide. The comments section refers to peptides predicted to be detected (+) or not detected (-), with (?) indicating uncertainty of detection.

Each *in silico* digest alone produced a small population of peptides that were either too hydrophilic to bind to the C18 chromatographic column effectively or too hydrophobic to efficiently elute from the C18 during the reverse-phase gradient. However, the combined theoretical sequence coverage guaranteed by two distinct digestions using either AspN or trypsin was expected to potentially yield detection of all ST residues.

2.3.3 Analysis of RGA by LCMS

LC-MS/MS Methodology

Cysteine residues on whole RGA bound to anti-FLAG agarose beads were reduced with dithiothreitol and then irreversibly capped by alkylation using iodoacetamide. Following on-beads digestion, AspN- or tryptic-generated RGA peptides were loaded onto a C18 pre-column (PC), separated on a C18 analytical column (AC) by HPLC, and mass analyzed on an LTQ-FT-ICR or LTQ-Orbitrap mass spectrometer. Main beam or full MS (MS^1) scans of the intact peptides were acquired as they eluted according to their unique hydrophobicity along a 90 min reverse-phase chromatographic gradient. In this way, a powerful two-dimensional separation of the peptide mix was achieved; firstly by chromatographic retention and secondly by mass. In addition, attomole sensitivity could be achieved due to the use of micro-capillary reverse-phase chromatography columns, packed and conditioned in-house and equipped with micro-emitter tips to facilitate nanoflow micro-electrospray. The full MS scans of intact peptides were obtained by the FT-ICR (50,000 FWHM) or Orbitrap (60,000 FWHM) mass analyzers to take

advantage of their superior mass resolution and accuracy, resolving m/z differences of <0.01 Da at 400 m/*z* with mass accuracy errors of <5 parts per million (ppm).

Two internal standard peptides, Angiotensin (Angio) and Vasoactive Intestinal Peptide (Vaso) were spiked into every sample run at 100 fmol prior to loading onto the PC. These internal standards were used to assess peptide loading, chromatographic quality, retention, mass accuracy, CAD and ETD fragmentation efficiency, and relative quantitation. Vaso and Angio were expected to maintain peak widths of ~15-30 seconds (50% peak height) and observed retention times of ~23 and ~32 min, respectively, for a 90 min gradient. The experimental MS^1 masses of intact peptide standard Vaso and Angio, with the theoretical monoisotopic masses of $[M+3H]^{+3}$: 475.8846 m/*z* and $[M+3H]^{+3}$: 432.8998 m/*z* respectively, were generally detected to be accurate to <5 ppm, and the mass accuracy shifts of sample peptides were expected to agree with any mass accuracy shifts in the internal standards within this same <5 ppm window of error.

Mass resolution was calculated using **Equation 2.1** where $W_{½}$ = the full width of the m/*z* peak at half the maximum height (FWHM) and mass accuracy error was calculated in ppm using **Equation 2.2**.

(**Equation 2.1**) $$mass\ resolution = \frac{\left(\frac{m}{z}\right)}{\left(W_{\frac{1}{2}}\right)}$$

By rearranging **Equation 2.1** and solving for $W_{½}$, we find that a species with 400 m/*z* would be resolved from a separate species ≤399.992 m/*z* and ≥ 400.008 m/*z* using a mass resolution of 50,000 FWHM:

$$\left(W_{\frac{1}{2}}\right) = \frac{(400)}{(50{,}000)} = 0.008\frac{m}{\mathrm{z}}$$

(**Equation 2.2**) $ppm\ mass\ error = \frac{\left(\left(experimental\ \frac{m}{z}\right) - \left(theoretical\ \frac{m}{z}\right)\right)}{\left(theoretical\ \frac{m}{z}\right)} \times 10^6$

By rearranging **Equation 2.2** and solving for (experimental m/*z* – theoretical m/*z*), we find that a species with 400 m/*z* would be detected at 399.998-400.002 m/*z* if the mass accuracy error is <5 ppm:

$$\frac{(5)(400)}{10^6} = \left(\left(experimental\ \frac{m}{z}\right) - \left(theoretical\ \frac{m}{z}\right)\right) = 0.002\frac{m}{z}$$

Given these same resolution and mass accuracy parameters in the case of standard peptide Vaso, its $[M+3H]^{+3}$ peak at 475.8846 m/*z* could be resolved from neighboring species ≤475.8751 m/*z* and ≥ 475.8941 m/*z*, and would be detected between 475.8822-475.8870 m/*z*:

$$\left(W_{\frac{1}{2}}\right) = \frac{(475.8846)}{(50{,}000)} = 0.0095\frac{m}{\mathrm{z}}$$

$$\frac{(5)(475.8846)}{10^6} = \left(\left(experimental\ \frac{m}{z}\right) - \left(theoretical\ \frac{m}{z}\right)\right) = 0.0024\frac{m}{z}$$

The ability to resolve fine differences in mass with mDa mass accuracy dramatically increases the ability to confidently identify RGA peptides and post-translational modifications, even when species very close in mass are present. One invaluable aspect of using such high resolution and mass accuracy is that it accesses the ability to distinguish between atomic isotopes. These isotopes differ by the mass of one or more neutrons (1.0087 Da/neutron) minus a small mass defect caused by the amount of nuclear binding energy needed for the neutron to form part of the nucleus of the element [50]. As these mass defects differ for each isotope of

every element, Carbon-12 has been defined as the element having no mass defect and a monoisotopic mass of 12.0000 Da. The two elements present in all proteins that most contribute to multiple peptidic isotopes are carbon and nitrogen. Carbon's most abundant natural isotope, carbon-12, is present at 98.9% abundance while 1.1% is present as carbon-13, with a mass of 13.0034 Da. The most abundant natural isotope of nitrogen, nitrogen-14 with a mass of 14.0031, is present at 99.6%, with its lesser abundant isotope, nitrogen-15 at mass 15.0001, present at 0.37%. On average, peptides contain five carbons per residue but only one nitrogen. Carbon-13 has a greater effect than nitrogen-15 on the isotopic distribution of peptides, because it is a more abundant isotope and because of the disparity between the number of carbon and nitrogen atoms. For this reason, reasonable first-order approximations of a peptide's isotopic distribution may be made by only considering carbon-13 isotopes. The probability of a monoisotopic peptide, having only carbon-12 atoms (P_M), is given by **Equation 2.3**, where c is the abundance of carbon-13 and w is the number of total carbon items in the peptide [51]:

(**Equation 2.3**) $P_M = (\frac{100-c}{100})^w$

Likewise, the probability of a peptide containing one carbon-13 atom (P_{M+1}) is given by **Equation 2.4**:

(**Equation 2.4**) $P_{M+1} = w(\frac{c}{100-c})(\frac{100-c}{100})^w$

For Vaso, bearing 61 carbons, the probability that one of its molecules contains only carbon-12 is 51%:

$$P_M = \left(\frac{100 - 1.1}{100}\right)^{61} = \left(\frac{98.9}{100}\right)^{61} = 51\%$$

The probability that Vaso contains exactly one carbon-13 atom is therefore 34%:

$$P_{M+1} = 61\left(\frac{1.1}{100-1.1}\right)\left(\frac{100-1.1}{100}\right)^{61} = (61)(0.011)(.989)^{61} = 34\%$$

Alternatively, the probability of isotopic peaks caused by carbon-13 may be determined by binomial probability [52]. The probability P_k of a peptide containing exactly k carbon-13 atoms with n total carbons atoms may be calculated from **Equation 2.5**:

(**Equation 2.5**) $P_k = \binom{n}{k}(p^k)(1-p)^{n-k}$

Therefore, the probabilities of Vaso containing exactly zero, one, and two carbon-13 atoms are calculated as 51%, 34%, and 12% respectively:

$$P_k = (0.011^0)(1-0.011)^{61-0} = 51\%$$

$$P_k = \frac{61!}{1!\,(61-1)!}(0.011^1)(1-0.011)^{61-1} = 34\%$$

$$P_k = \frac{61!}{2!\,(61-2)!}(0.011^2)(1-0.011)^{61-2} = 12\%$$

These probabilities translate to the amount of total Vaso ion current that will be represented in each isotopic state. One half of the ions will represent the first isotopic peak pertaining to carbon 12 only, a third will represent the second peak, and the remaining third of the ion current will contain the remaining peaks, which contain two or more carbon-13 atoms (i.e. $2^{13}C$, $3^{13}C$, etc.). The relative abundances of each carbon-13 peak vs. carbon-12 only are easily determined by using the products of **Equation 2.5** to solve **Equation 2.6** where the peptide contains k carbon-13 atoms:

(**Equation 2.6**) $Relative\ abundance\ of\ k^{13}C = \frac{P_k}{P_0}$

Using probabilities P_0, P_1, P_2, find that ions containing one and two carbon-13 atoms are present at 67% and 24% respectively, relative to those containing only carbon-12:

$$Relative\ abundance\ of\ 1^{13}C = \frac{34\%}{51\%} = 67\%$$

$$Relative\ abundance\ of\ 2^{13}C = \frac{12\%}{51\%} = 24\%$$

When the identity of a peptide is unknown, an estimation of the number of residues, and therefore the number of carbons may be made. The average molecular weight of all 20 amino acids is 110 Da, so an approximation of the total number of carbons can be made using **Equation 2.7**:

(Equation 2.7) $\#\ Carbon\ atoms = (\frac{Total\ Molecular\ weight\ Da}{110\ Da})(5)$

Applying this calculation to Vaso yields

$$\#\ Carbon\ atoms = \left(\frac{1425}{110}\right)(5) = 65$$

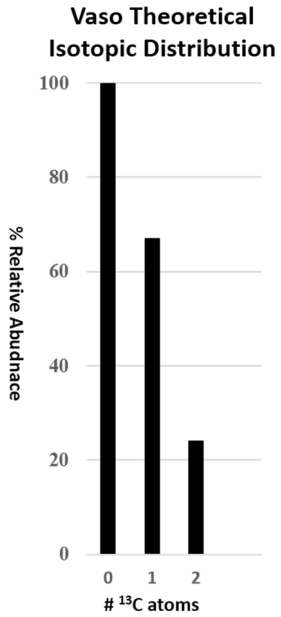


Figure 2.14 – *Theoretical distribution of the first three isotopic peaks of Vasoactive Intestinal Peptide fragment, HSDAVFTDNYTR. The distribution was computed binomially using equations 2.5 and 2.6.*

which is close to the actual number of 61 carbons. The theoretical isotopic distribution of $[M+3H]^{+3}$ Vaso, computed binomially from **Equations 2.5 and 2.6**, is shown in (**Figure 2.14**).

By comparison, Vaso's experimental isotopic distribution (**Figure 2.15**) matches well with the

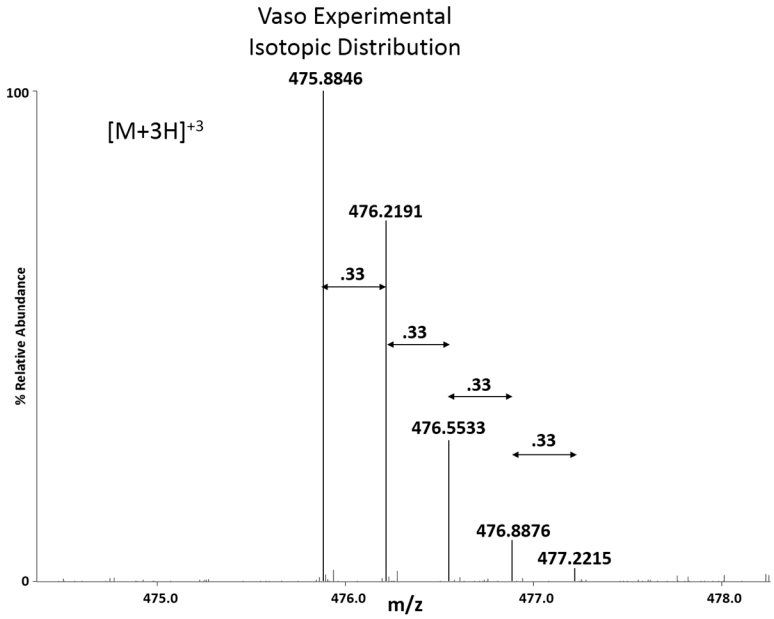


Figure 2.15 – *Isotopic distribution of Vasoactive Intestinal Peptide fragment, HSDAVFTDNYTR, experimentally determined using LTQ-FT at 50,000 FWHM resolution.*

statistical approximation. It is important to note that the isotopic distribution of Vaso and all

peptides will be completely independent of charge state because the only difference in primary molecular structure is the degree of protonation. However, the ability to resolve isotopes has a critical implication for determining charge states by mass spectrometry. Because each successive isotopic peak represents a molecule separated by ~1 Da and the mass spectrometer measures m/*z*, the pattern of separation between isotopic peaks within a molecule is indicative of its charge state. One may simply solve for charge state by **Equation 2.8** when the difference in m/*z* between isotopic peaks (Δ isotopes) is known:

(Equation 2.8) $z = \left(\frac{1}{\Delta\frac{m}{z}\, between\ isotopes}\right)$

For the most abundant charge state of Vaso (*z*=3), the distance between isotopes is 0.33 m/*z* so this equates to:

$$Vaso\ z = \left(\frac{1}{0.33}\right) = 3$$

Due to the high resolving power of mass analyzers like the FT-ICR and Orbitrap, the charge states of peptides from RGA samples could be determined, allowing not only a highly accurate m/*z* measurement, but that of mass itself. In addition, isotopic distributions were used as an analytical tool, giving an additional piece of identifying evidence for alleged peptides. Likewise, analysis of isotopic distribution helped determine if species similar in mass were potentially related as these distributions should be virtually identical barring the addition of atoms with highly abundant isotopes, uncommon to most peptides, or low abundance/poor detection of ions that skews the statistics of isotopes detected. These observations were made both qualitatively and quantitatively with the aid of the PNNL Isotope Pattern Calculator, created by Michael Cusack. It is important to note that one distinct disadvantage encountered in resolving isotopes is that a peptide's ion current, detected as one distinct peak under low resolution, is

split amongst multiple signals in high-resolution, lowering the signal-to-noise ratio by a certain factor. As previously calculated, 51% of Vaso's ion current is due to its ^{12}C peak. The abundance of the ^{12}C peak will decrease as peptide mass increases and the overall number of isotopic peaks detected will increase as they become more abundant. This, combined with the tendency for peptides to also gain a larger charge state distribution as they increase in mass, can drastically divide ion current. This is one of the greatest analytical challenges associated with analyzing large peptides and whole proteins, techniques termed middle- and top-down, respectively.

RGA peptides were quantified relative to the levels of standard peptides Angio and Vaso detected. Approximately 100 fmols of each peptide standard were spiked into every sample before they were loaded onto the pre-column. The monoisotopic peaks corresponding to each charge state of each standard were manually observed by isolating them in an extracted ion chromatogram (XIC) with a 0.01 m/*z* window centered on the theoretical m/*z* value). **Figure 2.16** demonstrates standards from a representative run of trypsin digested RGA. The ion current corresponding to 100 fmol was calculated by integrating the total C12 isotope area of each standard and averaging the two using **Equation 2.9**:

(Equation 2.9) $$100fmol\ ion\ current = \left(\frac{\sum Area_{100fmol\ Angio} + \sum Area_{100fmol\ Vaso}}{2}\right)$$

Using the same methodology and the data from **Equation 2.9**, **Equation 2.10** was used to calculate the abundance of RGA peptides (including all PTM forms) and other species in a single analysis:

(Equation 2.10) $$Peptide\ abundance\ (fmol) = \frac{100(\sum Area_{sample\ peptide})}{(100\ fmol\ ion\ current)}$$

For the tryptic RGA peptide run shown in **Figure 2.16**, 100 fmol equated to:

$$100fmol\ ion\ current = \left(\frac{6.09E8\ +\ 6.72E8}{2}\right) =\ 6.41\text{E}8$$

and the abundance of the unmodified tryptic RGA peptide, FGSSGLAPAHLGSNAFK, was calculated to be:

$$Peptide\ abundance\ (fmol) = \frac{100(4.75E9)}{(6.41E8)} = 741\ fmol$$

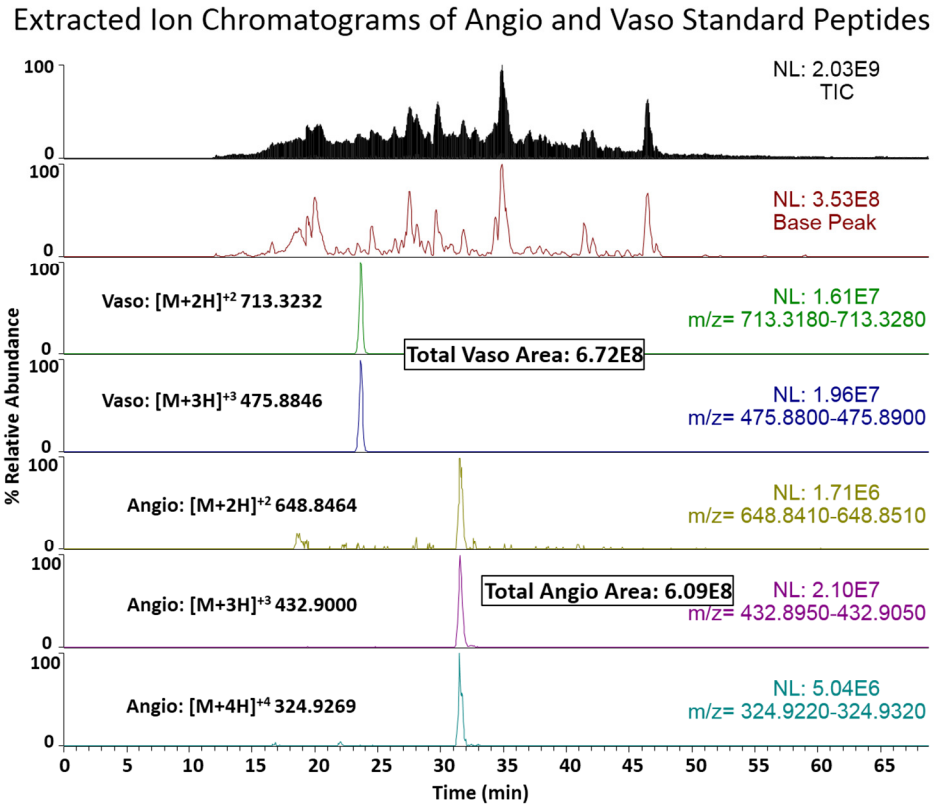


Figure 2.16 – *XICs of internal peptide standards Vasoactive Intestinal Peptide (Vaso) and Angiotensin (Angio) used to calculate relative quantitation of RGA peptides.*

While variations in isotopic distribution between peptides exist due to different sizes and sequences, the effect on abundance calculations was negligible considering the population of RGA peptides analyzed. Most peptides ranged from 10-25 residues in length, with the percentage of monoisotopic peaks in the overall isotopic distribution differing by a factor ≤ 2. Therefore, comparison of the monoisotopic abundances of known amounts of standards relative

to those of sample peptides offered a robust method to calculate sample peptide abundances. Still, variations in the levels of peptides detected varied as is expected due to factors such as digestion efficiency and homogeneity, complete cysteine reduction and alkylation, peptide solubility, binding and elution efficiency on C18, variable sample complexity throughout chromatographic gradient, and ionization efficiency by ESI. In order to estimate the approximate amount of RGA present in each digest, the abundances of the top five most abundant RGA peptides were summed and averaged **(Equation 2.11)**:

$$RGA\ abundance\ (fmol) = \overline{x}_{top\ 5\ abundant\ RGA\ peptides}$$

In order to control for the variability in peptide levels detected, the percent total abundance of post-translationally modified peptides was quantified relative to the total abundance of all forms of those peptides detected (**Equation 2.12**), rather than against the estimated protein abundance as a whole.

(Equation 2.12) $$\%\ Abundance\ Modified\ Peptide = \frac{\sum Area_{Modified\ peptide}}{\sum Area_{All\ peptide\ forms}} \times 100$$

Detection of low-level PTMs

High sample purity, intelligent instrumentation methods, and care interpreting data are all critical for the successful detection of O-GlcNAcylated peptides, as they are usually detected at low levels (<10%) relative to their unmodified counterparts. The dynamic range (>1000) and sensitivity (attomole detection) of the high-resolution trap-based mass analyzers utilized allows for the detection of PTMs at ≤1%, but this sensitivity will be lost if a significant amount of other ions, including unmodified forms of the desired peptide to be analyzed, fill the trap's capacity. This prevents one from simply analyzing a larger amount of material to improve O-GlcNAc detection. For MS^2 analysis, the LTQ is filled exclusively with ions that display the specific m/*z*

value to be fragmented, bypassing this issue unless an abundant contaminating ion has a mass-to-charge ratio within the isolation window (typically 3 m/*z* in total). Still, low PTM abundance presents a challenge for obtaining quality fragmentation and MS^2 spectra for two reasons. Firstly, the mass spectrometer is generally operated in the data-dependent mode whereby ions detected in MS^1 are automatically chosen for MS^2 analysis based on parameters described in Methods to obtain MS^2 data for as many analytes as possible. Low level species like most O-GlcNAcylated peptides may still be too low to be chosen over higher abundant species co-eluting with them. For individual cases where the m/*z* value of a particular PTM peptide is known, this may be circumvented by operating in data-independent mode in which ions that fall within the targeted m/*z* window will continually undergo MS^2 regardless of its abundance in the preceding MS^1 spectra. Secondly, O-GlcNAc's low abundance is often not sufficient to fill the ion trap to its optimal capacity for conducting MS^2, resulting in poor spectral quality that prevents confident sequencing and PTM site-localization. In addition to these challenges, some O-GlcNAcylated peptides appear to suffer from "ion suppression" whereby the O-GlcNAcylated peptide, which has only a slightly lower retention by reverse-phase chromatography than its unmodified counterpart, fails to ionize as efficiently (**Figure 2.17**).

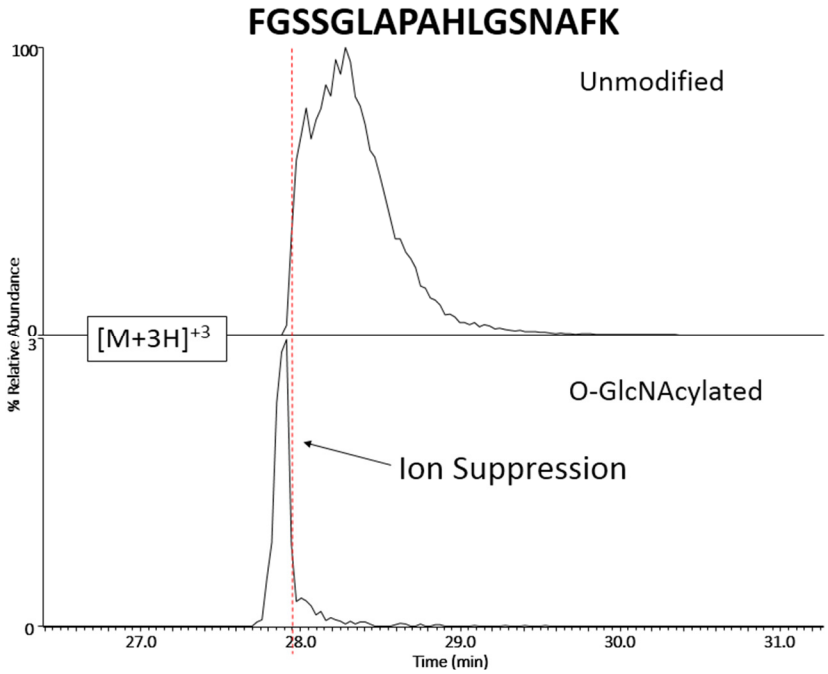


Figure 2.17 – *XICs of unmodified and mono-O-GlcNAcylated forms of peptide FGSSGLAPAHLGSNAFK co-eluting. The O-GlcNAcylated peptide experiences ion suppression relative to the unmodified form, presumably because the former is more buried inside the hydrophilic interior of the electrospray droplet.*

One hypothesis for this phenomenon fits well with the ion desorption model of electrospray ionization [53] postulating that analyte ions are created when the surface electric fields of the charged electrospray droplets become great enough to allow an analyte ion to escape the energy barrier holding it inside the droplet. Under this model, it is conceivable that O-GlcNAc ion suppression could occur, because the slightly more hydrophilic O-GlcNAcylated peptides may favor the aqueous interior of the ESI droplet over the more hydrophobic liquid-gas interphase, where desorption and ionization occurs. Researchers have developed various O-GlcNAc enrichment methods to overcome these limitations. These include immunoprecipitation by the pan-specific O-GlcNAc antibody, CTD110.6, raised against the carboxyl terminal domain of O-GlcNAcylated RNA polymerase II [54, 55], lectin weak affinity chromatography [56], and chemoenzymatic tagging to enrich O-GlcNAc using biotin-avidin affinity [31, 54, 55]. While all of these methods have shown some success using large amounts of whole cell or nuclear lysates,

none have demonstrated the ability to enrich O-GlcNAcylation with high recovery or reproducibility. While efforts are continually being made to development an enrichment method that will efficiently eliminate non-GlcNAcylated material while recovering O-GlcNAcylated peptides in high yield, none yet exists that is appropriate for characterizing GlcNAcylation across a whole protein like RGA. To maximize the probability of successfully detecting O-GlcNAcylation, we relied on the combination of SEC overexpression, immunoprecipitation efficiency, appropriate sample preparation, sensitive LCMS instrumentation, intelligent MS methodology, and thoughtful, knowledgeable data analysis.

Sequencing Peptides using MS/MS analysis

Because ESI is a soft ionization technique, ionized peptides are left intact and may be analyzed accordingly to generate the numerous data already discussed. Still, the absence of fragmentation yields little information to investigate the primary structure and sequence of these peptides. MS/MS analysis enabled RGA peptide sequencing by fragmenting the peptide ions to generate product fragment ions, and mass analyzing those ions to generate sequence ladder data. The resulting spectra were interpreted to determine peptide primary structure and locate PTM sites. CAD and ETD MS/MS were both utilized and conducted in the LTQ in order to take advantage of its speed and sensitivity. Approximately 3-5 MS/MS spectra at 1 m/*z* unit resolution (400 FWHM) and ± 0.2 m/z mass accuracy could be acquired at ~10-fold higher sensitivity in the same time frame (~1 sec) as 1 MS^1 scan in the FT-ICR or Orbitrap. Open Mass Spectrometry Search Algorithm (OMSSA) software aided in data analysis to search for O-GlcNAcylation and phosphorylation PTMs, as well as non-specifically immunopurified tobacco proteins. Still, theoretical monoisotopic masses for all potentially modified RGA peptides were

manually investigated and all RGA MS/MS spectra manually validated for peptide sequencing and PTM site-mapping.

MS/MS Sequencing by CAD

Fragmenting peptides by CAD has enabled mass spectrometrists to unambiguously identify peptide sequences, and locate certain PTM sites. However, the presence of labile O-linked PTMs like O-GlcNAc and phosphorylation on a peptide severely decrease the amount of random peptide bond cleavages, therefore suppressing the number and abundance of b and y ions available to facilitate sequencing. This is due to the presence of significantly more labile O-linkages (relative to peptide bonds) that upon CAD, preferentially result in the neutral loss of phosphoric acid (phosphorylation) or charge-reduced loss of an oxonium ion (O-GlcNAcylation) [54] from modified peptides. This phenomenon can be exploited as an analytical tool, enabling one to search for these diagnostic ions to identify modified peptides. While phosphorylation can sometimes be site-mapped using CAD due to incomplete fragmentation of the labile Ser/Thr-phosphate linkage, it is typically impossible to determine sites of O-GlcNAcylation by CAD fragmentation. O-GlcNAcylated peptides commonly fragment completely and exclusively at the Ser/Thr-O-GlcNAc linkage and prevent any amide bond fragmentation. The difficulty in site localization of O-GlcNAcylated peptides is compounded by the fact that O-GlcNAcylation often exists in Ser/Thr-rich domains of proteins [7], causing further ambiguity in locating the modified site as multiple putative locations exist in high density. RGA contains four such regions (**Figure 2.18**), and it is imperative to generate a nearly perfect series of fragment ions to identify PTMs in them.

Site-mapping O-GlcNAcylation by ETD

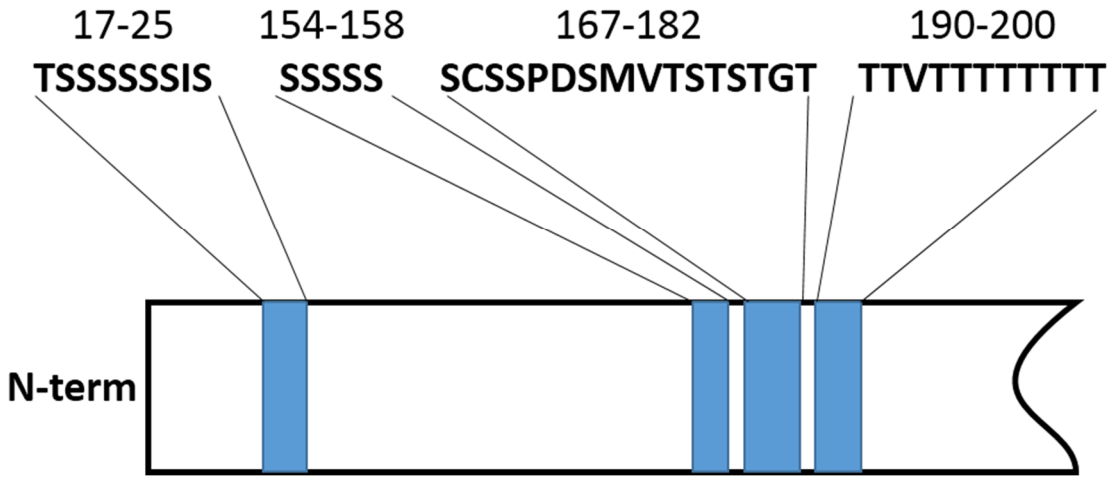
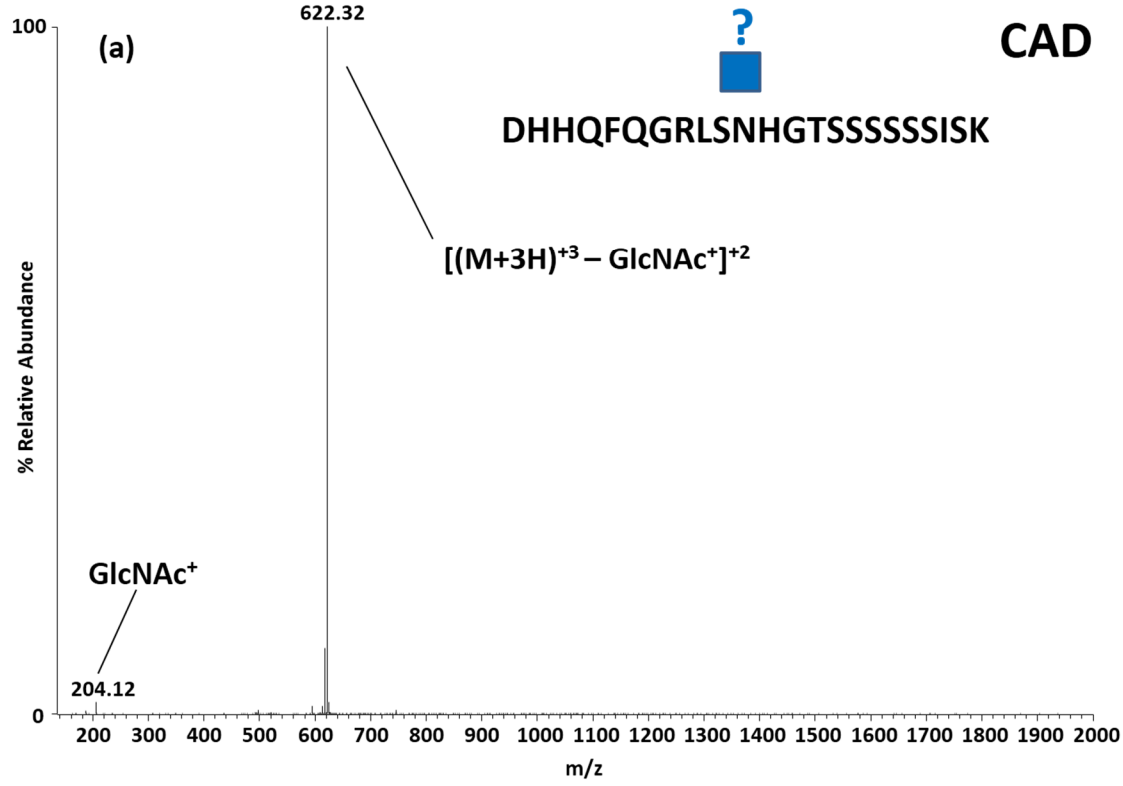
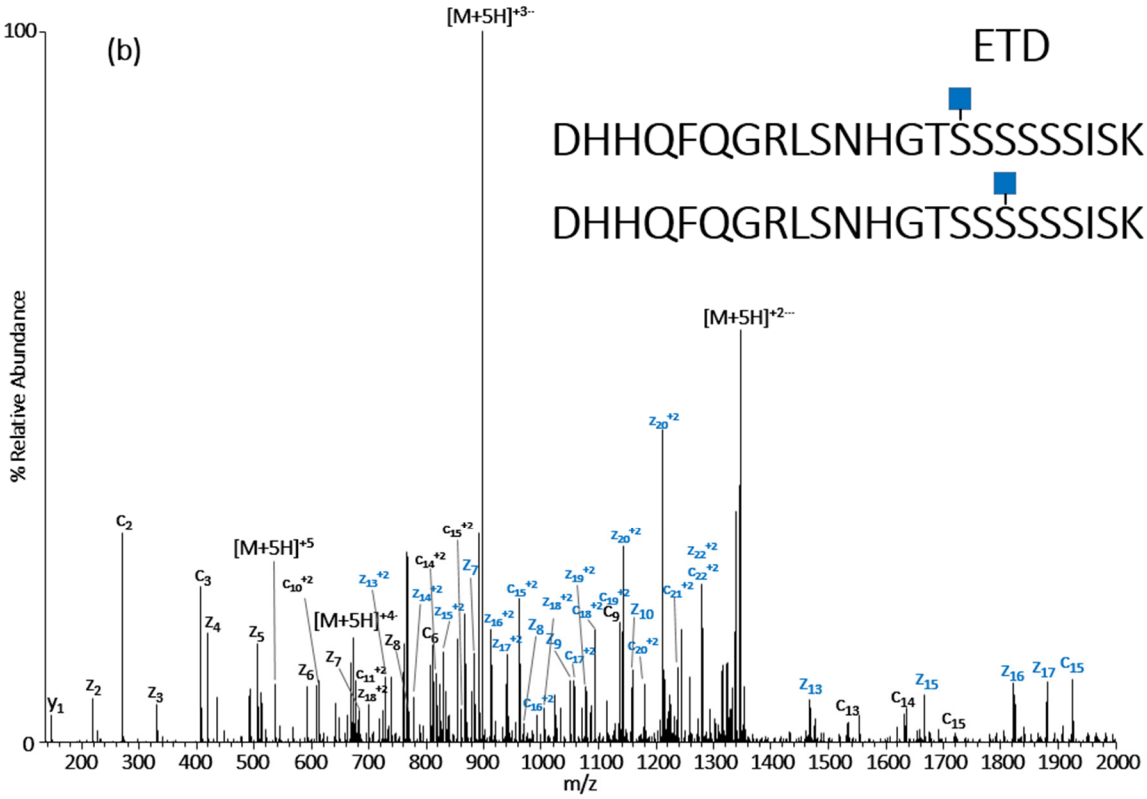


Figure 2.18 – *Schematic of 4 Ser/Thr-rich domains in RGA. All 4 are located in the N-terminal half of RGA.*

Fortunately, ETD offers an alternative MS/MS method with the ability to fragment peptides along the backbone and retain O-GlcNAcylation on the modified residue by utilizing a fragmentation pathway based on free-radical chemistry rather than collisions with helium atoms. The O-GlcNAc site(s) can be localized using ETD while CAD produces only diagnostic fragment ions (**Figure 2.19**).


(a)
CAD
?
DHHQFQGRLSNHGTSSSSSSISK
622.32
[(M+3H)+3 – GlcNAc+]+2
GlcNAc+
204.12
% Relative Abundance
m/z


(b)
ETD
DHHQFQGRLSNHGTSSSSSSISK
DHHQFQGRLSNHGTSSSSSSISK
[M+5H]+3
[M+5H]+2
[M+5H]+5
[M+5H]+4
% Relative Abundance
m/z

(c)

GlcNAc		Unmodified					Unmodified			GlcNAc		
+2 c ion	+1 c ion	+2 c ion	+1 c ion	c		z	+1 z ion	+2 z ion	+3 z ion	+1 z ion	+2 z ion	+3 z ion
		67	133	1	D	23	2484	1243	829	2687	1345	897
		136	270	2	H	22	2353	1178	786	2556	1279	853
		204	407	3	H	21	2216	1109	740	2419	1211	808
		268	535	4	Q	20	2079	1041	694	2282	1142	762
		342	682	5	F	19	1951	977	651	2154	1078	719
		406	810	6	Q	18	1804	903	602	2007	1005	670
		434	867	7	G	17	1676	839	560	1879	940	627
		513	1023	8	R	16	1619	810	541	1822	912	608
		569	1137	9	L	15	1463	732	489	1666	834	556
714	1427	613	1224	10	S	14	1350	676	451	1553	777	519
771	1541	670	1338	11	N	13	1263	632	422	1466	734	490
840	1678	738	1475	12	H	12	1149	575	384	1352	677	451
868	1735	767	1532	13	G	11	1011	507	338	1215	608	406
919	1836	817	1633	14	T	10	954	478	319	1158	580	387
963	1923	861	1720	15	S	9	853	427	285	1056	529	353
1006	2010	904	1807	16	S	8	766	384	256	969	486	324
1050	2097	948	1894	17	S	7	679	340	227	882	442	295
1093	2184	992	1981	18	S	6	592	297	198	795	398	266
1137	2271	1035	2068	19	S	5	505	253	169	708	355	237
1180	2358	1079	2155	20	S	4	418	210	140	621	311	208
1237	2471	1135	2268	21	I	3	331	166	111	534	268	179
1280	2558	1179	2355	22	S	2	218	110	73	421	211	141
1345	2687	1243	2484	23	K	1	131	66	44			

Figure 2.19 – *CAD (a) and ETD (b) spectra of mono-O-GlcNAcylated peptide DHHQFQGRLSNHGTSSSSSISK. Signature fragmentation of the GlcNAc moiety indicates the peptide is O-GlcNAcylated, but no sequence data was generated. The c and z· fragment ions produced by ETD (b & c) indicated that two mono-O-GlcNAcylated forms co-eluted and were fragmented together. One form contained O-GlcNAc on S18 and another on S20, highlighted in blue in (c). Unmodified fragment ions are indicated in gray/black and O-GlcNAcylated in blue (b & c). Blue square indicates site of O-GlcNAcylation.*

Another advantage of ETD is that it can effectively fragment peptides of larger size if they contain sufficient charge, whereas CAD fragmentation becomes less efficient as peptide length increases and vibrational energy diffuses across a larger mass. The ETD fragmentation mechanism requires that the carbonyl radical anion (generated by electron transfer from radical anion reagent) abstract a proton from a nearby protonated nitrogen, causing the radical site located on the resulting charge-reduced species to trigger fragmentation of the N-Cα bond resulting in ions of type c and z· [57]. Because of this, the most critical requirements for candidate peptides are that they contain z≥+3 with the charge ideally spread evenly across the peptide. This allows one charge to remain on each c and z· ion after charge-reduction, and ensures that protons will be available to facilitate fragmentation along the entire peptide backbone. Increasing the charge also has the added benefit of increasing the ETD reaction rate as the rate

of electron transfer from the reagent radical anion to the cationic peptide increases quadratically with the peptide precursor charge state [58] (e.g. a peptide of z = +30 will react 100x faster than that of z = +3). Therefore, ETD is not only superior to CAD for purposes of site-mapping labile PTMs like O-GlcNAcylation, it is also superior in that it can be used on longer (>15 residues), more highly-charged (z≥+3) peptides, giving a greater understanding of what PTMs exist simultaneously or individually on the protein as a whole.

2.3.4 His-Flag-RGA (+SEC) Sequence Coverage

Figures 2.20A & 2.20B respectively demonstrate representative base peak chromatograms of AspN and tryptic RGA digest screens, reporting the most abundant ion detected in each MS1 scan. Screens of 500-1000 fmol RGA were extrapolated to determine that 6xHis-3xFLAG RGA(+SEC) samples each contained 10-50 pmols of RGA.

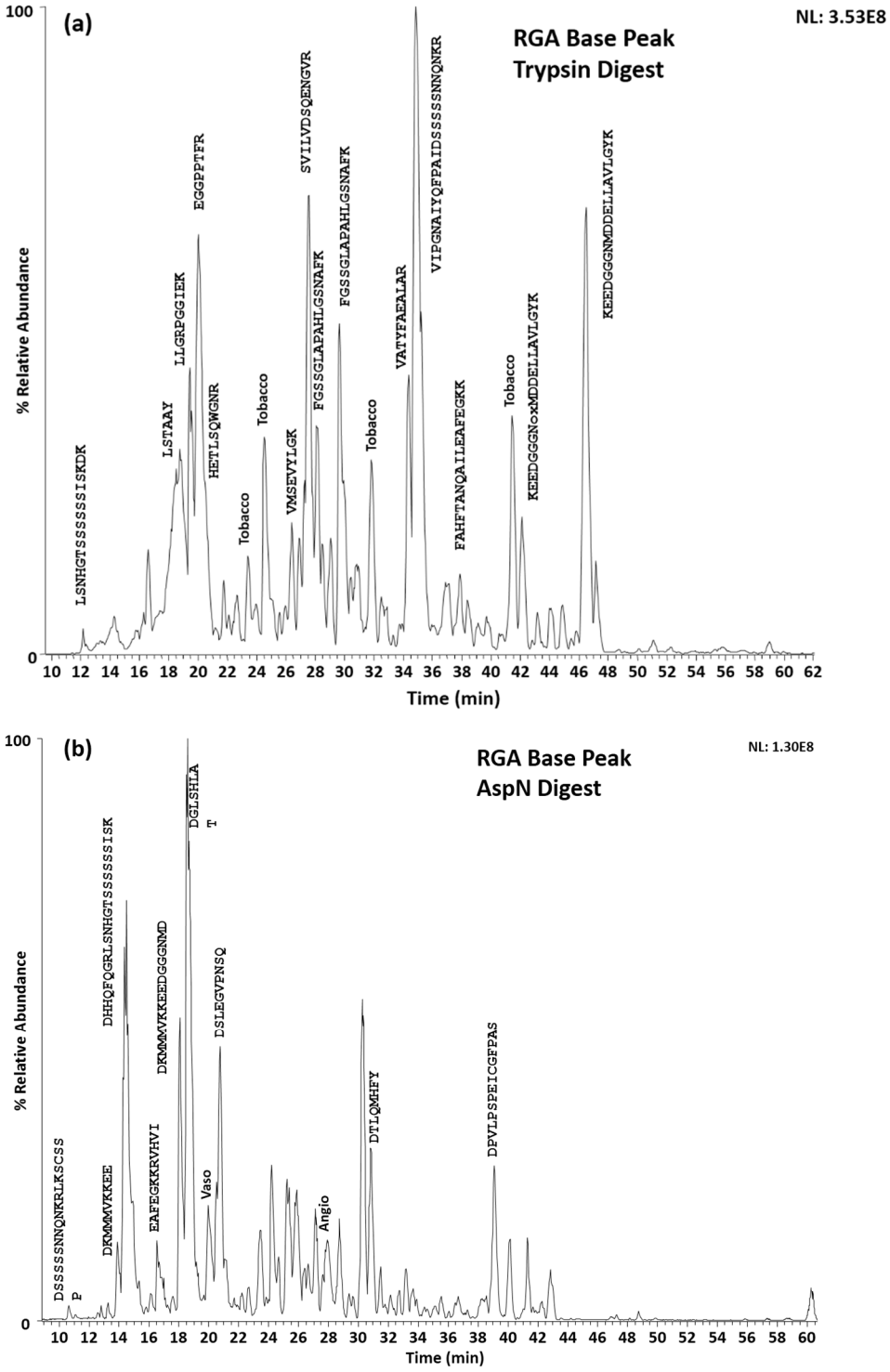


Figure 2.20 –*Base Peak Chromatograms of representative trypsin (a) and AspN (b) RGA digests.*

Notably, the tandem purification of RGA, utilizing the 6x His- and 3x FLAG-tags, resulted in relatively pure RGA samples in all preps, indicated by >90% of the base peak chromatogram belonging to RGA peptide signal. While neither digest could exclusively achieve such a high level of coverage, the combination of peptides detected between trypsin and AspN digests yielded 91% sequence coverage of RGA (**Figure 2.21**)

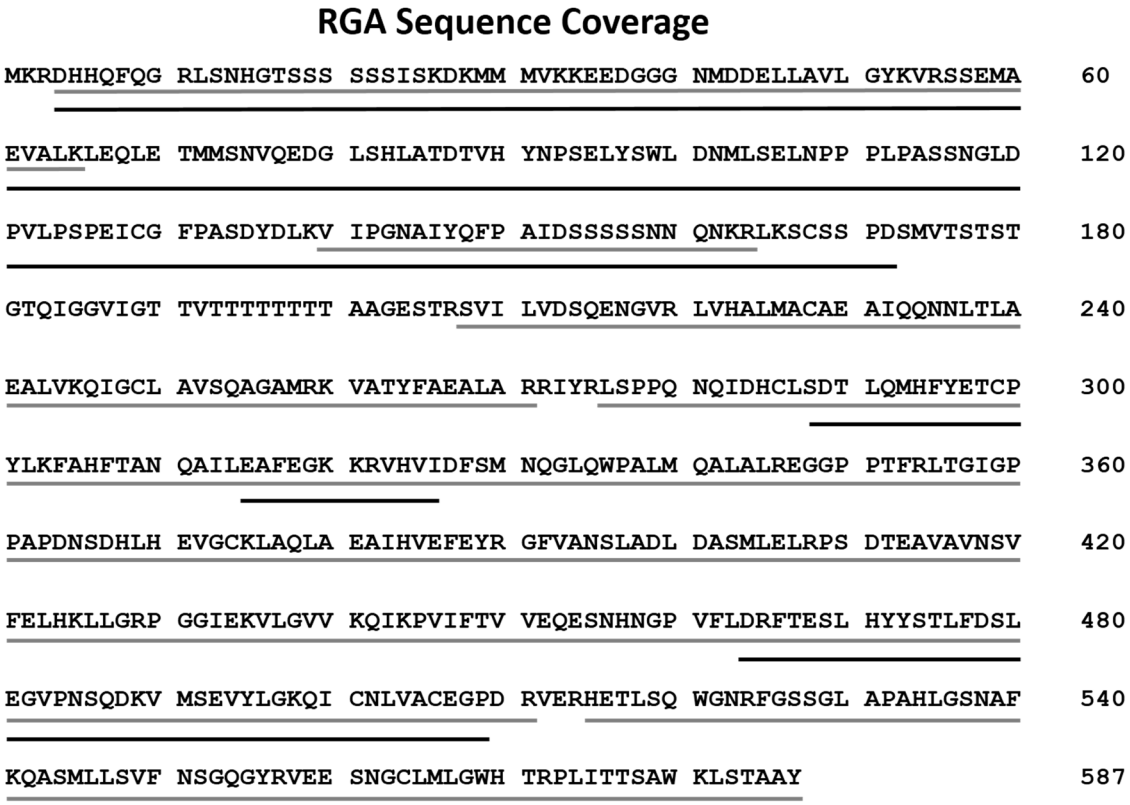


Figure 2.21 – *RGA Sequence coverage from trypsin (gray) and AspN (black) digests totaled 91% including 74% of Ser and Thr residues.*

Still, only 74% of all Ser and Thr residues were detected. This was because the tryptic peptide, SCSSPDSMVTSTSTGTQIGGVIGTTVTTTTTTTTAAGESTR, and the AspN peptide, DSMVTSTSTGTQIGGVIGTTVTTTTTTTTAAGESTRSVILV, contain two Ser/Thr-rich domains. Neither peptide covering this poly-Ser/Thr region was detected. In order to probe the efficiency of digestion in this region, the flanking peptides in the tryptic digest of RGA were examined. The

N-terminally and C-terminally flanking peptides, VIPGNAIYQFPAIDSSSSSNNQNKR and SVILVDSQENGVR respectively, were both detected at ~100% level relative to the most abundant RGA peptides detected as a whole, demonstrating that trypsin targeted and cleaved this region successively and the entire population of these peptides were accounted for and identified. Detection of this region remained a high priority as it contained 19 Ser/Thr residues and was central to the hypothesis that RGA co-expressed with SEC is heavily O-GlcNAcylated.

Detection of the S/T-rich domain by alternative digests

To further examine the potential success of alternative methods to detect the missing Ser/Thr-rich domain, recombinant 6x His- and 3x FLAG-tagged RGA expressed in *E. coli* was used. RGA bound to anti-FLAG agarose beads was reduced and alkylated with DTT and IAA as described previously, but with the addition of 4 M urea to the buffer to encourage further RGA denaturation and increase protease accessibility. RGA was cleaved C-terminally to its lysine residues by endoproteinase LysC, an enzyme that retains activity in high concentrations of urea. The resulting peptides were extracted, diluted to 0.5 M urea (amenable to trypsin), and sub-digested with trypsin to generate the peptide sequence, SCSSPDSMVTSTSTGTQIGGVIGTTVTTTTTTTTAAGESTR. The digest as a whole remained similar in efficiency to that from the tobacco sample. Still, SCSSPDSMVTSTSTGTQIGGVIGTTVTTTTTTTTAAGESTR was now detected in the analysis of this digest, but only at <5% abundance relative to the RGA peptides as a whole and with a very poor elution profile, proving that the peptide was surprisingly retentive despite a heavy concentration of hydroxyl groups and polar residues, tailing for longer than 20 min under high organic phase. If modified forms of this peptide present at <10% were to elute in this manner, they would fall below the detectable dynamic range of the mass spectrometer. Moreover, the

same peptide was again undetectable when applying this method to the RGA (+SEC) sample expressed in tobacco, providing further evidence that protease accessibility and efficiency was sufficient, so urea was not necessary for digestion. However, this observation was consistent with the hypothesis that the Ser/Thr-rich region could be heavily O-GlcNAcylated, because if the peptide existed in several different modified forms (i.e. different masses), the ion current for each species would be reduced relative to the solely unmodified form present in the recombinant sample, and this could cause all RGA (+SEC) forms to fall below the limit of detection. An AspN digestion, conducted in the absence of urea, yielded DSMVTSTSTGTQIGGVIGTTVTTTTTTTTAAGESTRSVILV as well, but it suffered from the same poor elution and abundance as the peptide generated in the same region by trypsin. This peptide was further sub-digested with endoproteinase Glu-C, cleaving C-terminally to glutamic acid, with the idea that the shorter DSMVTSTSTGTQIGGVIGTTVTTTTTTTTAAGE peptide may exhibit more hydrophilicity and adopt a sharper elution profile. Unfortunately, the peptide was no longer detected. One may reason that in actuality, the absence of the positively-charged Arg had the opposite of the intended chromatographic effect. Accordingly, tryptic RGA peptides were sub-digested with AspN to generate the similar DSMVTSTSTGTQIGGVIGTTVTTTTTTTTAAGESTR peptide still retaining the basic Arg on the C-terminus. The peptide was once again detected but with no improvement on elution profile or sensitivity. (**Figures 2.22A and 2.22B**) give representative XICs of the Ser/Thr-rich domain when using alternative chromatography stationary phases, digests, and chemical derivatization.

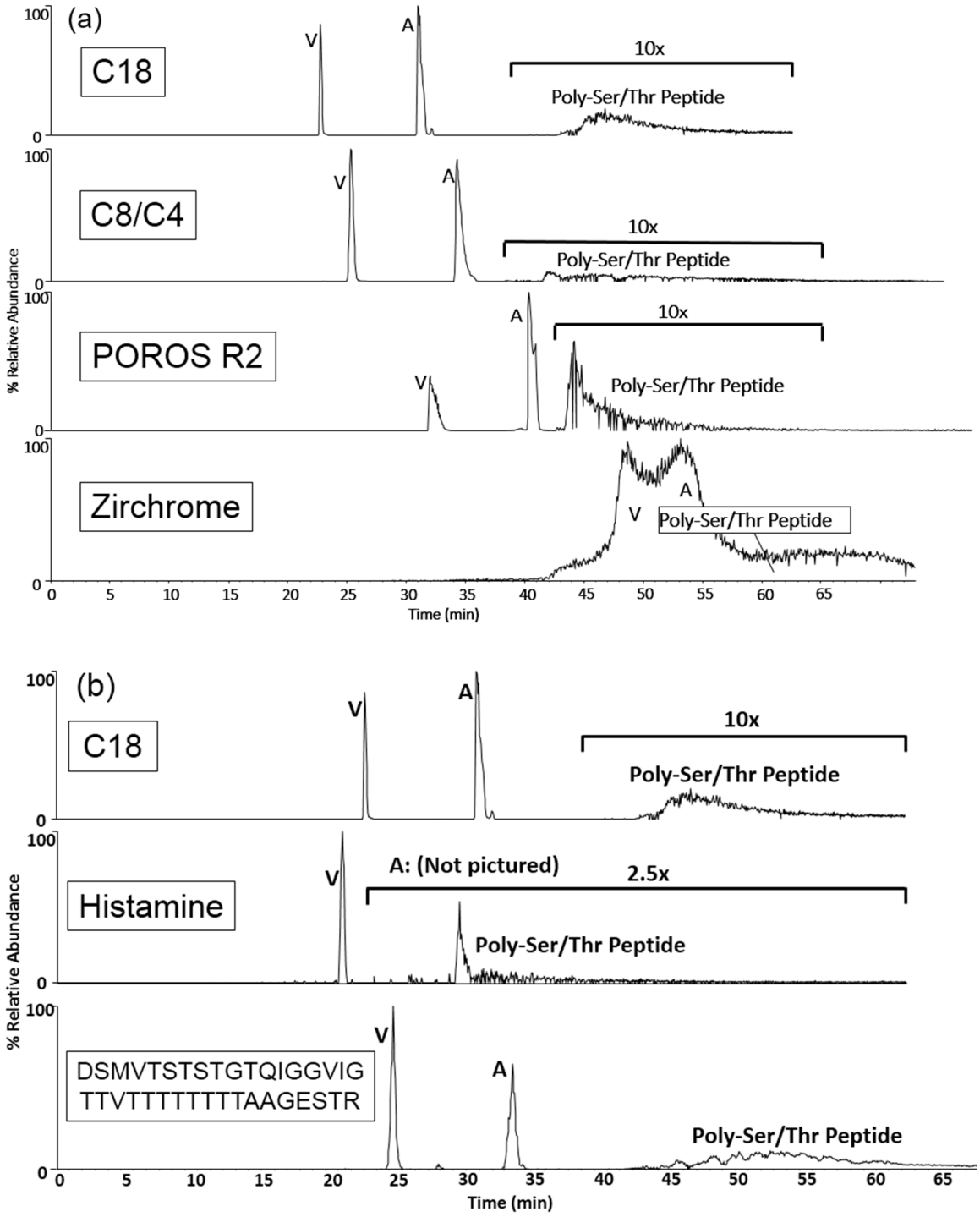


Figure 2.22– *XICs of Angiotensin (A) and Vasactive Intestinal Peptide (V) internal standards and Poly-Ser/Thr peptide SCSSPDSMVTSTSTGTQIGGVIGTTVTTTTTTTTAAGESTR separated by (a) four different chromatographic stationary phases, and (b) C18 after histamine derivatization or Trypsin+AspN digestion.*

Detection of the S/T-rich domain by alternative stationary phases

Several alternative chromatographic stationary phases were chosen to potentially improve the chromatographic profile of the tryptic peptide, SCSSPDSMVTSTSTGTQIGGVIGTTVTTTTTTTTAAGESTR, by decreasing band broadening, tailing, and overall retention. Octyl (C8) and butyl (C4) carbon chains were employed for packing of a PC and AC, respectively. C8, less hydrophobic in nature relative to C18, is traditionally employed for separation of especially hydrophobic peptides, with the even less retentive C4 generally being applied to whole proteins. This setup used C8 for loading and binding of the target hydrophobic peptide, and C4 for analytical separation. A separate ACPC setup was packed with POROS® R2 (Invitrogen) a Styrene-Divinylbenzene polymer phase, sometimes used for biomolecules as an alternative base support to silica-supported carbon chains [59]. POROS® R2 is not further derivatized and is marketed to be suitable for peptides. A third type of stationary media, Zirchrome®-CARB, was selected for its reported use in separations of oligosaccharide isomers by its graphitized carbon-coated surface on ZrO_2 [60]. An AC and PC were both packed with Zirchrome®-CARB with the hypothesis that if carbohydrates eluted effectively, the density of hydroxyls on and potential glycosylation of the target tryptic peptide may give it the appropriate binding affinity to do the same. However, none of the alternate stationary phase setups improved the elution profile and recovery of the target peptide by reverse-phase chromatography.

Detection of the S/T-rich domain by HILIC

One alternative to reverse-phase chromatography is hydrophilic interaction liquid chromatography (HILIC), a technique comparable to normal-phase in which a polar stationary phase is employed. Although the mechanism of separation is not fully understood, it has been

empirically demonstrated that the most hydrophilic residues on peptides tend to contribute the most to retention while the most hydrophobic have little effect or even demote retention [61, 62]. One disadvantage observed is that peptides do not ionize as efficiently sprayed under HILIC mobile phase relative to the aqueous dilute acid employed for reverse-phase and sensitivity suffers as a result. Peptides were loaded onto a C18 micro-capillary column and desalted with 0.1% acetic acid prior to HILIC, because the influence from ionic species could have a significant and undesirable effect on HILIC chromatography and load retention. Though this method of desalting is commonly used for purposes of desalting small amounts (pmols-fmols) of peptides with >50% recovery, it was already known that the peptide in question did not elute with high recovery from C18 by acetonitrile. In an attempt to bypass this problem, the desalted peptides were eluted using an 80% 2-propanol solution as 2-propanol is commonly used as a solvent to stringently clean and regenerate reverse-phase packing material. Unfortunately, a screen of the eluent revealed no peptides were recovered from this cleanup.

Detection of the S/T-rich domain by Derivatization of Carboxylic Acids with Histamine

Finally, a chemical derivatization strategy of derivatizing carboxylic acid residues (Asp, Glu, and C-terminus) with histamine was implemented. This technique was developed by Dr. Michelle English in the Hunt Laboratory with the original purpose of increasing peptide basicity thereby increasing protonation and charge to improve ETD fragmentation [45]. Another consistent effect of this derivatization is decreased retention under reverse-phase, which could potentially be exploited to enhance detection. Improving upon similar techniques [63-65], carboxylic acids on peptides were acid amidated with excess histamine using the water-soluble 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC) coupling reagent in 1 M pyridine-HCl (pH

5-5.5) (**Figure 2.23**) as described in [45].

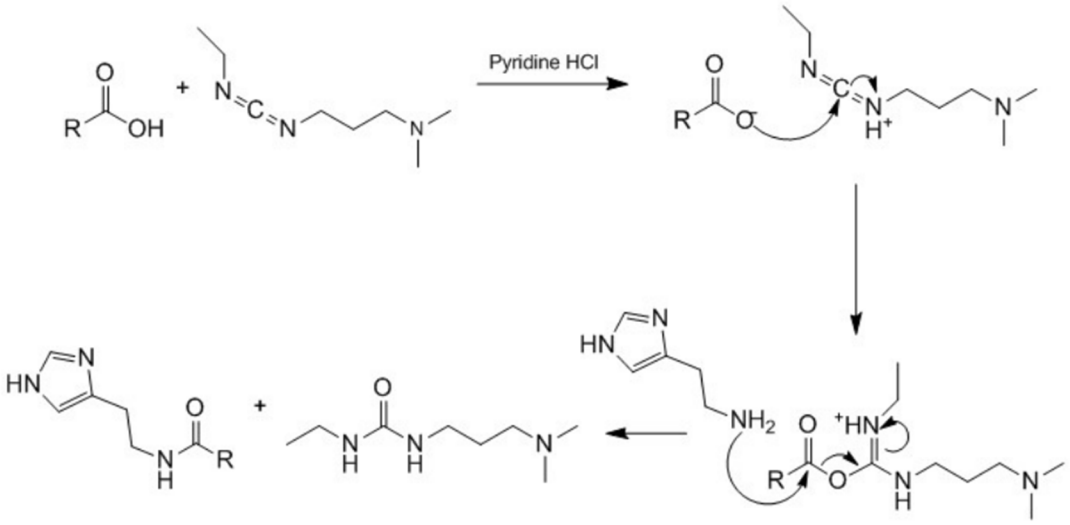


Figure 2.23 – Reaction mechanism for derivatization of carboxylic acids with histamine.

Peptides are derivatized and loaded directly onto C18 columns as the pyridine buffer, histamine, and urea-like byproduct washed through the column. This minimizes the need for further sample handling and potential peptide loss.

SCSSP**D**SMVTSTSTGTQIGGVIGTTVTTTTTTTTAAG**E**ST**R** contained 3 carboxylic acid-containing residues (bolded) to be modified. The peptide was fully derivatized with 3 histamines to 50%, decreasing retention on C18 by 15 min while moderately mitigating tailing. This yielded moderate improvement in chromatographic profile, but sensitivity remained poor.

Detection of the S/T-rich domain by Lysine Insertion

All generated forms of the Ser/Thr-rich region were inadequate as a platform for O-GlcNAc detection and site-mapping due to irreversible binding, extreme chromatographic tailing, and/or poor ionization. Because of this, we employed a site-directed mutagenesis strategy to generate a new tryptic cleavage site that would break the peptide into two relatively

equal halves. A new 6x His 3x FLAG-RGA vector was expressed containing a lysine insertion (bolded) between G185 and G186: SCSSPDSMVTSTSTGTQIG**K**GVIGTTVTTTTTTTTAAGESTR. This site was chosen because it was located in the middle of the longest continuous sequence of residues not containing Ser or Thr and because it would be adjacent to two glycine residues. Glycine is unique amongst the amino acids as it only contains a hydrogen for a side chain, giving it unique flexibility and the least and 4th-least propensity to participate in common secondary structures such as alpha helices and beta sheets respectively, as empirically classified by Chou and Fasman [66]. Because of these factors, this lysine insertion site was hypothesized to be the least likely to cause disruption in any potential RGA secondary or tertiary structure or SEC binding. The 6x His 3x FLAG RGA-G185_G186insK was over-expressed with SEC in tobacco as previously described. Western Blot analysis demonstrated that the decreased mobility of RGA (+SEC) by gel electrophoresis was characteristic of RGA-G185_G186insK as well (**Figure 2.24**).

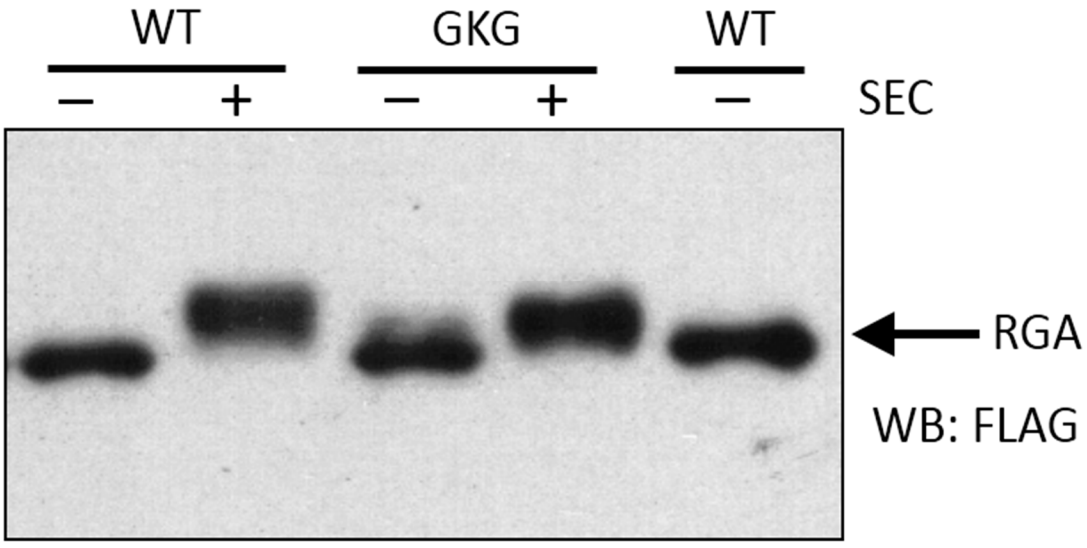


Figure 2.24 – Western Blot with anti-FLAG-HRP of RGA run on 6% acrylamide SDS-PAGE demonstrated that the shift characteristic of RGA(+SEC) still remained for RGA-G185_G186insK (+SEC) samples.

Following reduction, alkylation, and trypsin-digestion, peptides were separated by C18 and analyzed my MS. Contrary to their intact counterpart from RGA (+SEC), peptides SCSSPDSMVTSTSTGTQIGK and GVIGTTVTTTTTTTTAAGESTR were now both detected with a 10+ min decrease in retention, sharp symmetric peak shape for the former, and markedly improved peak shape for the latter.

2.3.5 Posttranslational Modification of RGA(+SEC)

Detection of RGA O-GlcNAcylation

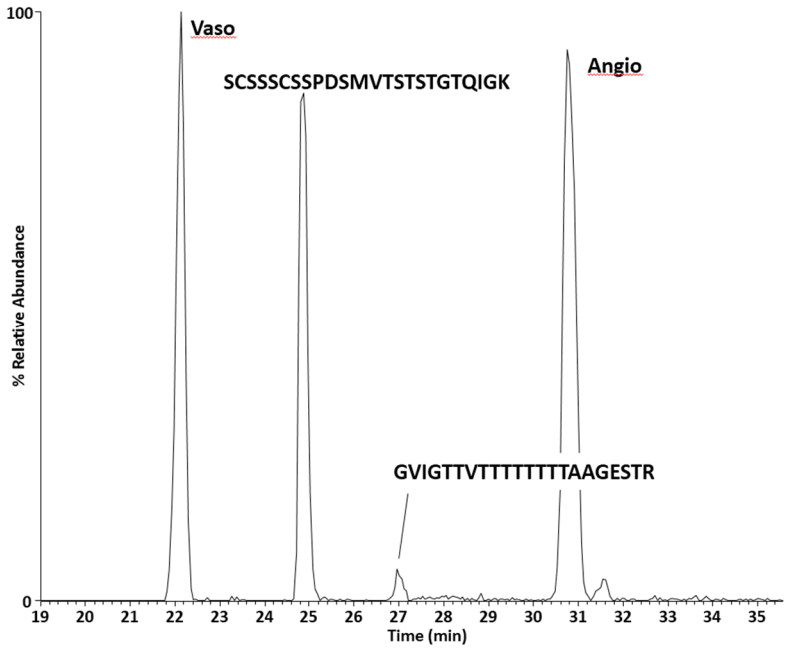


Figure 2.25 – *XICs of the previously undetectable RGA(+SEC) poly-Ser/Thr region eluting as two peptides amongst peptide standards Vaso and Angio after trypsin cleaved it in half at the inserted lysine.*

Through OMSSA-assisted data searches and manual searches and interpretation of MS^1 and MS^2 spectra, 13 unique O-GlcNAcylated peptides were discovered on RGA and RGA-G185_G186insK (+SEC). These were comprised of 6 AspN-generated and 7 trypsin-generated peptides, including SCSSPDSMVTSTSTGTQIGK and GVIGTTVTTTTTTTTAAGESTR, containing a dense population of Ser and Thr residues hypothesized to be heavily O-GlcNAcylated. Owing to

the lysine insertion creating a new tryptic cleavage site, these modifiable residues were now detected by mass spectrometry (**Figure 2.25**) and mass spectral analysis determined that SCSSPDSMVTSTSTGTQIGK contained one, two, and three simultaneously occurring O-GlcNAc sites while GVIGTTVTTTTTTTTAAGESTR contained all combinations of one through six coinciding O-GlcNAc sites (**Figure 2.26**).

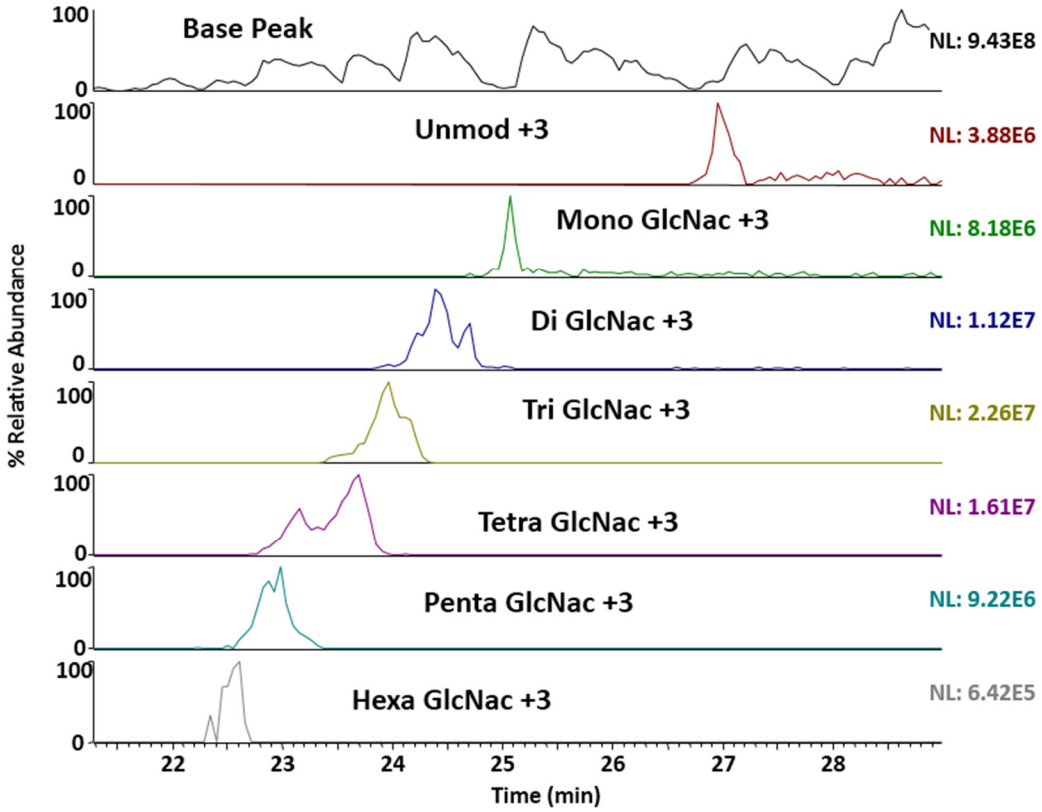


Figure 2.26 - *XICs of 1-6 O-GlcNAcs detected on GVIGTTVTTTTTTTTAAGESTR.*

The two peptides comprised of the remaining two RGA regions that contain >3 sequential Ser/Thr residues were also found to contain O-GlcNAc. Asp-N-generated peptide DHHQFQGRLSNHGTSSSSSSISK was mono-, di-, tri-, and tetra-O-GlcNAcylated while tryptic peptide VIPGNAIYQFPAIDSSSSSNNQNKR was mono-O-GlcNAcylated. Outside of these poly-Ser/Thr domains, an additional nine regions of O-GlcNAcylation were detected, two of which also bore multiple simultaneous O-GlcNAcs.

Detection of RGA Phosphorylation

Yet, O-GlcNAcylation was not the only PTM detected on RGA. Of the 13 O-GlcNAcylated peptides detected, 4 also contained phosphorylated forms. It was previously unknown that RGA could be phosphorylated, but not entirely unexpected as an estimated 1/3 of the proteome may be phosphorylated and the vast majority O-GlcNAcylated proteins are also phosphorylated at some point during their lifetimes, whether separately from or simultaneously with O-GlcNAcylation [3, 7]. While it was known that one or more tobacco kinases were responsible for this phosphorylation of RGA, the exact kinases responsible were unknown. Interestingly, three of the four phosphorylated peptides contained poly-Ser/Thr domains. DHHQFQGRLSNHGTSSSSSSISK contained mono-, di-, and tri-phosphorylation while DPVLPSPEICGFPAS, VIPGNAIYQFPAIDSSSSSNNQNKR, and SCSSPDSMVTSTSTGTQIGK contained mono-phosphorylation. Of the poly-Ser/Thr containing RGA peptides generated, only GVIGTTVTTTTTTTTAAGESTR contained no detectable phosphorylation.

Detection of an unexpected and unprecedented glyco-PTM

By examining MS^1 spectra, it was observed that several species with mass shifts of Δ41 Da were co-eluting with the O-GlcNAcylated forms of GVIGTTVTTTTTTTTAAGESTR (**Figure 2.27**).

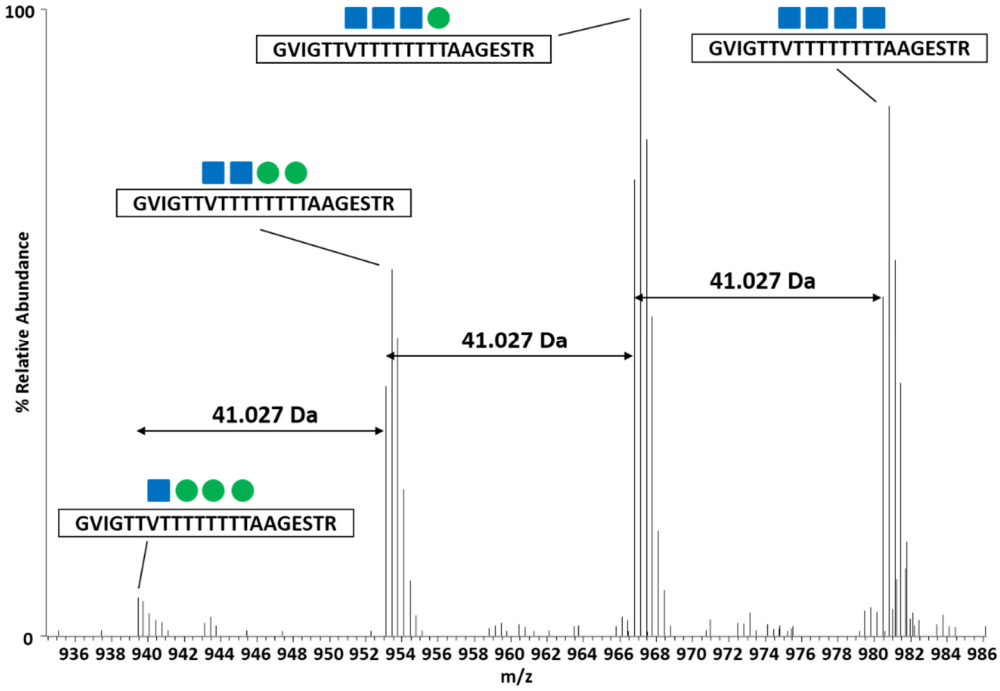


Figure 2.27 - *Masses related to O-GlcNAcylated poly-Thr/Ser domain were due to a previously undescribed nuclear protein PTM, O-hexosylation. O-GlcNAcylation and O-hexosylation are indicated by blue squares and green circles, respectively.*

Interestingly, no such mass shift occurred relative to the unmodified peptide form. Given their similar mass, charge states, isotopic distribution, abundance, and retention, these Δ41 Da shifted-peptides were further investigated for their potential relationship to the O-GlcNAcylated peptides. Because the intact peptides were detected in the Orbitrap with high resolution and accurate mass, the mass shifts were precisely calculated to 41.027 Da. While this Δmass pertained to two known theoretical modifications (i.e. Ser>Gln substitution and amidination of lysines or N-terminal amines with methyl acetimidate [67], neither of these possibilities fit the circumstances. The number of Δ41 Da shifts on numerous single peptides outnumbered the single Ser present on GVIGTTVTTTTTTTTAAGESTR and the sample was never treated with methyl acetimidate (**Figure 2.28**). Given that the shift occurred only in relation to O-GlcNAcylated forms, Δ41 Da was examined relative to O-GlcNAc itself. In fact, a loss of 41 Da equated to the

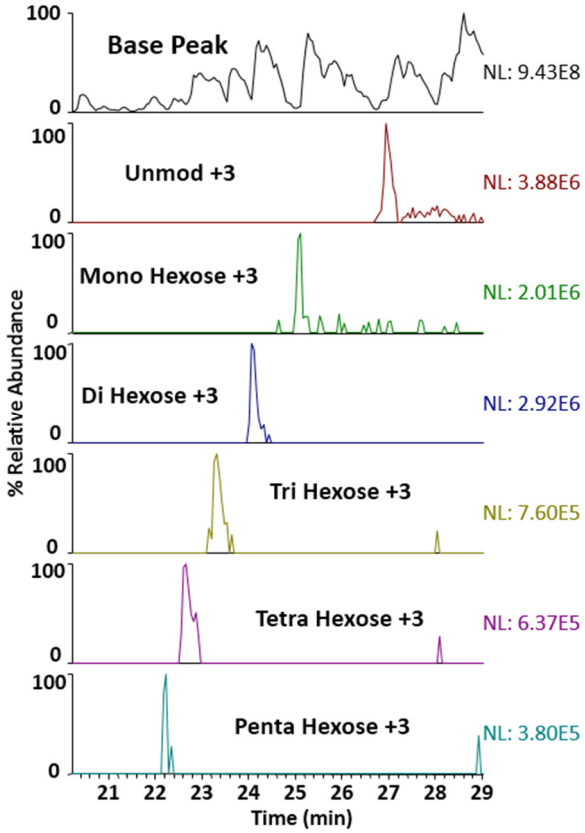


Figure 2.28 – *XICs of 1-5 unexpected Δ162 masses on GVIGTTVTTTTTTTTAAGESTR determined to be O-hexose PTMs.*

mass difference (Δ41.0266 Da) between GlcNAc and a regular hexose sugar (potentially glucose) lacking the N-acetyl moiety.

A full analysis of the data revealed that in total, peptides comprised of six distinct regions of RGA contained one or more hexose moieties. All six of these regions also contained O-GlcNAcylated peptide forms and of these, two were modifiable with all three PTMs found in this study. CAD of hexosylated peptides generated signature ions from the loss of the hexose (**Figure 2.29**), indicating they were O-linked analogous to O-GlcNAc. Several hexose sites were able to be localized by ETD and unambiguously verified to contain O-linkages to Ser or Thr residues. In all cases, O-hexose was detected as a monosaccharide PTM only, with no elongation.

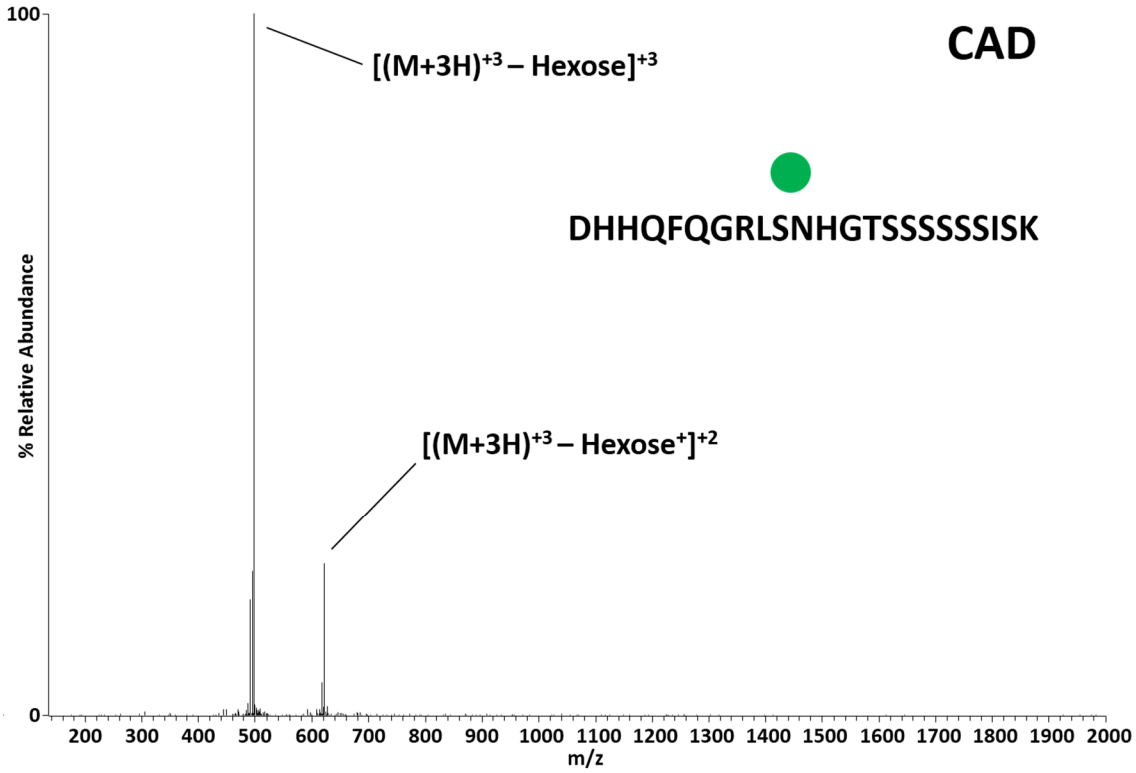


Figure 2.29 – *Hexose PTMs also produced signature fragment ions upon CAD, characteristic of O-linkage. This spectrum indicates that mono-O-hexosylation is present within RGA residues 4-26.*

Prior to the discovery of O-GlcNAc, it was widely believed that nuclear (and cytoplasmic) proteins contained no glycosylation at all. Reports of glycosylation of nuclear and cytoplasmic proteins apart from O-GlcNAcylation remain a minority in the literature, yet there is a growing body of evidence supporting additional glycosylation of nucleocytoplasmic proteins and the nucleocytoplasmic domains of transmembrane proteins [68]. Most of these studies have used lectins (sugar-binding proteins) or anticarbohydrate antibodies to probe for these alleged glycoproteins. These results can be difficult to validate though due to the lack of structural data confirming the glycans and the capability for lectins and antibodies to bind to biomolecules non-specifically (e.g. via hydrophobic interactions). Nucleocytoplasmic glycosylation is also counterintuitive given the present understanding that most glycosyltransferases are membrane

proteins whose active sites are located within the lumen of the Endoplasmic Reticulum (ER) or Golgi. A few possibilities exist to justify the existence of these reports. There is a possibility that unknown and unconventional glycosyltransferases reside in the cytoplasm and/or nucleus. There are indeed reports of glycosyltransferase activity in highly purified rat liver nuclei, although even minor contamination with the ER could yield misleading results. Alternatively, some proteins have been found to relocate to the cytoplasm after entering the ER and there is evidence that endocytosed glycoproteins may enter the cytoplasm in this way. For a third possibility, some membrane-associated glycosyltransferases' are located on the cytoplasmic side of the ER and are responsible for the synthesis of glycolipids and dolichyl-linked N-glycosylation precursors that are afterwards transported into the ER lumen for attachment to proteins. It is conceivable that one or more of these glycoproteins may have developed the capability to glycosylate proteins located in the cytoplasm [68].

Despite these limited reports of other nucleocytoplasmic glycosylation, the detection of an O-linked hexose PTM on RGA was unexpected for two reasons: (1) it is a nuclear protein and O-GlcNAcylation is the only nuclear protein glyco-PTM that has been rigorously validated and widely accepted as discussed above, and (2) simple O-linked monosaccharide glyco-PTMs apart from O-GlcNAc are virtually undocumented regardless of cellular or extracellular location.

A few reported exceptions to this occur for the glycosylation of *Drosophila* and mammalian proteins, containing epidermal growth factor (EGF)-like repeat domains. These EGF repeats consist of 30-40 amino acids including 6 conserved cysteine residues that form 3 intramolecular disulfide bonds [69]. About 25 years ago, the atypical O-linked glycol modification O-glucose was discovered as a trisaccharide sugar chain linked to Ser52 in bovine blood coagulation factors VII and IX [70]. Soon after, the atypical O-linked O-fucose was detected and

localized to the nearby residue, Ser60, on human blood coagulation factor VII (O-glucose oligosaccharide was also again detected on Ser52 in this study) [71]. In the same time period, O-fucose was also detected on factor XII [72] and human Prourokinase, a serine protease first discovered in urine [73]. Later, both modifications were also discovered on the extracellular EGF-domains of Notch1, one of four mammalian notch proteins important for cellular signaling [74]. In all cases, O-fucose maintained the consensus site of Cys_2-X-X-Gly-Gly-**Ser/Thr**-Cys_3 where the site of modification is indicated on the bolded Ser or Thr between the 2nd and 3rd Cys residue of the EGF repeat and X is any residue. Depending on the clotting factor, it was determined to exist as either a monosaccharide or oligosaccharide modification, and was detected in both forms on Notch1. O-glucose held the consensus sequence Cys_1-X-**Ser**-X-Pro-Cys_2 where the site of modification is indicated on the bolded Ser between the 1st and 2nd Cys residue of the EGF repeat and X is any residue. It was also detected as both a monosaccharide and oligosaccharide on Notch1, but was further elongated on all clotting factors described. Many additional O-glucose and O-fucose glycosylation sites have been discovered on Notch protein, mainly from studies using *Drosophila*, with the modifications all adhering to their previously characterized consensus motifs [75, 76]. These rare glyco-modifications are interesting in light of the O-hexosylation discovered by this study of RGA, especially as O-glucose is an O-hexose modification. However, these literature modifications differ from RGA O-hexose in a few significant ways: (1) Notch and coagulation factors are restricted to metazoans, (2) these proteins are modified on consensus sequences of EGF domains by glycosyltransferases in the ER or Golgi, (3) the modified domains are extracellularly located, and (4) the monosaccharide modifications can still be elongated into oligosaccharides in many cases. It is worth noting that Notch proteins also contain the only documented sites of extracellular O-GlcNAc, the addition of which is catalyzed by an epidermal O-GlcNAc transferase (EOGT) in both Drosophila [77] and

mammals [78]. In both cases O-GlcNAc was mapped to a Thr residue located between the 5th and 6th Cys residues of EGF repeats, and evidence suggests that it may be elongated at least in mammals. This EOGT is distinct from what is traditionally referred to as OGT, still the only known enzyme responsible for catalyzing the addition of O-GlcNAc to nucleocytoplasmic proteins of metazoans. For the reasons described above, it is reasonable to conclude that like EOGT and nucleocytoplasmic O-GlcNAcylation, the glycosyltransferases responsible for O-fucose and O-glucose are entirely separate from what is responsible for the O-hexosylation of RGA described in this dissertation. This new O-hexose is unprecedented for the fact that it resides on a nuclear protein and one that is expressed by plants.

Stoichiometry and Site-locations of RGA O-GlcNAcylation, Phosphorylation, and O-hexosylation

An in-depth analysis of RGA PTMs revealed 23 O-GlcNAc, 6 phosphate, and 14 O-hexose sites based on the detection of these modifications on intact peptides (MS^1) and ETD fragment ions (MS^2). For reasons previously discussed, fragmentation by ETD preserved all three of these labile PTMs during MS^2 and enabled the majority of these sites to be localized to the exact residue modified as summarized in **TABLE 2.3**.

PTM Site	O-GlcNAc	Phosphate	O-Hexose
S13		X	
T17	X	X	X
S18	X		X
S19	X		X
S20	X	X	X
S23		X	
S56	X		X
S57	X		X
S82	X		X
S115	X		X
S116	X		X
S170		X	
T176	X		
T177	X		
T178	X		
S253	X		
S537	X		
	Total	Total	Total
	14	5	9

Table 2.3 - 14 O-GlcNAc, 5 phosphate, and 9 O-hexose sites were unambiguously localized to the following Ser or Thr residues as indicated by "X".

Peptides were observed to be modified in a wide range of stoichiometries (<1% to >50%) and to various degrees of combinatorial PTM complexity including the presence of all three modifications at once. A summary of all modified peptides and corresponding PTM sites detected, totaling 89 distinct forms in all, is presented in **Table 2.4**.

Residues	Peptide sequence	# Mods			Site(s)	PPM	z	Relative Abundance[a]	Digest			IMAC	Histamine
		GlcNAc	Phospho	Hexose					Trypsin	AspN	CNBr		
4-26	DHHQFQGRLSNHGTgSSgSSSSISK	1			S18 & S20	-0.3	5	++++ +		X			
4-26	DHHQFQGRLSNHGg[TgSgSSS]SSISK	2			S18, 19, & []d	-0.1	5	++++		X			
4-26	DHHQFQGRLSNHGgTgSg[SSS]SSISK	3			T17, S18, & []d	<0.1	5	++		X			
4-26	DHHQFQGRLgggg(SNHGTSSSSSSIS)K	4			bc	0.4	5	+		X			
4-26	DHHQFQGRLpSNHGp[TSSS]SSSISK		1		S13 & []d	-0.4	4	+		X			
4-26	DHHQFQGRLpSNHGp[TSSpSSSSIS]K		2		S13 & S20	-0.3	4	+		X		X	
4-26	DHHQFQGRLppp(SNHGTSSSSSSIS)K		3		bc	0.7	4	+		X			
4-26	DHHQFQGRLSNHGThSSSSSSISK			1	S18	0.5	5	++		X			
4-26	DHHQFQGRLSNHGh[TSS]SSSSISK			1	d	-0.9	6	++		X			X
4-26	DHHQFQGRLSNHGThShSSSSSISK			2	S18 & S19	0.6	5	+		X			
4-26	DHHQFQGRLhhh(SNHGTSSSSSSIS)K			3	bc	1.3	5	+		X			
4-26	DHHQFQGRLpSNHGgTgSgSSSSSISK	1	1		pS13 & gT17, gS18, gS19	<0.1	5	++		X		X	
4-26	DHHQFQGRLpSNHGpg[TSSS]SSpSISK	1	2		pS13, gp[]d, & pS23	-0.8	4	+		X		X	
4-26	DHHQFQGRLpSNHGpg[TSSSSS]pSISK	1	3		pS13, gp[]d, & pS23	-1.6	4	+		X		X	
4-26	DHHQFQGRLpSNHGpgTgSgSSSSISK	2	1		pS13, pgT17, gS18, & gS19	0.7	5	++		X			
4-26	DHHQFQGRLgggp(SNHGTSSSSSSIS)K	3	1		bc	3.7	5	+		X			
4-26	DHHQFQGRLggpp(SNHGTSSSSSSIS)K	2	2		bc	1.0	4	+		X			
4-26	DHHQFQGRLSNHGpgh[TSS]SSSSISK	1	1	1	pgh[]d	-1.7	4	+		X			
4-26	DHHQFQGRLSNHGTghSghSghSSSSISK	1		1	ghS18, ghS19, ghS20	-0.3	5	++		X			

4-26	DHHQFQGRLghh(SNHGTSSSSSSIS)K	1		2	[bc]	-1.1	5	+		X			
4-26	DHHQFQGRLpSNHGphThShSSSSSISK		1	1	pS13, phT17, hS18, & hS19	-1.5	4	+		X			
54-65	VRSSEMAEVALK	1			gS56 & gS57	-0.3	3	++++	X				
54-65	VRSSEMAEVALK	2			gS56 & gS57	-0.3	3	+	X				
44-59	DhistELLAVLGYKVRSShistEM-hsl	1			S56	0.2	4	++++		X	X		
44-59	DhistELLAVLGYKVRSShistEM-hsl	1			S57	0.2	4	++		X	X		
44-59	DhistELLAVLGYKVRSShistEM-hsl			1	S56	0.8	4	++		X	X		
44-59	DhistELLAVLGYKVRSShistEM-hsl			1	S57	0.8	4	+		X	X		
79-86	DGLgSHLAT	1			S82	<0.1	2	+++		X			
79-86	DGLhSHLAT			1	S82	-2.1	2	+		X			
87-100	Dg(TVHYNPSELYS)WL	1			[bc]	3.9	2	+		X			
120-134	DPVLPg(SPEICGFPAS)	1			[bc]	2.3	2	++		X			
120-134	DPVLPp(SPEICGFPAS)		1		[bc]	1.9	2	+		X			
104-119	LShistELNPPPLPASgSNGL-hist	1			gS116	0.1	3	+++++		X	X		X
104-119	Lgg(ShistELNPPPLPASS)NGL-hist	2			[bc]	0.6	3	+		X	X		X
104-119	LShistELNPPPLPAShSNGL-hist			1	hS116	-0.2	3	++++		X	X		X
104-119	LShistELNPPPLPAhSgSNGL-hist	1		1	hS115, gS116	-1.0	3	++		X	X		X
104-119	LShistELNPPPLPAgShSNGL-hist	1		1	g115, hS116	-1.0	3	+		X	X		X
137-171	DLKVIPGNAIYQFPAIDg(SSSSSNNQNKRLKSCSS)P	1			[c]	2.0	5	+		X			
140-164	VIPGNAIYQFPAIDSSg(SSS)NNQNKR	1			[c]	3.2	4	++	X				
140-164	VIPGNAIYQFPAIDp(SSSSS)NNQNKR		1		[bc]	-0.2	3	+	X				
153-171	DSSSSSNNQNKRLKg(SCSS)P	1			[c]	1.1	3	+++		X			
153-171	Dgg(SSSSSNNQNKRLKSCSS)P	2			[bc]	3.9	3	+++		X			
153-171	DSSSSSNNQNKRLKp(SCSS)P		1		[c]	1.0	3	++++		X			
153-171	Dh(SSSSSNNQNKRLKSCSS)P			1	[bc]	1.6	3	+		X			
167-185	SCSSPDSMVgTSTSTGTQIGK	1			T176	-0.3	3	+++	X				
167-185	SCSSPDSMVTSgTSTGTQIGK	1			T178	-0.3	3	+++	X				
167-185	SCSSPDSMVg[TSTS]TGTQIGK	1			g[][d]	-0.3	3	+	X				
167-185	g(SCSSPDS)MVTSTSTGTQIGK	1			[c]	-0.9	3	+++	X				
167-185	g(SCSSPDSMVTSTSTGT)QIGK	1			[bc]	-0.1	3	+++	X				
167-185	gg(SCSSPDSMVTSTSTGT)QIGK	2			[ce]	0.4	3	+++	X				

167-185	ggg(SCSSPDSMVTSTSTGT)QIGK	3			[ce]	0.0	3	++	X				
167-185	SCSpSPDSMVTSTSTGTQIGK		1		S170	0.2	3	++	X				
167-185	h(SCSSPDSMVTSTSTGT)QIGK			1	[bc]	-0.8	3	++	X				
167-185	SCSpSPDSMVgTSTSTGTQIGK	1	1		pS170 & gT177	-2.1	3	++	X				
167-185	pg(SCSSPDSMVTSTSTGT)QIGK	1	1		[ce]	0.6	3	++	X				
167-185	pgg(SCSSPDSMVTSTSTGT)QIGK	2	1		[ce]	-0.3	3	++	X				
167-185	pggg(SCSSPDSMVTSTSTGT)QIGK	3	1		[ce]	-0.2	3	+	X				
167-185	gh(SCSSPDSMVTSTSTGT)QIGK	1		1	[bc]	0.4	3	++	X				
186-207	GVIGg(TTVTTTTTTTTAAGEST)R	1			[ce]	-0.8	3	++	X				
186-207	GVIGgg(TTVTTTTTTTTAAGEST)R	2			[ce]	-1.0	3	++	X				
186-207	GVIGggg(TTVTTTTTTTTAAGEST)R	3			[ce]	-0.7	3	+++	X				
186-207	GVIGgggg(TTVTTTTTTTTAAGEST)R	4			[ce]	-1.4	3	+++	X				
186-207	GVIGggggg(TTVTTTTTTTTAAGEST)R	5			[ce]	-0.2	3	++	X				
186-207	GVIGgggggg(TTVTTTTTTTTAAGEST)R	6			[ce]	-0.3	3	+	X				
186-207	GVIGh(TTVTTTTTTTTAAGEST)R			1	[bc]	-0.3	3	+	X				
186-207	GVIGhh(TTVTTTTTTTTAAGEST)R			2	[bc]	-0.4	3	++	X				
186-207	GVIGhhh(TTVTTTTTTTTAAGEST)R			3	[bc]	-0.3	3	+	X				
186-207	GVIGhhhh(TTVTTTTTTTTAAGEST)R			4	[bc]	-1.2	3	+	X				
186-207	GVIGhhhhh(TTVTTTTTTTTAAGEST)R			5	[bc]	-1.7	3	+	X				
186-207	GVIGgh(TTVTTTTTTTTAAGEST)R	1		1	[ce]	-0.8	3	++	X				
186-207	GVIGghh(TTVTTTTTTTTAAGEST)R	1		2	[ce]	0.8	3	++	X				
186-207	GVIGghhh(TTVTTTTTTTTAAGEST)R	1		3	[ce]	-0.4	3	++	X				
186-207	GVIGghhhh(TTVTTTTTTTTAAGEST)R	1		4	[ce]	-2.1	3	++	X				
186-207	GVIGghhhhh(TTVTTTTTTTTAAGEST)R	1		5	[ce]	-0.1	3	+	X				
186-207	GVIGggh(TTVTTTTTTTTAAGEST)R	2		1	[ce]	-0.5	3	+++	X				
186-207	GVIGgghh(TTVTTTTTTTTAAGEST)R	2		2	[ce]	-0.8	3	+++	X				
186-207	GVIGgghhh(TTVTTTTTTTTAAGEST)R	2		3	[ce]	-1.9	3	++	X				
186-207	GVIGgghhhh(TTVTTTTTTTTAAGEST)R	2		4	[ce]	-1.1	3	+	X				
186-207	GVIGgggh(TTVTTTTTTTTAAGEST)R	3		1	[ce]	-1.0	3	+++	X				
186-207	GVIGggghh(TTVTTTTTTTTAAGEST)R	3		2	[ce]	0.4	3	+++	X				
186-207	GVIGgggghhh(TTVTTTTTTTTAAGEST)R	3		3	[ce]	-0.7	3	+	X				
186-207	GVIGggggh(TTVTTTTTTTTAAGEST)R	4		1	[ce]	0.8	3	+++	X				
186-207	GVIGgggggh(TTVTTTTTTTTAAGEST)R	4		2	[ce]	-2.2	3	++	X				
186-207	GVIGggggggh(TTVTTTTTTTTAAGEST)R	5		1	[ce]	0.3	3	+	X				
246-259	QIGCLAVgSQAGAMR	1			gS253[e]	-4.1	2	+	X				

466-489	Fg(TESLHYYSTLFDSLEGVPNS)QDK	1			[ce]	1.8	3	+	X				
535-541	FGgSSGLAPAHLGSNAFK	1			gS537	-1.0	3	++	X				
581-587	Lg(ST)AAY	1			[ce]	1.1	1	++	X				

Table 2.4 – A summary of the number of modified residues and site-locations of O-GlcNAc, phosphate, and/or O-hexose PTMs on RGA peptides from (a) RGA(+SEC). [a]PTM levels are reported as % total peptide abundance detected, (+++++) = 50+% (++++) = 20-49%, (+++) = 10-19%, (++) = 1-9%, (+) = <1%, (-) no modified peptide detected. % Total abundances were calculated from ion currents observed in the main beam mass spectra [(modified peptide ion current)/ (modified peptide + unmodified peptide ion current) x (100)]. [b]Modified peptide containing this PTM site(s) was assigned by HPLC retention time and correct accurate mass, only. [c]PTM site is located on S/T/Y (Y – phosphorylation only) within (). [d]A mix of PTM sites were detected within brackets []. [e]Site was determined by correct elution profile, accurate mass, and signature CAD fragmentation. [f]Mass accuracy errors are reported in parts per million (PPM) relative to the mean experimental accurate mass of Angiotensin and Vasoactive Intestinal Peptide standard peptides. All cysteine residues were detected as carbamidomethylated. The digest(s) used to generate each peptide is denoted "X" and the positive charge state (z) reported. Peptides enriched using IMAC or derivatized with histamine are also denoted "X". O-GlcNAcylation, phosphorylation, and O-hexosylation are abbreviated by "q", "p", and "h", respectively.

Two of the localized phosphorylation sites, S23 and S170, matched motifs Ser-X-X-**Ser/Thr** and/or Ser/Thr-X-X-X-**Ser** of Casein kinase 1 (CK1), a ubiquitous kinase in eukaryotes involved in signal transduction. No distinct consensus site exists for O-GlcNAc, but it is often located in Ser/Thr-rich regions of proteins [11]. This is the case for RGA as all four poly-Ser/Thr regions contain at least one O-GlcNAc. Peptides DHHQFQGRLSNHGTSSSSSISK, SCSSPDSMVTSTSTGTQIGK, and GVIGTTVTTTTTTTTAAGESTR, representing three of these four regions, are all heavily modified with multiple combinations of PTMs. Peptides containing multiple Ser and Thr residues, especially in long stretches, present a significant analytical challenge because fragmentation must be nearly perfect to determine the exact site of modification, and these areas don't contribute any charge to help facilitate the ETD fragmentation that is needed. This problem is compounded by the fact that peptides modified on different sites but with the same number and types of PTMs (e.g. mono O-GlcNAcylation of S18 vs. mono O-GlcNAcylation of S20 for peptide DHHQFQGRLSNHGTSSSSSISK) often coelute. Therefore, when the intact peptides are isolated in preparation for MS^2, both forms are isolated and fragmented together, generating mixed ETD spectra as the ETD spectrum of mono O-GlcNAcylated DHHQFQGRLSNHGTSSSSSISK shown in **Figure 2.19b** demonstrates. The z ion

series fragments contain an addition of 203 Da starting with z_7 because one peptide form contains an O-GlcNAc at S20 (**Figure 2.19c**). However, there are also z_7 and z_8 fragments that do not contain the +203 Da addition because another form of the peptide contains the O-GlcNAc on S18. The same phenomenon occurs with the c ion series as one sequences from the N-terminus to the C-terminus. Due to the strong abundance of these two sites and the ability of AspN to generate the modified peptide that also contains several basic His/Arg/Lys residues (predominantly $z = 5$), these sites were localized by ETD despite the mixed spectrum. Still, it is important to consider that alternative mono O-GlcNAcylated forms that are present at much lower levels are very difficult to detect and validate, so there is a possibility that these exist too.

Several tools were implemented to aid in the detection and sequencing of these three PTMs. Immobilized Metal Affinity Chromatography was used to enrichment for phosphorylated peptides, an alternative site of cleavage (C-terminus of Met) was exploited using cyanogen bromide (CNBr) to generate select peptides more amenable to ETD, and derivatization of carboxylic acids with histamine was used to enhance the charge density and overall charge states of certain peptides, also for the purpose of improving ETD efficiency.

Enrichment of Phosphorylated Peptides by IMAC

Due to the detection of RGA phosphorylation by tobacco kinases, enrichment by IMAC was employed to further improve the detection and localization of these sites. This takes advantage of the ability of Fe (III) to bind phosphate and later release it when it is reduced to Fe (II) (**Figure 2.30**). To prevent carboxylic acids from also binding to Fe (III), peptides were first esterified with methanolic HCl then loaded onto an IMAC column constructed in-house as described in Methods. Phosphorylated peptides were bound, extraneous peptides washed

through, and after connecting the IMAC column to a C18 pre-column, Fe (III) was reduced to Fe (II) with ascorbic acid to elute the enriched phosphorylated peptides onto the pre-column.

By enriching several different phosphorylated forms of DHHQFQGRLSNHGTSSSSSSISK, thereby

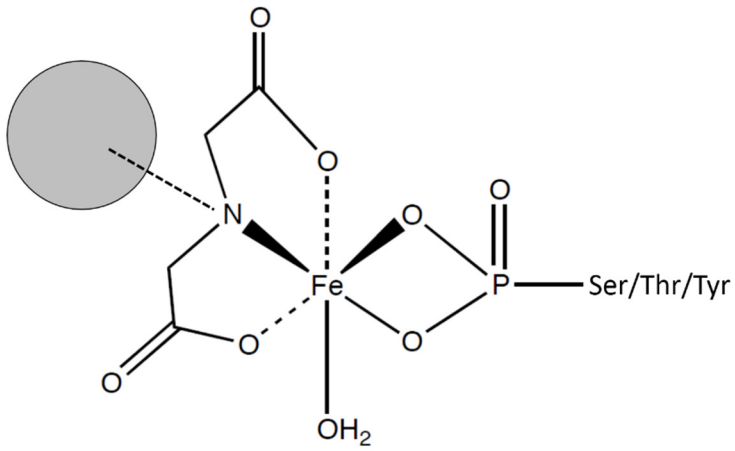


Figure 2.30 – *Fe (III) coordinated to imidodiacetate on a solid support (POROS®MC20) specifically binds phosphorylated peptides. Non-phosphorylated peptides are washed away and the enriched phosphorylated peptides are released by reducing Fe (III) to Fe (II) using ascorbic acid.*

mitigating dynamic range limitations that decrease isolation and sequencing ability, modified sites were elucidated for multiple PTM forms that would otherwise have been very difficult to map.

An alternative peptide cleavage technique using CNBr

Several PTMs were detected on tryptic or Asp-N-generated peptides that were difficult to localize primarily because of a lack of charge to facilitate ETD. These included peptides GVIGTTVTTTTTTTTAAGESTR, SSEMAEVALK, and LSTAAY. The first, encompassing the poly-Thr region, had its ion current split between two charge states: z = 2 and z = 3. The first two protonation sites occurred on the N-terminal amine and the guanidinium group of arginine. The third region available for protonation was the series of amides along the peptide backbone itself. Despite half of GVIGTTVTTTTTTTTAAGESTR charging to z=3 upon ESI, ETD fragmentation

remained poor. This was likely due to two main factors: (1) the lack of internal basic and other residues empirically shown to frequently be associated with regions of high ETD fragmentation, and (2) low charge density relative to mass (i.e. relatively high m/*z*). Li et al compared fragmentation efficiencies between different residues using K-means clustering applied to several thousand peptides generated from LysC, GluC, and trypsin protein digests [79]. Fragmentation involving Gly, Ile, and Val tended to be relatively low in frequency, those involving Ala, Ser, and Thr average, and those involving residues Glu, His, Asn, Gln, Arg, and Lys were relatively high. The former two groups constitute 20 of the 22 residues in this peptide. A separate study by Good et al investigated the relationship between total charge, total mass, and charge density on ETD fragmentation efficiency using several thousand yeast and *Arabidopsis* peptides (z = 3-5) generated from Lys-C or trypsin digests [80]. They found that for charges states z= 3-5, ETD fragmentation efficiency was independent of the total peptide mass or number of charges (**Figure 2.31**).

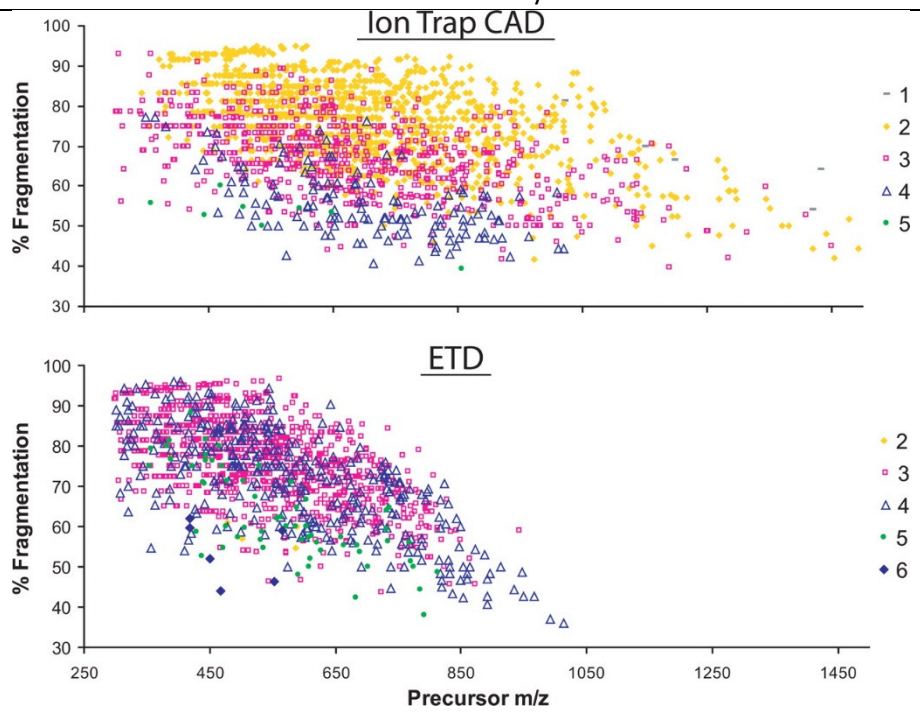


Figure 2.31 – *ETD % Fragmentation (defined by the number of observed fragment ions divided by the number of theoretical fragments) is relatively independent of charge state and peptide mass, but is highly dependent on charge density. Fragmentation efficiency decreases in a linear fashion as m/z increases. Figure reproduced from [80].*

The key factor was charge density, with ETD fragmentation decreasing linearly with decreasing charge density. For m/*z* > ~850, or residue-to-charge ratios >~8, ETD fragmentation efficiency suffered greatly with less than half of the theoretical fragment ions observed. The poly-Thr peptide and its PTM forms fit this challenging demographic as the *z* = 3 form has 7.33 residues per charge and an unmodified mass-to-charge ratio of 710 m/*z*. The most abundant PTM forms, containing mono O-hexosylation and either di or tri O-GlcNAcylation, respectively hold mass-to-charge ratios of 899 and 967 m/*z*. It is possible that the large sugar modifications themselves may contribute to diminished ETD efficiency by sterically hindering the interaction between the electron transfer reagent and carbonyl groups along the peptide backbone. Similarly, the large potential for hydrogen bonding may be realized in a way that partially folds and constricts this entire region of the peptide away from reagent. Also, the amide moieties on each O-GlcNAc are

exact mimics of the electrophilic amides along the peptide backbone, and can contribute to "wasted" electron transfers that do not contribute to backbone fragmentation that generates c and z ions.

Although other enzymes exist in the proteomics repertoire to generate alternative peptides in these regions, none were suitable to generate peptides that would exhibit appropriate chromatographic elution and ETD fragmentation. One strategy used to address this challenge was to take advantage of a non-enzymatic cleavage method first described by Gross Witkop [81] in which CNBr cleaves proteins C-terminally to Met residues converting them into homoserine lactones (**Figure 2.32**).

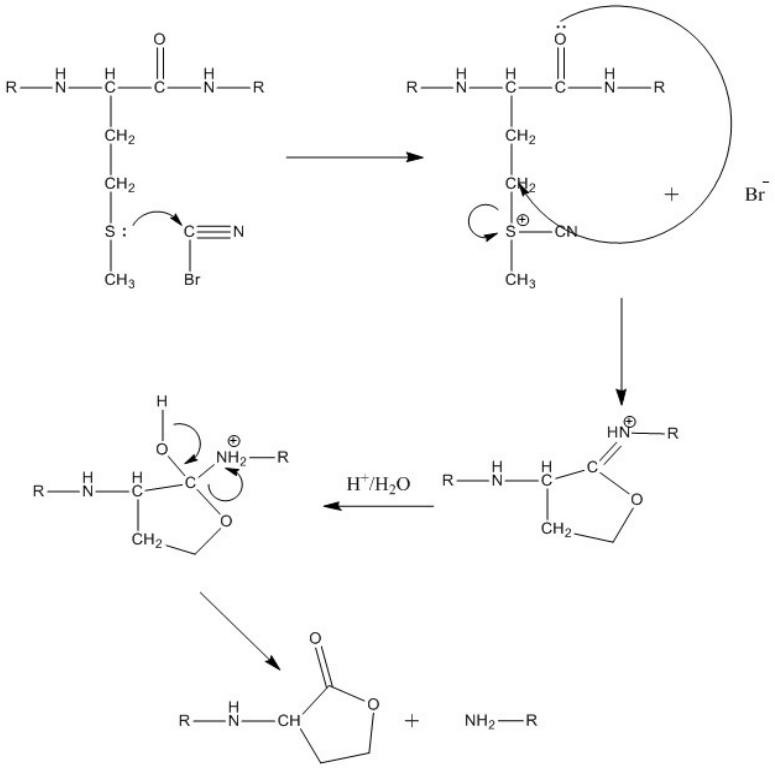


Figure 2.32 – *Reaction scheme for cleavage of Met-X bonds by CNBr where X is any amino acid.*

Because Met is a lower frequency amino acid in most proteins, cleavage by CNBr can be used to cleave proteins into a few large fragments very efficiently, as the mechanism does not require that a large enzyme access the sites to be cleaved. This may help further protein denaturation in preparation for efficient enzymatic digestion. Here, cleavage of Met's C-terminus was used to increase the library of peptides that would encompass the PTMs already detected, but also bear additional charge to enhance ETD fragmentation. RGA contains 19 Met residues and *in silico* cleavage by CNBr is reported in (**Table 2.5**) with peptides containing known modified regions highlighted in gray.

Mass	Position	Peptide	# HKR	# ST	Comments
3093.5	2-29	KRDHHQFQGRLSNHGTSSSSSSISKDKM	8	9	+
1114.5	32-42	VKKEEDGGGNM	2	0	+
1876	43-59	DDELLAVLGYKVRSSEM	2	2	+
1425.8	60-72	AEVALKLEQLETM	1	1	+
3386.5	74-103	SNVQEDGLSHLATDTVHYNPSELYSWLDNM	2	6	+
7424.6	104-174	LSELNPPPLPASSNGLDPVLPSPEICGFPASDYDLKVIPG NAIYQFPAIDSSSSSNNQNKRLKSCSSPDSM	4	14	-
5171.7	175-226	VTSTSTGTQIGK*GVIGTTVTTTTTTTTAAGESTRSVILV DS QENGVRLVHALM	4*	20	?
3180.6	227-258	ACAEAIQQNNLTLAEALVKQIGCLAVSQAGAM	1	2	+
4056.1	259-293	RKVATYFAEALARRIYRLSPPQNQIDHCLSDTLQM	4	6	+
4339.2	294-330	HFYETCPYLKFAHFTANQAILEAFEGKKRVHVIDFSM	7	3	+
1108.6	331-340	NQGLQWPALM	0	0	+
6763.4	341-404	QALALREGGPPTFRLTGIGPPAPDNSDHLHEVGCKLA QLA EAIHVEFEYRGFVANSLADLDASM	7	5	?

9657.1	405-491	LELRPSDTEAVAVNSVFELHKLLGRPGGIEKVLGVVKQIK PVIFTVVEQESNHNGPVFLDRFTESLHYYSTLFDSLEGV P NSQDKVM	11	10	-
5796.8	492-545	SEVYLGKQICNLVACEGPDRVERHETLSQWGNRFGSS GLA PAHLGSNAFKQASM	7	7	?
2254.1	546-566	LLSVFNSGQGYRVEESNGCLM	1	3	+
2372.3	567-587	LGWHTRPLITTSAWKLSTAAY	3	6	+

*Table 2.5 – Theoretical peptides generated by in silico digest of RGA by CNBr. Peptide masses (Da) and position numbers are listed along with the number of basic (# HKR) and GlcNAc-modifiable (# ST) residues for each theoretical peptide. The comments section refers to peptides predicted to be detected (+) or not detected (-), with (?) indicating uncertainty of detection. *Indicates lysine insertion.*

CNBr cleavage theoretically generates yet another form of two adjacent and heavily modified poly-Ser/Thr regions, consisting of residues T190-T206. This peptide, VTSTSTGTQIGKGVIGTTVTTTTTTTTAAGESTRSVILVDSQENGVRLVHALm where lowercase "m" represents the newly formed homoserine lactone residue, would contain three more basic residues (including the lysine insertion) than its tryptic counterpart. In addition, it would contain three carboxylic acids able to be derivatized with histamine to carry charge. Then, if all basic functional groups plus the peptide backbone were protonated upon ESI, this CNBr-generated peptide would contain 9 charges for a residue-to-charge ratio of 5.8. Non-PTM forms of the z = 9 peptide would hold a mass-to-charge ratio of 611 m/z.

To test Met cleavage by CNBr, a slightly modified method from Yaroslav et al. [82] was used to cleave the single Met residue on human recombinant protein H4 and generate extreme C-terminal peptide, DVVYALKRQGRTLYGFGG. Dissolving CNBr in acetonitrile and adding a 20-100 fold excess over Met residues to the sample dissolved in 70% tri-flouroacetic acid (TFA), as described in the literature, failed to generate the target peptide. Instead, as described in Materials and Methods, a small crystal (<1 mg) of CNBr was added directly to the sample. Mass spectrometric analysis revealed dominant signal from the CNBr-derived peptide,

DVVYALKRQGRTLYGFGG, as theoretically predicted. The enhancement of cleavage was probably due to the very large excess of CNBr over Met (>1000x) and potentially increased water access to Met, because no organic solvent was used.

Next, the method was applied to whole RGA-G185_G186insK, eluted from anti-FLAG beads after immunoprecipitation with acetic acid. The elution was performed to avoid potential degradation of the agarose beads by strong acid which could unnecessarily complicate the sample. Unfortunately, the theoretical poly-Ser/Thr peptide was not detected, apparently being too lengthy to efficiently elute from the chromatographic columns. However, *in silico* CNBr subcleavage of an AspN digest revealed three unique peptides (one synonymous with CNBr-only cleavage) that would contain previously detected PTMs. These peptides are highlighted in gray in **Table 2.6**.

Mass	Position	Peptide	# HKR	# ST	Comments
344.2	27-29	DKM	1	0	-
631.4	32-36	VKKEE	2	0	-
1760.9	44-59	DELLAVLGYKVRSSEM	2	2	+
1425.8	60-72	AEVALKLEQLETM	1	0	+
575.3	74-78	SNVQE	0	1	?
330.1	101-103	DNM	0	0	-
1604.8	104-119	LSELNPPPLPASSNGL	0	3	+
1519.8	213-226	DSQENGVRLVHALM	2	1	-
3180.6	227-258	ACAEAIQQNNLTLAEALVKQIGCLAVSQAGAM	1	2	-
2960.6	259-283	RKVATYFAEALARRIYRLSPPQNQI	5	2	-
558.3	289-293	DTLQM	0	1	?
3907.0	294-326	HFYETCPYLKFAHFTANQAILEAFEGKKRVHVI	7	2	+
450.2	327-330	DFSM	0	1	-
1108.6	330-339	NQGLQWPALM	0	0	+
2315.3	340-363	QALALREGGPPTFRLTGIGPPAP	2	2	+
374.1	401-404	DASM	0	1	-
713.4	405-410	LELRPS	1	1	?

443.2	488-491	DKVM	1	0	-
1921.9	492-509	SEVYLGKQICNLVACEGP	1	1	+
3892.9	510-545	DRVERHETLSQWGNRFGSSGLAPAHLGSNAFKQ ASM	6	6	+
2254.1	546-566	LLSVFNSGQGYRVEESNGCLM	1	3	+
2372.3	567-587	LGWHTRPLITTSAWKLSTAAY	3	6	+

Table 2.6 - Theoretical peptides generated by in silico digest of RGA by AspN + CNBr. Peptide masses (Da) and position numbers are listed along with the number of basic (# HKR) and GlcNAc-modifiable (# ST) residues for each theoretical peptide. The comments section refers to peptides predicted to be detected (+) or not detected (-), with (?) indicating uncertainty of detection.

Peptides DELLAVLGYKVRSSEm and LGWHTRPLITTSAWKLSTAAY contained two and three more basic residues respectively, than their tryptic counterparts, SSEMAVALK and LSTAAY. Peptide LSELNPPPLPASSNGL was a slightly more compact version of its AspN counterpart, having lost 3 residues from its N-terminus. While this improved the overall charge state potential for two out of these three peptides, all three still lacked a basic residue on the C-terminal side of the Ser/Thr residues in question. Electron transfer to the carbonyl group consumes one positive charge from the peptide precursor and a reaction cascade proceeds to break one amide bond per molecule, generating one c and z· ion each. If there is not at least one charge remaining on each fragment ion, it will not be detected by the mass spectrometer. Therefore for all three

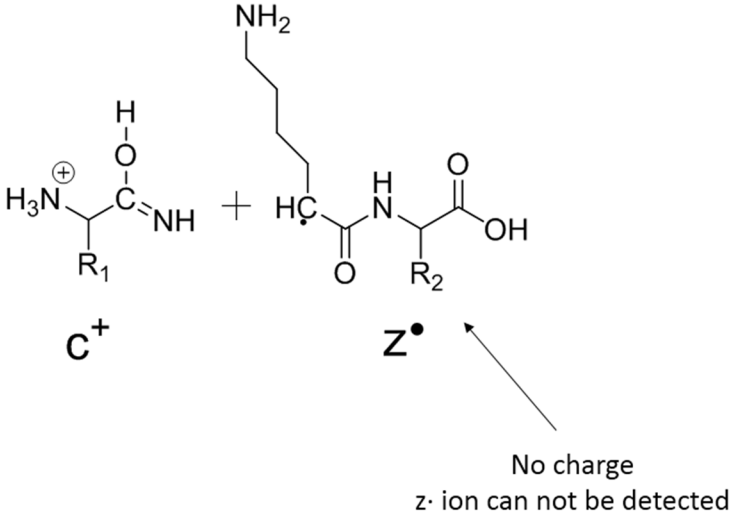


Figure 2.33 – *ETD consumes one charge, requiring a minimum of 3 charges for c and z· ions to both be detected. If no protonation exists C-terminally to the cleavage site, the z ion will not be detected.*

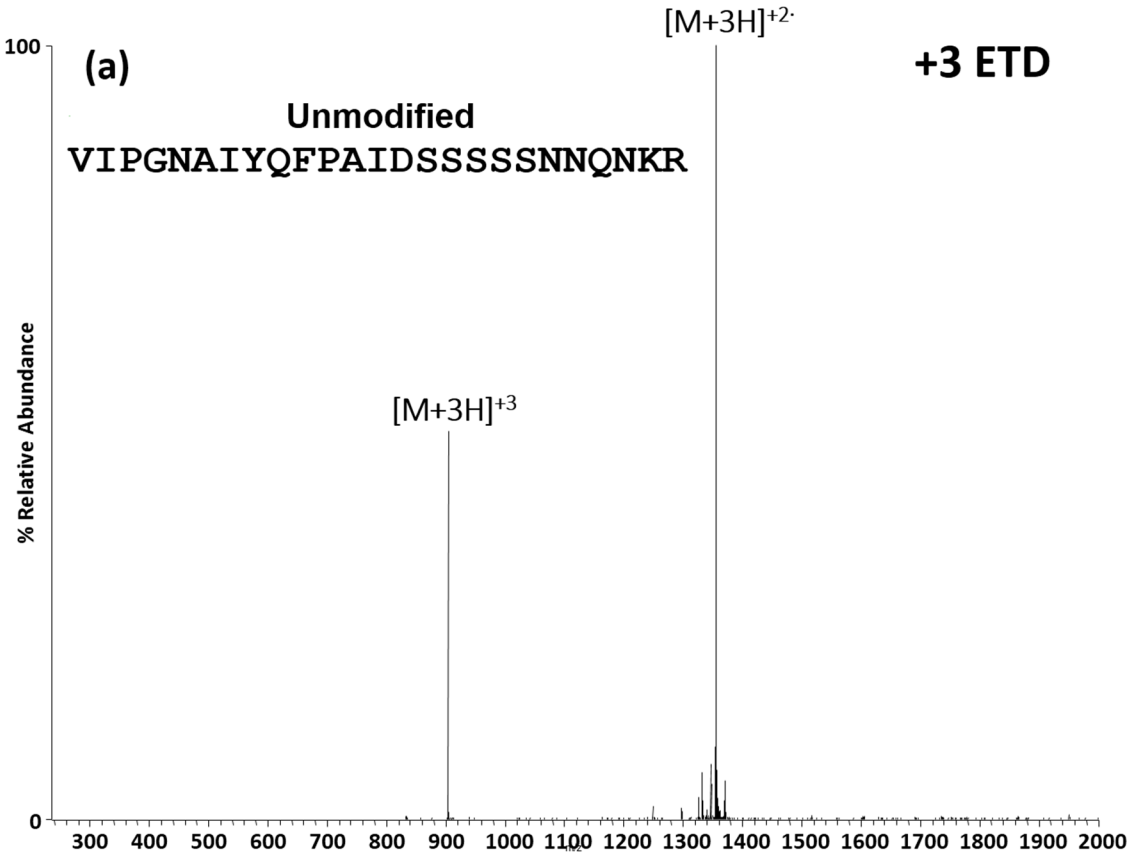
peptides above, it would be impossible to detect any z· ions within the Ser/Thr regions (**Figure 2.33**).

Loss of one half of the ions in this region severely hinders the ability to confidently assign the PTM site(s) and often precludes the ability to do so completely. Therefore, CNBr subdigestion was combined with chemical derivatization to add C-terminal charges to these modified peptides before ETD analysis.

Enhancement of ETD by Derivatization of Carboxylic Acids with Histamine

As previously described in this section, one technique that may be used to add charge is the derivatization of carboxylic acids with histamine [45]. The basic imidazole ring is a strong candidate for protonation, adding charge to the three types of acidic residues: Asp, Glu, and C-termini. Therefore, this method is useful for simultaneously adding charge throughout peptides in addition to adding C-terminal charge. Derivatization reactions were performed on Asp-N generated peptides sub-cleaved with CNBr in the same manner as previously described in this chapter. Tryptic RGA peptides were also derivatized with histamine, especially in the attempt to improve ETD fragmentation of poly-Ser/Thr peptides VIPGNAIYQFPAIDSSSSSNNQNKR and GVIGTTVTTTTTTTTAAGESTR, neither of whose PTMs could be site-mapped.

Following derivatization of Asp and the C-terminus with histamine, VIPGNAIYQFPAIDSSSSSNNQNKR was charge enhanced from its previous *z* = 3 to *z* = 4 or 5. **Figure 2.34** illustrates the improvement in ETD of VIPGNAIYQFPAIDSSSSSNNQNKR before and after full derivatization with histamine.


[M+3H]+2·
100
(a)
+3 ETD
Unmodified
VIPGNAIYQFPAIDSSSSSNNQNKR
[M+3H]+3
% Relative Abundance
0
300 400 500 600 700 800 900 1000 1100 1200 1300 1400 1500 1600 1700 1800 1900 2000

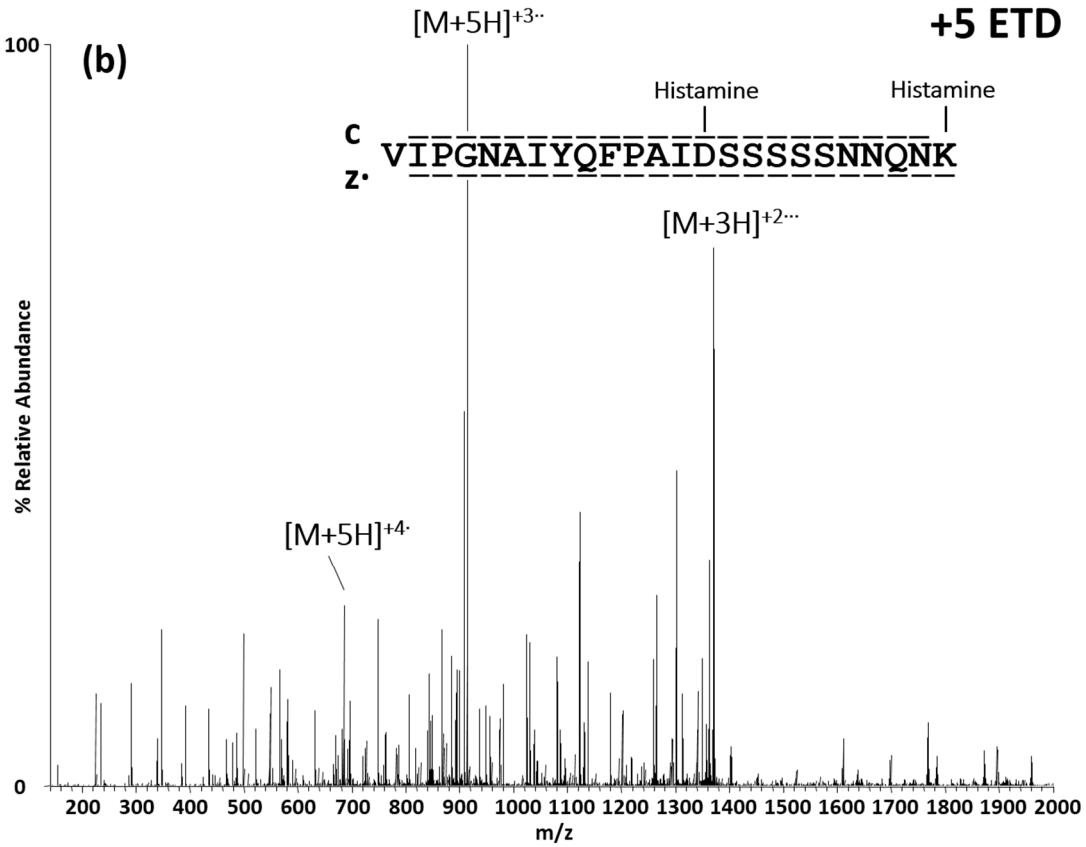


Figure 2.34 – *Derivatization of Asp and the C-terminus of unmodified VIPGNAIYQFPAIDSSSSSNNQNK with histamine increased the charge state from z = 3 (a) to z = 5 (b) and improved sequence coverage from nil to full. Black bars represent observed c and z· fragment ions.*

Unfortunately, the ion current of the fully derivatized, *z* = 5 peptide only accounted for ~10% of its total ion current detected. This was due to incomplete reaction efficiency, side reaction of Asp forming a succinimide (**Figure 2.35**), and ion current splitting evenly between charge states +4 and +5. These ion current divisions combined with the already low abundance of both O-GlcNAc and phosphate PTMs still prevented even the *z* = 5 PTM forms from fragmenting well by ETD. Because GVIGTTVTTTTTTTTAAGESTR similarly contained two derivitazable sites, Glu and the C-terminus, it was also charge enhanced from its predominant *z* = 3 to *z* = 4 or 5 as well. Fortunately, its PTM forms were more abundant than its unmodified form although the ion current was still split amongst 26 distinct combinations of PTMs, plus additional isobaric

combinations having the same m/*z* but different primary locations. The ion current was again additionally evenly divided between *z* = 4 and 5 following ESI. Despite this, tens of fmols of several PTM forms were available for ETD, but it did not improve substantially, and the sites remained virtually unmappable. The charge likely remained too localized near the peptides' C-terminus, as opposed to the N-terminus and center where the majority of modifiable residues were located.

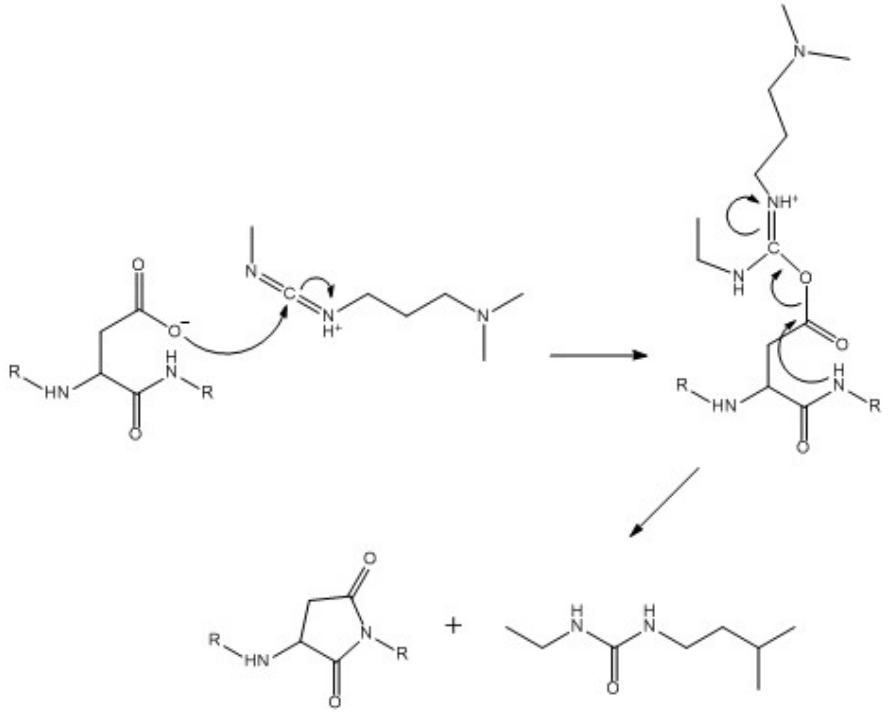
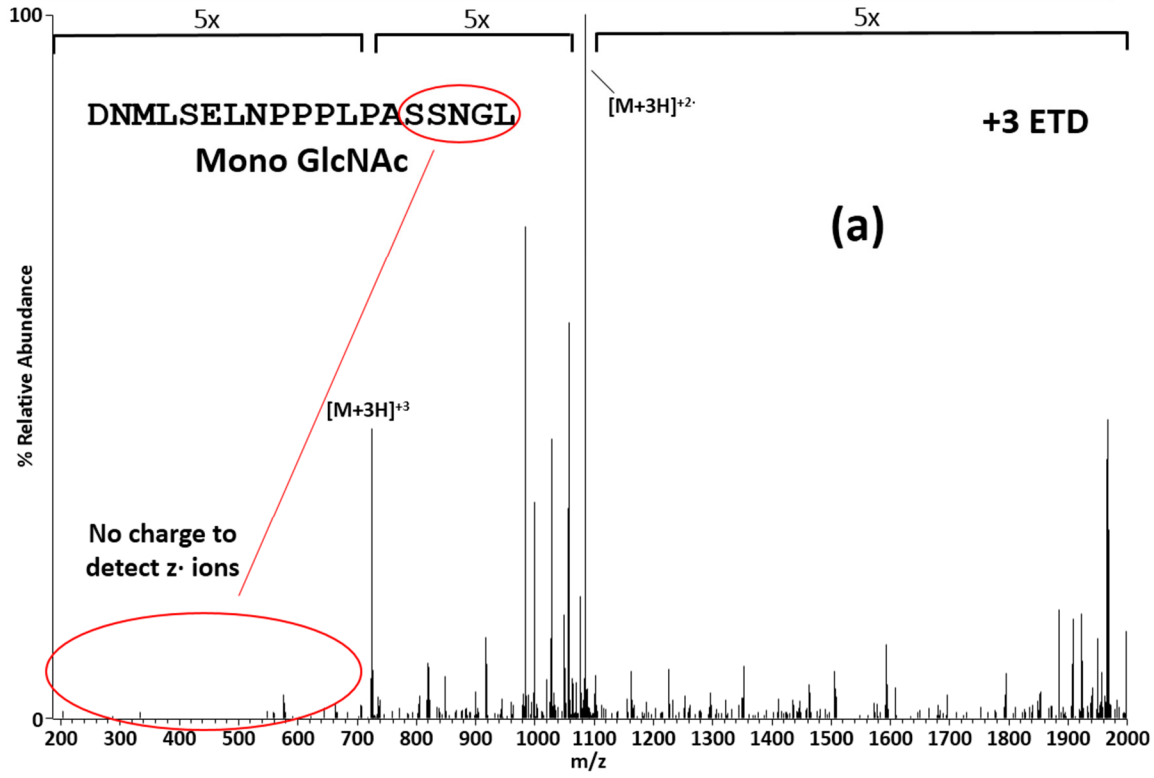
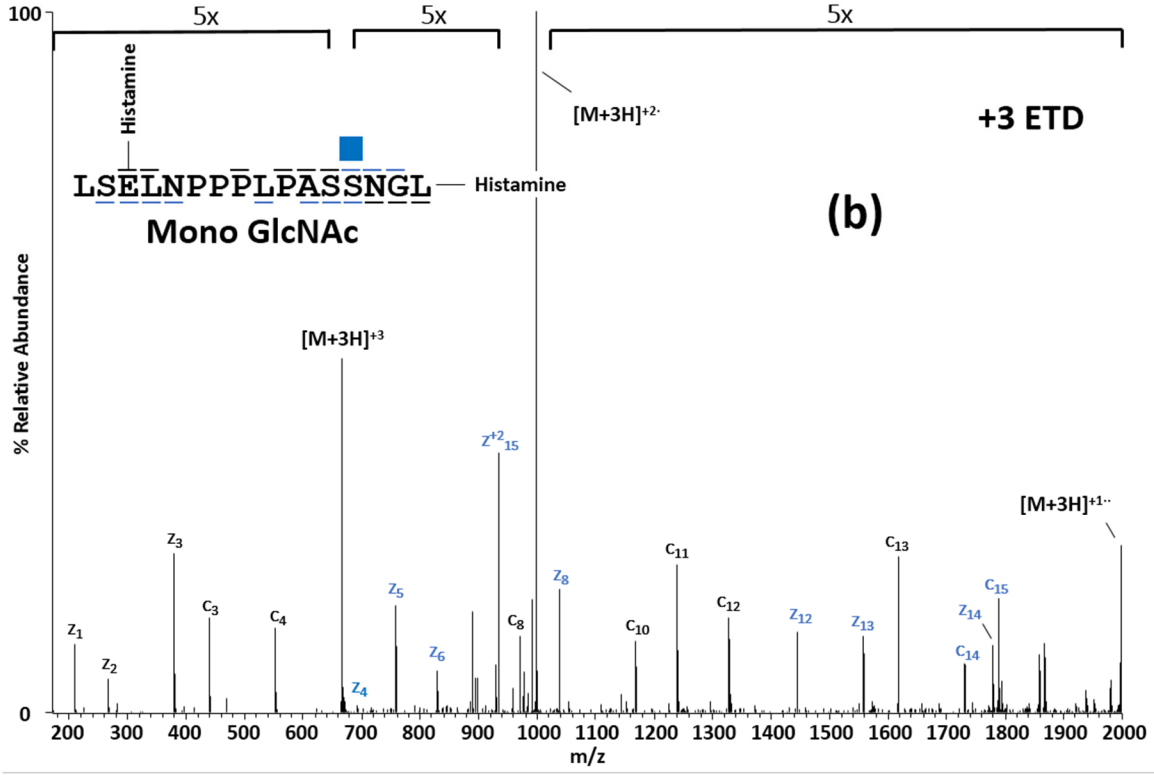


Figure 2.35 – *Reaction scheme of Asp residues forming stable succinimide side-products.*

Histamine derivatization was more successful in improving ETD fragmentation for the peptides cleaved by Asp-N and CNBr. In the case of DELLAVLGYKVRSSEm, which lacks the C-terminal carboxylic acid due to homoserine lactone formation, a carboxylic acid still exists C-terminally to the tandem Ser residues because of the adjacent Glu residue. Asp-N peptide DNMLSELNPPPLPASSNGL was subcleaved with CNBr which had two advantages. The first advantage was to avoid potentially splitting ion current at a site where charge enhancement

was not critical. Aspartic acids have a propensity to form backbone succinimides upon activation by EDC. A neighboring amide may attack the carbonyl carbon of Asp's carboxylic acid, forming the succinimide which is stable at the acidic pH (5.5) used for derivatization [45, 83]. The CNBr-subcleaved peptide, LSELNPPPLPASSNGL, contained an N-terminal primary amine two carboxylic acids, Glu and the C-terminus, with the potential to hold three charges following histamine derivatization. A second advantage was to gain a higher charge density to improve ETD. Derivatization of all three CNBr-subcleaved AspN peptides enabled the successful ETD sequencing of their unmodified forms. Peptide LGWHTRPLITTSAWKLSTAAY was enhanced from z = 5 to z = 6 with a new charge on the C-terminus, but the low abundance (<1%) of the O-GlcNAcylated form still prevented the generation of a quality O-GlcNAcylated ETD spectrum. Despite this, DaELLAVLGYKVRSSaEm (*z* = 4), where "a" precedes residues derivatized with histamine, could now be sequenced by ETD to reveal S56 as the major O-GlcNAcylation and O-hexosylation site in this region with O-GlcNAc being more abundant than O-hexose. A more minor site, S57 also separately contained both PTMs, with O-GlcNAc also in higher abundance than O-hexose. Finally, LSaELNPPPLPASSNGaL (z=3) was now sufficiently charge-enhanced to allow for unambiguous sequencing of four of its five PTM forms (**Figure 2.36**).


100
5x
5x
5x
DNMLSELNPPPLPASSNGL
Mono GlcNAc
[M+3H]+2·
+3 ETD
(a)
% Relative Abundance
[M+3H]+3
No charge to
detect z· ions
0
200 300 400 500 600 700 800 900 1000 1100 1200 1300 1400 1500 1600 1700 1800 1900 2000
m/z


100
5x
5x
5x
Histamine
[M+3H]+2·
+3 ETD
LSELNPPPLPASSNGL
Histamine
Mono GlcNAc
(b)
% Relative Abundance
[M+3H]+3
[M+3H]+1··
0
200 300 400 500 600 700 800 900 1000 1100 1200 1300 1400 1500 1600 1700 1800 1900 2000
m/z

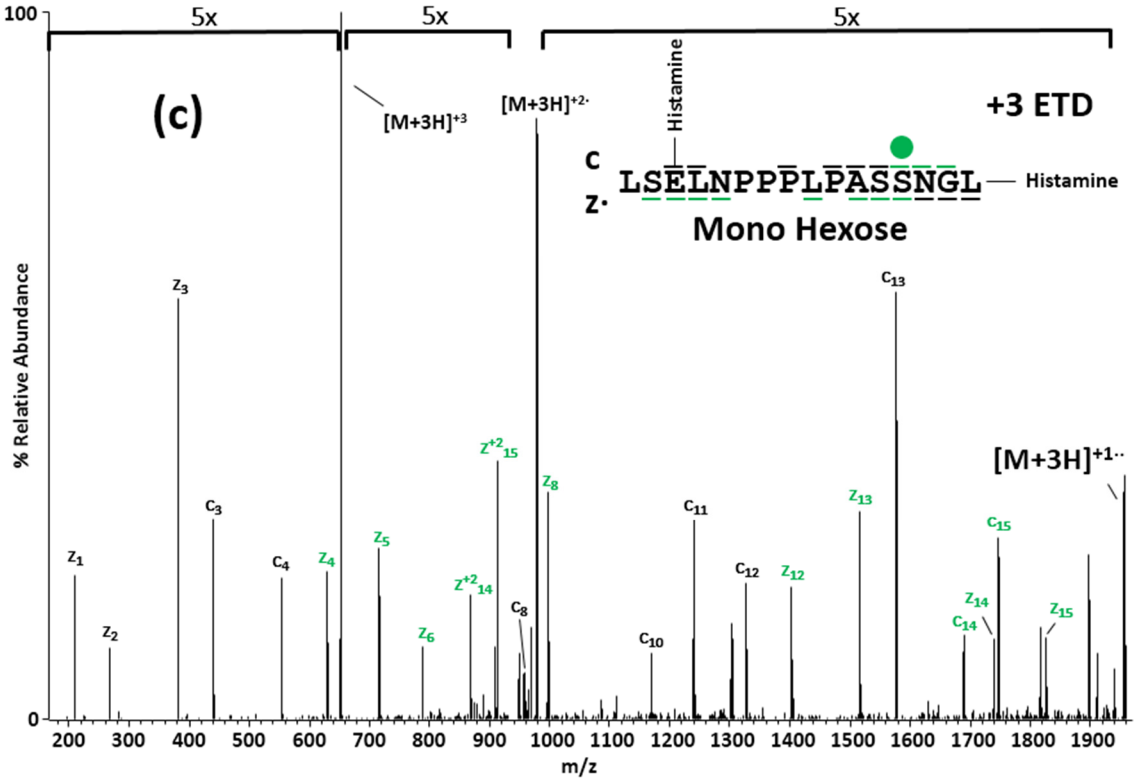


Figure 2.36 *–DNMLSELNPPPLPASSNGL does not efficiently fragment by ETD, and no charge exists on the C-terminus to detect low mass z ions (a) and confidently locate PTMs. Charge density was increased by subcleaving with CNBr and derivatizing carboxylic acids with histamine to place the charge at strategic locations. Mono-O-GlcNAcylated (b) and O-hexosylated (c) ETD spectra of derivatized peptide LSELNPPPLPASSNGL allows unambiguous localization of each PTM.*

Mono-O-GlcNAcylation and O-hexosylation were detected on residue S116. Interestingly, both PTMs were also detected simultaneously in two forms: O-hexose on S115 and O-GlcNAc on S116 was the major form and vice versa was the minor form. Subcleavage of Asp-N-digested peptides with CNBR, combined with derivatization by histamine, enabled the sequencing of 9 distinct RGA PTM peptides forms.

Analysis of RGA(+SEC) PTMs

Overall, the RGA(+SEC) tobacco system produced highly modified RGA protein with respect to the number of sites modified, the number of PTM combinations observed, and the abundance of these PTMs relative to the total abundance of the peptide detected. A total of 13

peptides were detected containing O-GlcNAcylation, phosphorylation, and/or O-hexosylation. These 13 peptides contained a combined total of 89 distinct PTM forms, with the majority due to 3 of the 4 poly-Ser/Thr peptides bearing 30, 12, and 26 PTM forms, respectively (**Table 2.7**).

Peptide	Forms containing:							
	O-GlcNAc		Phosphate		O-Hexose		Total	
	# Forms	Abundance	# Forms	Abundance	# Forms	Abundance	# Forms	Abundance
DHHQFQGRLSNHGTSSSSSSISK	19	+++++	16	++	11	++	30	+++++
DELLAVLGYKVRSSEM	3	++++			2	++	5	++++
DGLHLAT	1	++			1	+	2	++
DTVHYNPSELYSWL	1	+					1	+
DPVLPSPEICGFPAS	1	++	1	+			2	++
LSELNPPPLPASSNGL	4	+++++			3	++++	5	+++++
VIPGNAIYQFPAIDSSSSSNNQNKR	1	++	1	+			2	++
SCSSPDSMVTSTSTGTQIGK	10	+++++	4	+++	2	++	12	+++++
GVIGTTVTTTTTTTTAAGESTR	21	+++++			20	+++++	26	+++++
QIGCLAVSQAGAMR	1	+					1	+
FTESLHYYSTLFDSLEGVPNSQDK	1	+					1	+
FGSSGLAPAHLGSNAFK	1	++					1	++
LSTAAY	1	++					1	++
	Total		Total		Total		Total	
	65		22		39		89	
	73%		25%		44%			

Table 2.7 –Summary of the 13 modified RGA(+SEC) peptides that contain O-GlcNAc, phosphate, and/or O-hexose. The number of unique forms containing each PTM are reported as well as the total abundance of peptide containing each PTM. PTM levels are reported as % total peptide abundance detected, (+++++) = 50+% (++++) = 20-49%, (+++) = 10-19%, (++) = 1-9%, (+) = <1%. % Total abundances were calculated from ion currents observed in the main beam mass spectra [(modified peptide ion current)/ (modified peptide + unmodified peptide ion current) x (100)].

(a)

Peptide	Number of Forms containing:							
	O-GlcNAc	Phosphate	O-Hexose	O-GlcNAc & Phosphate	O-GlcNAc & O-Hexose	Phosphate & O-Hexose	O-GlcNAc, Phosphate, & O-Hexose	Total
DHHQFQGRLSNHGTSSSSSISK	6	5	4	8	4	2	1	30
DELLAVLGYKVRSSEM	3	0	2					5
DGLHLAT	1	0	1					2
DTVHYNPSELYSWL	1	0	0					1
DPVLPSPEICGFPAS	1	1	0					2
LSELNPPPLPASSNGL	2	0	1		2			5
VIPGNAIYQFPAIDSSSSSNNQNKR	1	1	0					2
SCSSPDSMVTSTSTGTQIGK	6	1	1	3	1			12
GVIGTTVTTTTTTTTAAGESTR	6	0	5		15			26
QIGCLAVSQAGAMR	1	0	0					1
FTESLHYYSTLFDSLEGVPNSQDK	1	0	0					1
FGSSGLAPAHLGSNAFK	1	0	0					1
LSTAAY	1	0	0					1
	Total	Total	Total	Total	Total	Total	Total	Total
	31	8	14	11	22	2	1	89
	35%	9%	16%	12%	25%	2%	1%	

(b)

Peptide	Abundance of Forms containing:							
	O-GlcNAc	Phosphate	O-Hexose	O-GlcNAc & Phosphate	O-GlcNAc & O-Hexose	Phosphate & O-Hexose	O-GlcNAc, Phosphate, & O-Hexose	Total
DHHQFQGRLSNHGTSSSSSISK	+++++	+	++	++	++	+	+	+++++
DELLAVLGYKVRSSEM	++++		++					++++
DGLHLAT	++		+					++
DTVHYNPSELYSWL	+							+
DPVLPSPEICGFPAS	++	+						++
LSELNPPPLPASSNGL	+++++		++++		++			+++++
VIPGNAIYQFPAIDSSSSSNNQNKR	++	+						++
SCSSPDSMVTSTSTGTQIGK	++++	++	++	++	++			+++++
GVIGTTVTTTTTTTTAAGESTR	++++		++		+++++			+++++
QIGCLAVSQAGAMR	+							+
FTESLHYYSTLFDSLEGVPNSQDK	+							+
FGSSGLAPAHLGSNAFK	++							++
LSTAAY	+							+

Table 2.8 - Summary of modified RGA(+SEC) peptides expanded to report the number and abundance of peptides containing O-GlcNAc, phosphate, and O-hexose exclusively or in combination

Of the 89 different PTM forms from the 13 peptides containing PTMS, 31 contained only O-GlcNAc, 8 only phosphate, 14 only O-hexose, 11 O-GlcNAc and phosphate, 22 O-GlcNAc and O-hexose, 2 phosphate and O-hexose, and 1 O-GlcNAc, phosphate, and O-hexose (**Table 2.8**).

O-GlcNAcylation was most prevalent, as all 13 peptides found to be post-translationally modified contained 1 or more O-GlcNAcylated forms, with 2 or more O-GlcNAcs detected simultaneously on 5 of them. Of these 5 peptides able to be multiply O-GlcNAcylated, 4 were >50% modified with O-GlcNAcylation. Forms of these same 4 peptides were also detected containing 1 or more O-GlcNAcs simultaneously with phosphorylation and/or O-hexosylation. In all, 23 O-GlcNAcylation sites were detected on RGA(+SEC) and 14 were unambiguously localized. It is likely that additional sites exist on PTM peptides bearing isobaric m/*z* values with those forms already sequenced, with the resulting mixed MS^2 spectra preventing their identification. O-hexosylation was the second most prevalent PTM with 20-50% of LSELNPPPLPASSNGL modified and >50% of GVIGTTVTTTTTTTTAAGESTR modified. The majority of the former's O-hexosylation was detected apart from O-GlcNAc, although the two forms were also present simultaneously in lesser abundance. In the latter, O-hexosylation was also present but the vast majority was detected simultaneously with O-GlcNAcylation in 15 different PTM combinations. A total of 14 confirmed O-hexosylation sites were detected on RGA(+SEC) with 9 of these site-localized. Phosphorylation was the least abundant PTM, yet six sites were still detected on four peptides, and five of these could be site-localized. On two of these, SCSSPDSMVTSTSTGTQIGK and DHHQFQGRLSNHGTSSSSSSISK, phosphorylation was detected on forms bearing and lacking O-GlcNAcylation. Likewise, these two peptides also contained O-hexose forms bearing and lacking O-GlcNAcylation. One peptide, DHHQFQGRLSNHGTSSSSSSISK, was detected containing all PTMs both alone and with all possible PTM combinations, including O-GlcNAc, phosphorylation, and O-hexose all simultaneously present. Notably, it was the only peptide to contain all potential combinations and was the most diversely modified region with 30 total PTM forms. It was O-GlcNAcylated in 19 forms of these forms, phosphorylated in 16 forms, and O-hexosylated in 11 forms detected. This is significant, because it reveals that at least 30

modified forms of RGA(+SEC) exist. In fact, >30 PTM forms of RGA were observed to exist because 3 other peptides were >50% modified with multiple PTM combinations. Multiple PTM forms of domains S13-S25, S167-T181, and T190-T206 must coexist on the same protein molecules.

Site Locations of RGA PTMs

The vast majority of RGA PTMs occur on the N-terminal half of the protein. All poly-Ser/Thr regions are contained in this region as well. Eight distinct regions of RGA were found to be modified by more than one type of PTM, and five of these regions contained multiple simultaneous PTMs of the same and/or different type.

The most diversely modified region is encompassed by residues S13-S25. The majority of this region is detected with an even mix of O-GlcNAcylation on S18 or S20. An abundant di-O-GlcNAcylated form was detected bearing O-GlcNAcs detected most strongly on S18 and S19, with others co-eluting forms that contained O-GlcNAc(s) between T17-S22. One interpretation of this is that O-GlcNAcylation of S18 can direct O-GlcNAcylation of S19 or vice versa. This conclusion must be considered cautiously though, because mixed isobaric spectra limit one's ability to detect PTMs on residues located in between two others found to contain PTMs. They also hurt the limit of PTM detection. Therefore, it is possible that this region was also mono-O-GlcNAcylated at S19 but wasn't detected because it would have been nearly impossible to distinguish whether the modified c_{16} and z_8 fragments ions were a result of the addition of O-GlcNAc on S19, or the mix on S18 and S20, respectively (**Figure 2.19b&c**). The data indicates additional sites were modified in the di-O-GlcNAcylated form, implying that the first O-GlcNAc may help facilitate recruitment of OGT and/or exert control of where it attaches additional O-GlcNAcs. A tri-O-GlcNAcylated form contained all the same potential sites as detected in the di-

O-GlcNAc form, yet it was clearer that T17 and S18 were modified with the first 2 O-GlcNAcs with a mix of the third on S19-S21. The tetra-O-GlcNAcylated peptide was of too little abundance to site-map. Mono-phosphorylation was found largely on S13, but a mix of sites amongst T17-S20 were also detected bearing phosphorylation. Di-phosphorylated forms were again modified on S13 and on a mix of sites amongst T17-S25. Phosphorylation of S20 could be elucidated, because the fragmentation pattern indicated a strong signal at that site in addition to S13. It is probable that these two sites coexist with other sites modified at the expense of S20. The tri-phosphorylated peptide signal was too weak for sequencing. Two forms of mono-O-hexosylation were partially separated chromatographically. The first contained O-hexose on S18, also a heavily modified O-GlcNAc site, and potential phosphorylation site. Separate co-eluting forms were detected in mixed spectra with O-hexose amongst T17-S19. Both S18 and S19 were modified in the di-O-hexosylated form and a tri-O-hexosylated form was detected by accurate mass and elution profile only. With the exception of S13 phosphorylation, the modifications mostly occurred on the N-terminal half of the Ser/Thr-rich domain. This implies they all have potential to compete or act synergistically to influence RGA confirmation, interactions, and overall function. It was useful to also elucidate and evaluate site-locations when these three types of PTMs appeared in tandem. Peptides with mono-O-GlcNAcylation and mono-phosphorylation also contained phosphorylation S13, with O-GlcNAc on T17, S18, and S19. Interestingly, with two O-GlcNAcs present, they were detected on the same residues, but phosphorylation was also detected on T17 in addition to S13. In this case, T17 appears to be a site of competition between phosphorylation and O-GlcNAcylation, potentially acting as a biochemical “switch”. With mono-O-GlcNAcylation and di-phosphorylation, the O-GlcNAc was detected as a mix between T17-S20 with additional sites of phosphorylation. Again, S13, a mix of T17-S20 similar to O-GlcNAc, and also S23. While there was evidence of phosphorylation

throughout the Ser/Thr-rich domain absent of O-GlcNAc, it is possible that the sugar influences phosphorylation at certain sites, such as S23, by helping to more readily direct a kinase there. The addition of a third phosphorylation appears to generate a similar collection of modified sites. When mono-O-GlcNAcylation and O-hexosylation were detected in tandem, each PTM could be found on S18, S19, and S20, so all 3 residues may be sites of competition between the two PTMs. As six theoretical forms containing these two PTMs existed, it was difficult to determine which were present due to the mixed spectra, but it was certain that multiple pairs were present and these may have differing roles for RGA simply based on their location. Mono-O-hexose was also detected with mono-phosphate in the absence of O-GlcNAc. Phosphorylation was detected on residues S13 and T17, consistent with that observed when mono- or di-, and di-O-GlcNAcylation was present, respectively. Interestingly, sites T17, S18, and S19 contained O-hexosylation, the same sites that contained O-GlcNAc when phosphorylation was present. Finally, mono-modified forms containing all 3 PTMs were detected in a mixture with the sites all localized to T17-S19, similar to many other modified forms of this peptides. Notably, phosphorylation on S13, consistently detected in all other phosphorylation forms, was not detected in these versions. It is possible that it was also present, but the low abundance (<1%) of this combination of PTMs affected the ability to detect it, or the presence of both O-GlcNAc and O-hexose simultaneously in this region may have directed phosphorylation only to the Ser/Thr-rich region. In summary, this region was the most dynamically modified by all three PTMs, representing at least 30 post-translationally modified forms of RGA. This bears strong evidence that the dynamic modification of this Ser/Thr-rich region is critical for events that facilitate RGA's function(s) in plants.

Peptide DELLAVLGYKVRSSEM contained mono-O-GlcNAcylation, and to a lesser degree, mono-O-hexosylation on S56. Site S57, the more minor site of modification relative to S56, also contained mono-O-GlcNAcylation and O-hexosylation with the latter again in lesser abundance than the former. Although these PTMs were not detected simultaneously on the Asp-N peptide, a tryptic miscleaved peptide (VRSSEMAEVALK) also containing residues S56 and S57 bore O-GlcNAcylation on both residues simultaneously. There was no obvious sign of constitutive action whereby the first modification dictated the addition of the second. It was notable that on both Ser residues, O-GlcNAcylation was the major PTM and O-hexose was the minor.

The Asp-N-generated peptide, DGLSHLAT, contained mono-PTM versions of O-GlcNAc and O-hexose on S82. Again, both PTMs modified the same site and O-GlcNAc was more abundant.

The majority of residue S116 was observed to be O-GlcNAcylated in peptide LSELNPPPLPASSNGL. A small amount (<1%) of this peptide was also di-O-GlcNAcylated but was unable to be site-mapped because of inadequate levels. Residue S115 was also abundantly mono-O-hexosylated, resulting in the vast majority of this residue being occupied by 1 of these 2 modifications. Because these modifications strongly compete for the same site, residue S116 presents another likely location for a switch-like PTM event where O-GlcNAc facilitates a specific function and O-hexose facilitates a different one. Adding another level of complexity, O-hexose was also detected on S115 in the presence of S116 O-GlcNAcylation, but was not present in its absence. Similarly, a lesser amount of O-GlcNAc was detected on S115 in the presence but not the absence of S116 O-hexosylation. Therefore, modification with O-hexose or O-GlcNAc on S116 appears to direct the post-translational modification of S115. While the lone modification of S116 is in vast abundance, the lone modification of S115 by any PTM was not detected.

Peptide SCSSPDSMTVTSTSTGTQIGK encompassed the Ser/Thr-rich domain N-terminally adjacent to the G185_G186insK lysine insertion. Following the primary sequence from the N-terminus to the C-terminus, it was the second of three such domains that were able to be modified by all three types of PTMs detected in this study. Mono-, di-, and tri-O-GlcNAcylated forms of this peptide were all detected in similar abundance. In addition, these same quantities of O-GlcNAcylation were each detected coinciding with mono-phosphorylation, and mono-O-GlcNAcylation was also detected in the presence of O-hexose. The peptide also bore mono-O-hexose alone. Unfortunately, fragmentation was poor for the majority of these forms, and was not improved by the techniques already discussed. Still, several PTM forms were able to be sequenced. Mono-O-GlcNAc was detected on T176 and also T178. Less abundantly, a mixture of O-GlcNAc sites located amongst T176-S179 were detected. Mono-phosphorylation occurred on S170, several residues upstream from O-GlcNAc. One form of O-GlcNAc and phosphate was determined to bear phosphate again on S170 and O-GlcNAc on T177. It was difficult to determine whether T177 was O-GlcNAcylated apart from S170 phosphorylation, as evidence for that mark was detected in mixed spectra. Given the site-location data known from this peptide, the post-translational modification of this region may be a case in which phosphorylation and O-GlcNAcylation act synergistically rather than antagonistically, because they do not compete for the same Ser/Thr residue. There are quite possibly several RGA functions influenced by the post-translational modification of this domain, because like the first Ser/Thr-rich domain of RGA, it was diversely modified at >50% level.

Immediately C-terminal to residues 167-185 is the fourth and final poly-Ser/Thr RGA domain, encompassed by peptide GVIGTTVTTTTTTTTAAGESTR (residues 186-207). It was the most abundantly (>90%) and second most diversely (26 PTM forms) modified peptide detected

from RGA(+SEC). Both O-GlcNAcylation and O-hexosylation were detected. The peptide was too charge deficient to fragment well by ETD, so spectra were of insufficient quality to site-localize the PTMs. However, because such a large number of multiply modified forms were detected, it was possible to determine that at least six O-GlcNAc and five O-hexose sites existed on this peptide. O-GlcNAcylation was present in greater abundance than O-hexosylation overall (**Figure 2.37**).

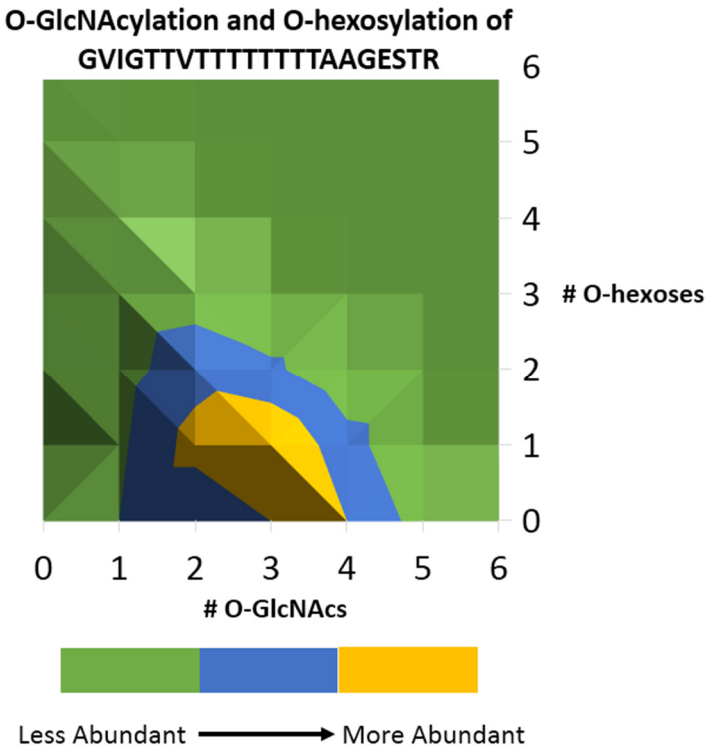


Figure 2.37 – *Heat map reporting abundances of the 26 post-translationally modified forms of GVIGTTVTTTTTTTTAAGESTR.*

Peptides containing only O-GlcNAcylation were more abundant than those containing only O-hexosylation as approximately one third of the region was modified in the former and <10% in the latter. Approximately two thirds of RGA contained combinations of both types of PTMs on this domain. Still, the abundance was skewed towards those forms containing more O-GlcNAcs relative to O-hexoses. The most abundant forms of the peptide contained 3-4 –GlcNAcs with 0-1 O-hexoses. Based on the observation of other abundantly and diversely modified regions of

RGA(+SEC), it is probable that many of the PTM combinations detected on residues 186-207 exist with multiple isobaric site-location patterns. Therefore, it is reasonable to hypothesize that many more than 26 actual PTM forms of this peptide, and thus RGA, exist.

In summary, RGA was abundantly and diversely modified in this system by three separate PTMs, and one of these was a previously unknown type of nuclear protein PTM. It is already known that over 30 forms of RGA exist in this system due to its post-translational modification. Also, the 89 PTM forms summed from all the modified peptides are present in various combinations with each other to form numerous macromolecular PTM combinations. The theoretical number of PTM combinations possible are very great (>10,000), so a conservative estimate for the actual number of unique RGA molecular resulting from its post-translational modification by O-GlcNAc, phosphate, and/or O-hexose, is easily in the hundreds. This data suggests that RGA function may be intricately regulated by key PTM events, and that the enzymes responsible for catalyzing their addition and removal work together in a complex interplay. The question remains as to what is responsible for the presence of the O-hexose modification. One hypothesis was that SEC could have unknown dual activity whereby it catalyzed the addition of O-hexose from UDP-hexose substrate, analogous to O-GlcNAc from UDP-GlcNAc. The fact that all hexose sites mapped were also sites of O-GlcNAcylation support this possibility as the enzyme appears to have very similar site specificity to SEC. An alternate expression of RGA was conducted and purified to explore this possibility.

2.3.6 The effect of SEC on O-hexosylation

To investigate the potential role SEC plays in RGA O-hexosylation, RGA-G185_G186insK was expressed in tobacco without SEC overexpression, and the samples were analyzed by mass

spectrometry. All three types of PTMs observed in RGA(+SEC) were also present in RGA(-SEC), but each to a lesser overall degree in the latter (**Tables 2.9 & 2.10**).

Peptide sequence	# Mods			Sites	z	PPM	Relative Abundance	Digest	
	GlcNAc	Phospho	Hexose					Trypsin	AspN
DHHQFQGRLSNHGT**gS**SSSSSISK	1			S18	5	1.0	++		X
DHHQFQGRL**gg(**SNHGTSSSSSSIS**)**K	2			[bc]	5	1.1	+		X
DHHQFQGRL**p[**SNHGTSSSSSSIS**]**K		1		[d]	4	0.7	++		X
DHHQFQGRLSNHG**p[**TS**pSpS**SS**pS]**ISK		2		S19, S20, S23, & [][d]	4	-0.1	+		X
DHHQFQGRLSNHGT**hS**SSSSSISK			1	S18	5	0.2	+		X
DELLAVLGYKVR**g(**SS**)**EMA	1			[bc]	3	0.3	++		X
DNML**g(**SELNPPPLPASS**)**NGL	1			[c]	3	1.0	+++		X
DNML**h(**SELNPPPLPASS**)**NGL			1	[bc]	3	3.3	++		X
GVIGTTV**g[gT**TTT**]**TTTTAAGESTR	1			T193 & [][d]	3	-0.1	++++	X	
GVIGTTV**gg(**TTTTTTTT**)**AAGESTR	2			[c]	3	-2.1	++	X	

Table 2.9 - Summary of the number of modified residues and site-locations of O-GlcNAc, phosphate, and/or O-hexose PTMs on RGA(-SEC) peptides. Data is reported as in Table 2.4.

(a)

Peptide	Number of Forms Containing: O-GlcNAc	Phosphate	O-Hexose	Total
DHHQFQGRLSNHGTSSSSSSISK	2	3	1	6
DELLAVLGYKVRSSEMA	1			1
DNMLSELNPPPLPASSNGL	1		1	2
GVIGTTVTTTTTTTTAAGESTR	2			2
	Total	Total	Total	Total
	6	3	2	11
	55%	27%	18%	

(b)

Peptide	Abundance of Forms Containing: O-GlcNAc	Phosphate	O-Hexose	Total
DHHQFQGRLSNHGTSSSSSSISK	++	++	+	++
DELLAVLGYKVRSSEMA	++			++
DNMLSELNPPPLPASSNGL	+++		++	++++
GVIGTTVTTTTTTTTAAGESTR	++++			++++

Table 2.10 - Summary of modified RGA(-SEC) peptides reporting the number and abundance of peptides containing O-GlcNAc, phosphate, and O-hexose exclusively or in combination.

O-GlcNAcylation detected in this sample resulted from tobacco OGT activity. Notably, there was a decrease in O-hexosylation that was comparable to the decrease in O-GlcNAcylation, implying that SEC does in fact catalyze O-hexose addition. However, this conclusion must be taken with caution as well, because if these PTMs do form a complex interplay with each other and their respective catalyzing enzymes, SEC down-regulation may still have been indirectly suppressing the function of an independent O-hexose transferase. For example, phosphorylation was also

decreased in overall abundance primarily due to the lack of any phosphorylation signal detected on peptide SCSSPDSMVTSTSTGTQIGK, so a change in SEC expression appeared to alter kinase activity and specificity.

A total of four peptides, all modified in RGA(+SEC), were detected with PTMs in the RGA(-SEC) sample. Three of these four were the most abundantly modified peptides in RGA(+SEC) and all four were in the top five most abundantly modified as well. The number of RGA(-SEC) PTMs discovered totaled 11 vs. 43 found on RGA(+SEC), and these consisted of 6 O-GlcNAcs, 3 phosphates, and 2 O-hexoses. Of these, 6 were able to be localized (**Table 2.11**).

PTM Site	O-GlcNAc	Phosphate	O-Hexose
S18	X		X
S19		X	
S20		X	
S23		X	
T193	X		
	Total	Total	Total
	2	3	1

Table 2.11 - PTM Sites unambiguously localized on RGA(-SEC).

All PTM sites were located in regions previously detected to be modified although two PTM sites not localized in RGA(+SEC), S19 phosphorylation and T193 O-GlcNAcylation, were able to be unambiguously site-mapped in RGA(-SEC). Multiple PTMs were detected on three of the four peptides, but not to the same scope as RGA(+SEC) and PTMs of different types were not simultaneously present on the same molecules as they were with the SEC overexpression.

Residues 4-26 only contained 6 detected PTM forms vs. 30 previously. The overall abundance of O-GlcNAc and O-hexose were also reduced by an order of magnitude. RGA(-SEC)

contained S18 mono-O-GlcNAcylation, 1 of 2 most abundant O-GlcNAc sites previously detected. A di-O-GlcNAcylated form at <1% was also detected. Mono-O-hexosylation was also again localized to S18 as before. Mono-phosphorylation was detected in similar abundance to S18 O-GlcNAcylation, although MS/MS spectra revealed a mix of many phosphorylation sites that ranged the entire poly-Ser/Thr domain. It is possible that the decrease in O-GlcNAcylation allowed kinases to adopt less stringent specificity in modifying this domain. Spectra resulting from di-phosphorylation signal also revealed a mix of sites. Three of these phosphorylation sites were able to be localized: S19, S20, and S23.

Mono-O-GlcNAcylation was detected on peptide DELLAVLGYKVRSSEMA, formed from conventional AspN cleavage N-terminally to D44 and an N-terminal E61 cleavage. Asp-N is known to bear some activity towards Glu residues albeit it is generally lower than for Asp [84]. The two modifiable residues contained in this peptide, S56 and S57, were both modified with O-GlcNAc and O-hexose in RGA(+SEC). RGA(-SEC) S56/57 O-GlcNAcylation abundance was decreased relative to RGA(+SEC) and O-hexosylation was undetected.

AspN peptide DNMLSELNPPPLPASSNGL containing S105, S115, and S116 was mono-O-GlcNAcylated and O-hexosylated to lesser degrees in RGA(-SEC) than RGA(+SEC). Also, these PTMs were not simultaneously detected in the former as they were in the latter.

Lastly, the Thr-rich peptide, GVIGTTVTTTTTTTTAAGESTR, was found to contain mono- and di-O-GlcNAcylation. While the O-GlcNAc abundance as a whole was reduced on this domain, the level of mono- and di-O-GlcNAcylation was increased on RGA(-SEC) relative to RGA(+SEC). O-GlcNAcylation was not detected in a series of up to 6 simultaneous modifications like in RGA(+SEC). Also, O-hexose was not detected on this peptide in RGA(-SEC) at all, contrasting with RGA(+SEC) in which >50% of the region contained a combination of O-GlcNAc

and O-hexose PTMs present together. Instead, in RGA(-SEC), OGT catalyzed the addition of 1 or 2 O-GlcNAcs on 1 quarter of RGA's poly-Thr domain, and successive additions either did not occur or were rapidly removed by O-GlcNAcase. ETD spectral quality of mono-O-GlcNAcylated GVIGTTVTTTTTTTTAAGESTR moderately improved for RGA(-SEC), presumably because this PTM form was present in higher abundance than any single PTM form detected in RGA(+SEC). As a result, O-GlcNAc was able to be localized to one site: T193, and evidence demonstrated these spectra were mixed, so other sites existed on the poly-Thr domain as well (**Figure 2.38**).

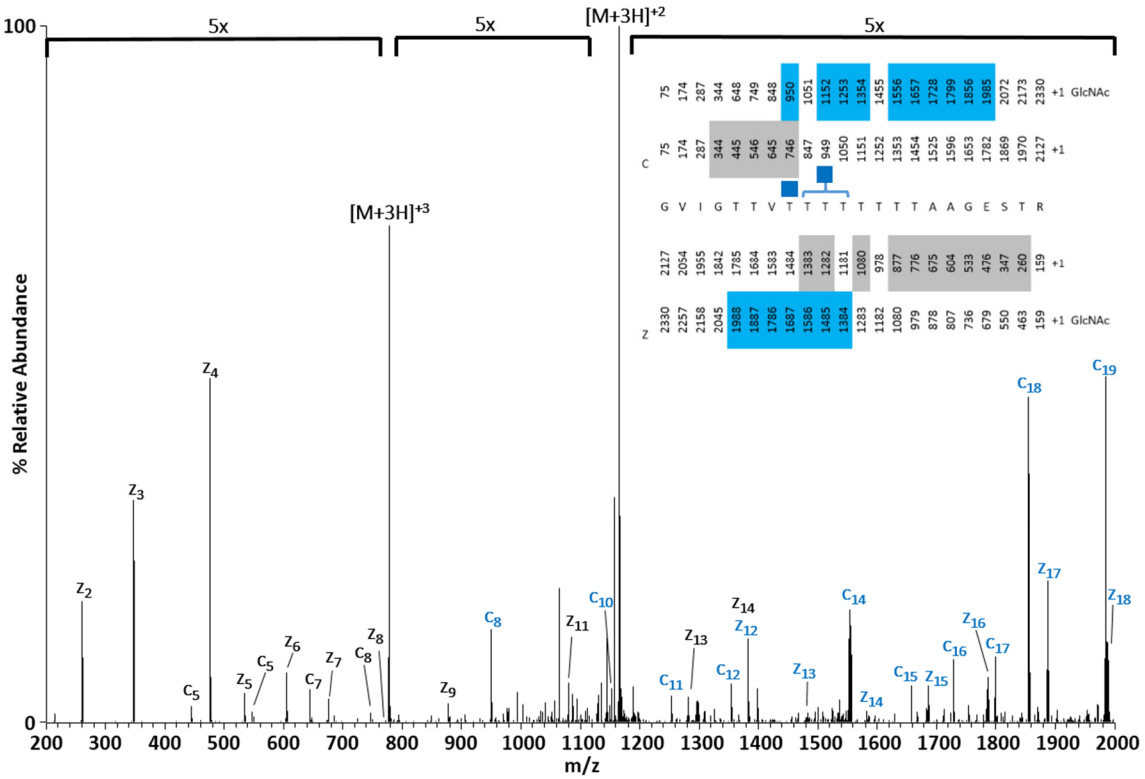


Figure 2.38 – *Mixed ETD spectrum of mono-O-GlcNAcylated GVIGTTVTTTTTTTTAAGESTR in RGA(-SEC). A GlcNAc was localized to T193 and one or more were detected between T194-T196.*

Summary and Future Analysis related to O-hexose

This analysis of RGA(-SEC) revealed that O-GlcNAcylation, O-hexosylation, and phosphorylation are all part of the repertoire of RGA PTMs, whether in the presence or the absence of overexpressed OGT, SEC. By comparing the locations and scale of RGA(-SEC) post-

translational modification to RGA(+SEC), we find that OGT expression level has a direct effect on the levels and scope of O-GlcNAcylation, phosphorylation, and O-hexosylation. Given that the relative decrease in O-hexose levels tracked closely with those of O-GlcNAc, it is clear that SEC has either an indirect if not direct effect on O-hexosylation. Indirectly, O-GlcNAcylation may be a positive regulator of one or more O-hexose transferases, or a negative regulator of one or more glycases responsible for O-hexose removal. Alternatively, this data bears strong evidence in support of SEC having a more direct role in O-hexosylation. The level of O-hexose decreased similarly to that of O-GlcNAc and O-hexose was only found in regions also found to be O-GlcNAcylated. In addition, every O-hexose site localized (whether subjected to SEC overexpression or not) was also determined to be an O-GlcNAc site. This strongly implies that SEC catalyzes the addition of O-hexose, or that it first modifies RGA with O-GlcNAc, and N-acetyl groups are subsequently removed from a portion of the O-GlcNAc moieties. All three of the cases above add an additional layer of complexity to the current understanding of what PTMs occur on nuclear and cytoplasmic proteins, and how those PTMs alter the confirmation, function, localization, etc. of those proteins.

Future studies will need to focus on more fully characterizing RGA O-hexose in 3 ways: (1) determining the exact type(s) of hexose sugar, (2) determining what enzyme(s) and processes are responsible for its addition, and (3) probing its effects on RGA function in the context of RGA O-GlcNAcylation and phosphorylation. The scope of the third question is especially vast. To answer the first question, enzymatic glyco-labeling experiments may be performed in which a known control hexose, such as glucose, is labeled with a radioisotope or fluorescent sugar by an appropriate glycosyltransferase, allowing detection of a successful transfer to the acceptor hexose. If RGA hexose responded accordingly, controlling for false

positive response from O-GlcNAc, one may claim identification of the hexose sugar type. Mass spectrometric analysis would provide a direct and an even more accurate assessment of the results and would allow one to determine the exact O-hexose moieties modified (allowing distinction between two or more types of O-hexoses if they existed). For example, galactosyltransferase, Gal-T6, transfers galactose to glucose on glucosylceramide to synthesize lactosylceramide [85]. Gal-T6 could be used to probe for glucose on RGA in this manner. If the hexose moiety was glucose, it would be an important insight, but would still leave the question of OGT's role unanswered. In an attempt to probe that question, both plant OGTs may be silenced to determine if O-hexose is still present on RGA. This method may be difficult to interpret if O-hexose were subsequently undetected though, because this RGA study has already shown that OGT expression has a drastic effect on the level of O-hexosylation, so its disappearance after OGT silencing could still be an indirect affect due to the lack of O-GlcNAc. Also, elimination of both OGTs has previously been demonstrated to have serious negative effects on plant development and vitality [39], so this could be unviable in practice. Much further exploration of this previously unreported nuclear O-hexose PTM will be required to gain a holistic characterization of its function on RGA, and its potential abundance and function in other proteins from plants or other kingdoms.

2.3.7 Phosphorylation of RGA by CK1

Multiple phosphorylation sites detected in RGA(+SEC) fit the casein kinase 1 (CK1) motifs, Ser-X-X-**Ser/Thr** or Ser/Thr-X-X-X-**Ser**. To further investigate the existence of phosphorylation and probe its effect on overall RGA post-translational modification, a third RGA sample was overexpressed with CK1 in tobacco. Samples digested with Asp-N or trypsin were first screened, then enriched for phosphorylation by IMAC as described previously. In all, 24

distinct PTM sites were detected on RGA(+CK1) consisting of 8 O-GlcNAcs, 12 phosphates, and 4 O-hexoses (**Tables 2.12&2.13**).

Residues	Peptide sequence	# Mods			Sites	z	PPM	Relative Abundance	Digest		IMAC
		GlcNAc	Phospho	Hexose					Trypsin	AspN	
4-26	DHHQFQGRLSNHG**g[**T**g**S**S**g**S**S**]**SSISK	1			S18, S20, & []^{d}	5	-1.0	++		X	
4-26	DHHQFQGRL**gg(**SNHGTSSSSSSIS**)**K	2			^{bc}	5	-1.9	+		X	
4-26	DHHQFQGRLSNHGTSSSSS**pS**ISK		1		S23	5	-1.0	++		X	
4-26	DHHQFQGRL**p[**SNHGTSSSSSSIS**]**K		1		^{d}	5	-1.9	+		X	X
4-26	DHHQFQGRLSNHGTS**pSpS**SSSISK		1		S19 & S20	5	2.9	++		X	X
4-26	DHHQFQGRLSNHGT**pSpS**SS**p[**SSIS**]**K		2		S18,S19 & []^{d}	5	0.3	+		X	X
4-26	DHHQFQGRLSNHGTS**pSpS**S**p[**SSIS**]**K		2		S19,S20 & []^{d}	5	1.4	++		X	X
4-26	DHHQFQGRL**ppp[**SNHGTSSSSSSIS**]**K		3		^{d}	5	-0.5	+		X	X
4-26	DHHQFQGRL**pppp(**SNHGTSSSSSSIS**)**K		4		^{c}	4	0.5	+		X	X
4-26	DHHQFQGRLSNHGT**hS**SSSSSISK			1	S18	5	-1.5	+		X	
44-60	DELLAVLGYKVR**g(**SS**)**EMA	1			^{bc}	3	-1.3	+		X	
44-60	DELLAVLGYKVR**pS**SEMA		1		S56	3	3.5	+		X	X
101-119	DNMLSELNPPPLPA**g(**SS**)**NGL	1			^{c}	2	0.5	++++		X	
101-119	DNML**p(**SELNPPPLPASS**)**NGL		1		^{bc}	2	-6.0	++		X	
101-119	DNML**h(**SELNPPPLPASS**)**NGL			1	^{bc}	2	-0.5	++		X	
140-164	VIPGNAIYQFPAID**p(**SSS**)**SSNNQNKR		1		^{c}	4	1.6	++	X		X
167-185	**g(**SCSSPDSMVTSTSTGT**)**QIGK	1			^{bc}	2	2.3	+	X		
167-185	**p(**SCSSPDSMVT**)**STSTGTQIGK		1		^{c}	3	0.2	++	X		X
186-207	GVIGTTV**gT**TTTTTTTAAGESTR	1			T193	3	5.4	+++	X		
186-207	GVIGTTV**g[**TTTTT**]**TTTAAGESTR	1			^{d}	4^{g}	1.2	++++^{g}	X		
186-207	GVIG**h(**TTVTTTTTTTTAAGEST**)**R			1	^{bc}	3	2.7	+	X		

Table 2.12 - Summary of the number of modified residues and site-locations of O-GlcNAc, phosphate, and/or O-hexose PTMs on RGA(+CK1). Data is reported as in Table 2.4.

PTM Forms of RGA(CK1)

(a)

Peptide	Number of Forms Containing: O-GlcNAc	Phosphate	O-Hexose	Total
DHHQFQGRLSNHGTSSSSSSISK	3	8	1	12
DELLAVLGYKVRSSEMA	1	1		2
DNMLSELNPPPLPASSNGL	1	1	1	3
VIPGNAIYQFPAIDSSSSSNNQNKR		1		1
SCSSPDSMVTSTSTGTQIGK	1	1		2
GVIGTTVTTTTTTTTAAGESTR	2		1	3
	Total	Total	Total	Total
	8	12	4	24
	33%	50%	17%	

PTM Forms of RGA(CK1)

(b)

Peptide	Abundance of Forms Containing: O-GlcNAc	Phosphate	O-Hexose	Total
DHHQFQGRLSNHGTSSSSSSISK	++	+++	+	+++
DELLAVLGYKVRSSEMA	+	+		+
DNMLSELNPPPLPASSNGL	++++	++	++	++++
VIPGNAIYQFPAIDSSSSSNNQNKR		++		++
SCSSPDSMVTSTSTGTQIGK	+	++		++
GVIGTTVTTTTTTTTAAGESTR	+++			+++

Table 2.13 - Summary of modified RGA(+CK1) peptides reporting the number and abundance of peptides containing O-GlcNAc, phosphate, and O-hexose. No combinations of different PTM types were detected.

As in RGA(-SEC), multiply modified peptides were detected, but they only contained PTMs of the same type. The number and overall abundance of O-GlcNAcs and O-hexoses decreased relative to RGA(+SEC). However, RGA(+CK1) phosphorylation increased relative to RGA(+SEC) on 4

peptides, and the number of phosphorylation sites detected more than doubled to 12. Of these, 9 PTMs were unambiguously localized (**Table 2.14**).

PTM Site	O-GlcNAc	Phosphate	O-Hexose
S18	X	X	X
S19		X	
S20	X	X	
S23		X	
S56		X	
T193	X		
	Total	Total	Total
	3	5	1

Table 2.14 - PTM Sites unambiguously localized on RGA(+CK1).

Two phosphorylation sites undetected in RGA(+SEC) and RGA(-SEC) were found in RGA(CK1). A small amount was detected on S18, a minor site of O-hexosylation and major site of O-GlcNAcylation in RGA(+SEC). In addition, S56, also modifiable with O-GlcNAc and O-hexose, was found to contain phosphorylation.

Peptide DHHQFQGRLSNHGTSSSSSSISK was detected with 12 PTM forms, 8 of which were phosphorylated. Evidence pointed to a wide mix of phosphorylation sites with much of the ETD spectra mixed. Also, a tetra-phosphorylated form was detected in RGA(CK1), whereas only up to tri-phosphorylation had been previously detected. Site S18 was only unambiguously detected in a di-phosphorylated form. This may indicate that S19 or a site near the C-terminus of the peptide facilitates the addition of phosphorylation to S18. This cannot be certain though, because a wide mix of mono-phosphorylation sites were also detected and could include site S18. The abundance of phosphorylation in this Poly-Ser/Thr region increased over that of

RGA(+SEC) overall. Also, a tetra-phosphorylated form was detected, when only up tri-phosphorylation was previously detected in this region.

A small amount of phosphorylation was detected on DELLAVLGYKVRSSEMA at S56, a peptide previously only detected bearing O-GlcNAcylation and O-hexosylation. Phosphorylation signal was uniquely detected on RGA(CK1) peptide DNMLSELNPPPLPASSNGL as well, a region already detected as O-GlcNAcylated and O-hexosylated. This furthers the hypothesis that these three PTM types and their corresponding enzymes all have an interplay in regulating RGA.

CK1 overexpression had an interesting effect on the poly-Ser/Thr region contained in peptide SCSSPDSMVTSTSTGTQIGK. In RGA(-SEC), phosphorylation levels were undetectable, but RGA(+CK1) restored them to levels comparable with RGA(+SEC). This indicates that SEC and CK1 act in a synergetic manner to increase phosphorylation of this region.

Overexpression of RGA along with CK1 allowed the detection of two additional phosphorylated peptides, and two additional phosphorylation sites compared to RGA(+SEC). It also clearly demonstrated that CK1 is an enzyme responsible for modifying RGA in at least four different regions of the protein.

2.4 Conclusions

Comprehensively, 92 distinct forms of modified peptides were detected on RGA, comprised of at least 46 PTM sites. In all, 32 of these modifications were able to be unambiguously localized (**Table 2.15**).

Comprehensive Site-Localized RGA PTMs			
PTM Site	O-GlcNAc	Phosphate	O-Hexose
S13		X	
T17	X	X	X
S18	X	X	X
S19	X	X	X
S20	X	X	X
S23		X	
S56	X	X	X
S57	X		X
S82	X		X
S115	X		X
S116	X		X
S170		X	
T176	X		
T177	X		
T178	X		
T193	X		
S253	X		
S537	X		
	Total	Total	Total
	15	8	9

Table 2.15 - Comprehensive list of PTMs localized between RGA(+SEC), RGA(-SEC), and RGA(+CK1)

Of the 15 O-GlcNAc sites identified on RGA, 9 were also sites of O-hexosylation. Of these, 5 were also sites of phosphorylation. Interestingly, all sites that were able to be modified by O-GlcNAc and phosphate were also O-hexose sites. Over 30 post-translationally modified forms of RGA may exist in a single plant, and these data suggest an interplay amongst all three of these modifications on RGA is likely critical to its proper function as a negative regulator of plant development (**Figures 2.39 & 2.40**).

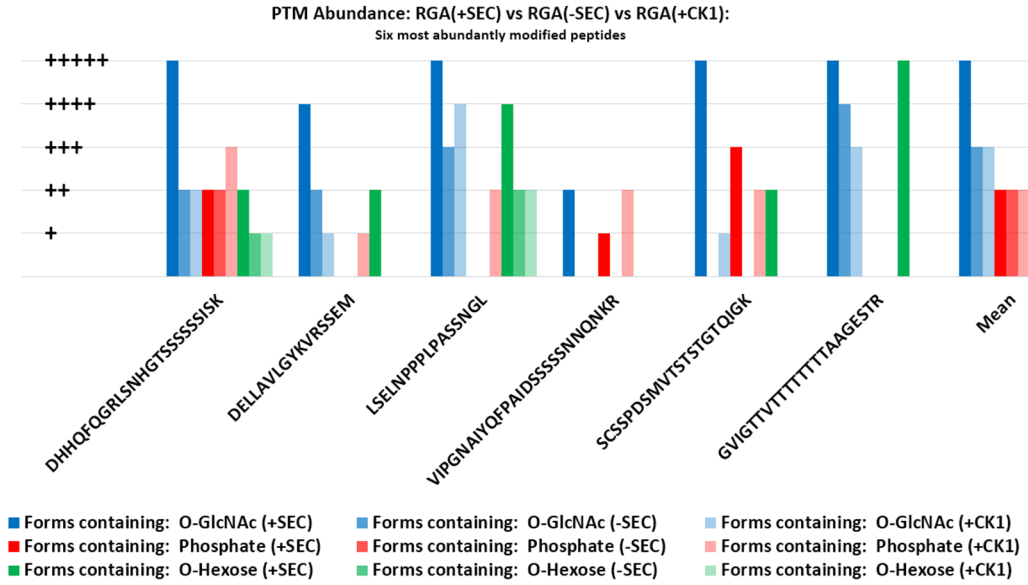


Figure 2.39 – *Sum total of O-GlcNAc, phosphate, and O-hexose in RGA(+SEC,-SEC,&+CK1) for the six most abundantly modified RGA peptides.*

(a) RGA(+SEC)

```
        10         20         30         40         50         60
MKRDHHQFQG RLSNHGTSSS SSSISKDKMM MVKKEEDGGG NMDDELLAVL GYKVRSSEMA

        70         80         90        100        110        120
EVALKLEQLE TMMSNVQEDG LSHLATDTVH YNPSELYSWL DNMLSELNPP PLPASSNGLD

       130        140        150        160        170        180
PVLPSPEICG FPASDYDLKV IPGNAIYQFP AIDSSSSSNN QNKRLKSCSS PDSMVTSTST

       190        200        210        220        230        240
GTQIGGVIGT TVTTTTTTTT AAGESTRSVI LVDSQENGVR LVHALMACAE AIQQNNLTLA

       250        260        270        280        290        300
EALVKQIGCL AVSQAGAMRK VATYFAEALA RRIYRLSPPQ NQIDHCLSDT LQMHFYETCP

       310        320        330        340        350        360
YLKFAHFTAN QAILEAFEGK KRVHVIDFSM NQGLQWPALM QALALREGGP PTFRLTGIGP

       370        380        390        400        410        420
PAPDNSDHLH EVGCKLAQLA EAIHVEFEYR GFVANSLADL DASMLELRPS DTEAVAVNSV

       430        440        450        460        470        480
FELHKLLGRP GGIEKVLGVV KQIKPVIFTV VEQESNHNGP VFLDRFTESL HYYSTLFDSL

       490        500        510        520        530        540
EGVPNSQDKV MSEVYLGKQI CNLVACEGPD RVERHETLSQ WGNRFGSSGL APAHLGSNAF

       550        560        570        580
KQASMLLSVF NSGQGYRVEE SNGCLMLGWH TRPLITTSAW KLSTAAY
```

(b) RGA (+CK1)

```
        10         20         30         40         50         60
MKRDHHQFQG RLSNHGTSSS SSSISKDKMM MVKKEEDGGG NMDDELLAVL GYKVRSSEMA

        70         80         90        100        110        120
EVALKLEQLE TMMSNVQEDG LSHLATDTVH YNPSELYSWL DNMLSELNPP PLPASSNGLD

       130        140        150        160        170        180
PVLPSPEICG FPASDYDLKV IPGNAIYQFP AIDSSSSSNN QNKRLKSCSS PDSMVTSTST

       190        200        210        220        230        240
GTQIGGVIGT TVTTTTTTTT AAGESTRSVI LVDSQENGVR LVHALMACAE AIQQNNLTLA

       250        260        270        280        290        300
EALVKQIGCL AVSQAGAMRK VATYFAEALA RRIYRLSPPQ NQIDHCLSDT LQMHFYETCP

       310        320        330        340        350        360
YLKFAHFTAN QAILEAFEGK KRVHVIDFSM NQGLQWPALM QALALREGGP PTFRLTGIGP

       370        380        390        400        410        420
PAPDNSDHLH EVGCKLAQLA EAIHVEFEYR GFVANSLADL DASMLELRPS DTEAVAVNSV

       430        440        450        460        470        480
FELHKLLGRP GGIEKVLGVV KQIKPVIFTV VEQESNHNGP VFLDRFTESL HYYSTLFDSL

       490        500        510        520        530        540
EGVPNSQDKV MSEVYLGKQI CNLVACEGPD RVERHETLSQ WGNRFGSSGL APAHLGSNAF

       550        560        570        580
KQASMLLSVF NSGQGYRVEE SNGCLMLGWH TRPLITTSAW KLSTAAY
```

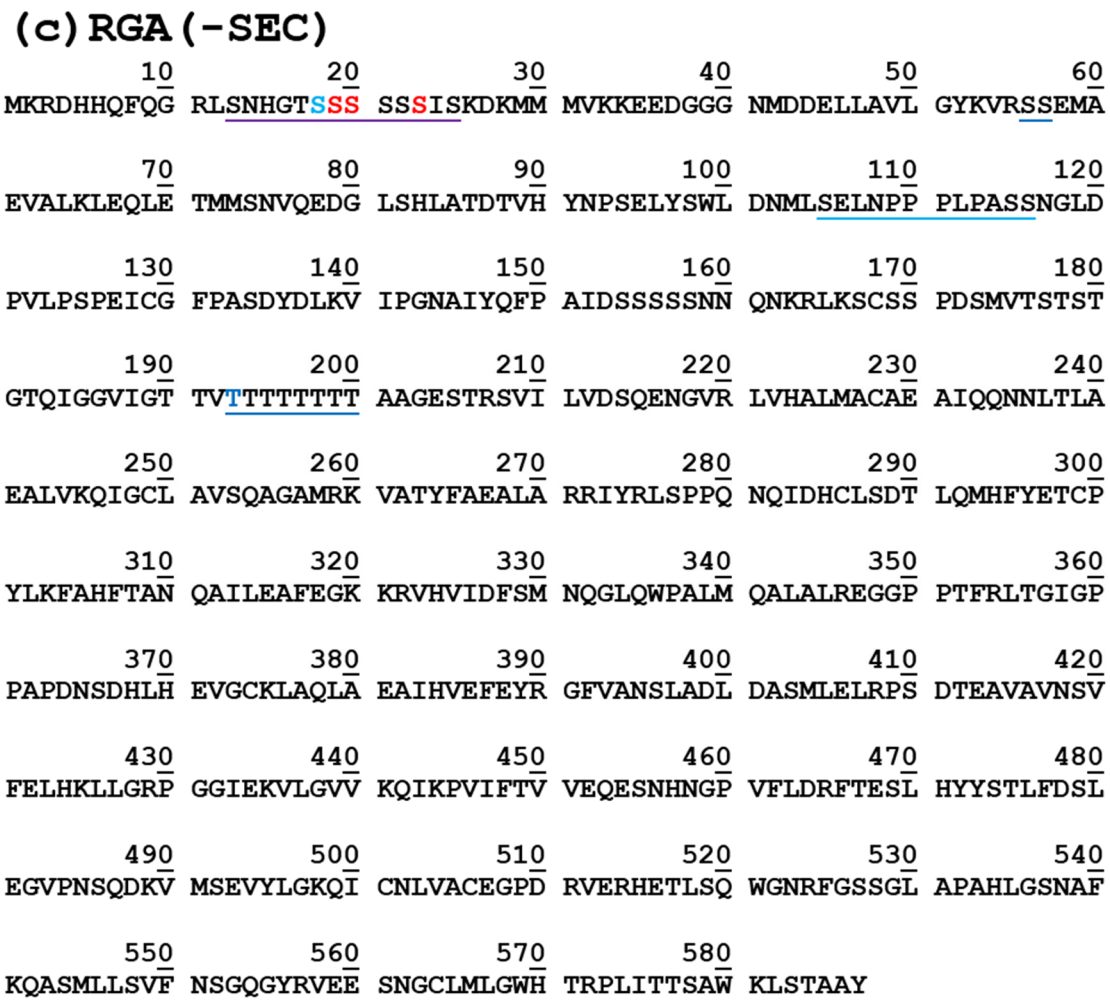


Figure 2.40 - *RGA sequences summarizing residues modified with O-GlcNAc, phosphate, O-hexose, O-GlcNAc & phosphate, O-GlcNAc & O-hexose, phosphate & O-hexose, and O-GlcNAc, phosphate, & O-hexose PTMs on (a) RGA(+SEC), (b) RGA(-SEC+CK1), and (c) RGA(-SEC-CK1). Underlines are color-coded as above to indicate regions in which one or more PTMs have been detected by accurate mass and elution profile.*

This study comprises the most complete characterization of post-translational modifications on any O-GlcNAcylated plant protein to date. It identifies numerous sites of O-GlcNAc, phosphate, and the previously undescribed nuclear PTM, O-hexose, while demonstrating that these PTMs have a complex interplay determining how they collectively contribute to RGA function.

2.5 References

1. Creasy, D. M. & Cottrell, J. S. Unimod: Protein modifications for mass spectrometry. *Proteomics* **4**, 1534-1536 (2004).

2. Farley, A. R. & Link, A. J. Identification and quantification of protein posttranslational modifications. *Methods Enzymol.* **463**, 725-763 (2009).

3. Manning, G., Whyte, D. B., Martinez, R., Hunter, T. & Sudarsanam, S. The protein kinase complement of the human genome. *Science* **298**, 1912-1934 (2002).

4. Hubbard, M. J. & Cohen, P. On target with a new mechanism for the regulation of protein phosphorylation. *Trends Biochem. Sci.* **18**, 172-177 (1993).

5. Cohen, P. The role of protein phosphorylation in human health and disease. *European Journal of Biochemistry* **268**, 5001-5010 (2001).

6. White, F. M. Quantitative phosphoproteomic analysis of signaling network dynamics. *Curr. Opin. Biotechnol.* **19**, 404-409 (2008).

7. Hart, G. W., Slawson, C., Ramirez-Correa, G. & Lagerlof, O. Cross talk between O-GlcNAcylation and phosphorylation: roles in signaling, transcription, and chronic disease. *Annu. Rev. Biochem.* **80**, 825-858 (2011).

8. Torres, C. R. & Hart, G. W. Topography and polypeptide distribution of terminal N-acetylglucosamine residues on the surfaces of intact lymphocytes. Evidence for O-linked GlcNAc. *J. Biol. Chem.* **259**, 3308-3317 (1984).

9. Ngoh, G. A., Facundo, H. T., Zafir, A. & Jones, S. P. O-GlcNAc Signaling in the Cardiovascular System. *Circulation Research* **107**, 171-185 (2010).

10. Ruan, H., Singh, J. P., Li, M., Wu, J. & Yang, X. Cracking the O-GlcNAc code in metabolism. *Trends in Endocrinology & Metabolism* **24**, 301-309 (2013).

11. Hart, G. W., Slawson, C., Ramirez-Correa, G. & Lagerlof, O. Cross talk between O-GlcNAcylation and phosphorylation: roles in signaling, transcription, and chronic disease. *Annu. Rev. Biochem.* **80**, 825-858 (2011).

12. Nolte, D. & Muller, U. Human O-GlcNAc transferase (OGT): genomic structure, analysis of splice variants, fine mapping in Xq13.1. *Mamm. Genome* **13**, 62-64 (2002).

13. Lazarus, M. B., Nam, Y., Jiang, J., Sliz, P. & Walker, S. Structure of human O-GlcNAc transferase and its complex with a peptide substrate. *Nature* **469**, 564-567 (2011).

14. Das, A. K., Cohen, P. T. W. & Barford, D. The structure of the tetratricopeptide repeats of protein phosphatase 5: implications for TPR-mediated protein-protein interactions. *EMBO J.* **17**, 1192-1199 (1998).

15. Cheung, W. D., Sakabe, K., Housley, M. P., Dias, W. B. & Hart, G. W. O-linked beta-N-acetylglucosaminyltransferase substrate specificity is regulated by myosin phosphatase targeting and other interacting proteins. *J. Biol. Chem.* **283**, 33935-33941 (2008).

16. Gao, Y., Wells, L., Comer, F. I., Parker, G. J. & Hart, G. W. Dynamic O-glycosylation of nuclear and cytosolic proteins: cloning and characterization of a neutral, cytosolic beta-N-acetylglucosaminidase from human brain. *J. Biol. Chem.* **276**, 9838-9845 (2001).

17. Toleman, C., Paterson, A. J., Whisenhunt, T. R. & Kudlow, J. E. Characterization of the Histone Acetyltransferase (HAT) Domain of a Bifunctional Protein with Activable O-GlcNAcase and HAT Activities. *J. Biol. Chem.* **279**, 53665-53673 (2004).

18. Comtesse, N., Maldener, E. & Meese, E. Identification of a Nuclear Variant of MGEA5, a Cytoplasmic Hyaluronidase and a β-N-Acetylglucosaminidase. *Biochem. Biophys. Res. Commun.* **283**, 634-640 (2001).

19. Kim, E. J., Kang, D. O., Love, D. C. & Hanover, J. A. Enzymatic characterization of O-GlcNAcase isoforms using a fluorogenic GlcNAc substrate. *Carbohydr. Res.* **341**, 971-982 (2006).

20. Hart, G. W. & Akimoto, Y. in *Essentials of Glycobiology* (eds Varki, A. et al.) (The Consortium of Glycobiology Editors, La Jolla, California, Cold Spring Harbor (NY), 2009).

21. Wells, L., Vosseller, K. & Hart, G. W. Glycosylation of nucleocytoplasmic proteins: signal transduction and O-GlcNAc. *Science* **291**, 2376-2378 (2001).

22. Hart, G. W., Housley, M. P. & Slawson, C. Cycling of O-linked beta-N-acetylglucosamine on nucleocytoplasmic proteins. *Nature* **446**, 1017-1022 (2007).

23. Love, D. C. & Hanover, J. A. The hexosamine signaling pathway: deciphering the "O-GlcNAc code". *Sci. STKE* **2005**, re13 (2005).

24. Hanover, J. A. Glycan-dependent signaling: O-linked N-acetylglucosamine. *FASEB J.* **15**, 1865-1876 (2001).

25. Copeland, R. J., Han, G. & Hart, G. W. O-GlcNAcomics - Revealing Roles of O-GlcNAcylation in Disease Mechanisms and Development of Potential Diagnostics. *Proteomics Clin. Appl.* (2013).

26. Butkinaree, C., Park, K. & Hart, G. W. O-linked β-N-acetylglucosamine (O-GlcNAc): Extensive crosstalk with phosphorylation to regulate signaling and transcription in response to nutrients and stress. *Biochimica et Biophysica Acta (BBA) - General Subjects* **1800**, 96-106 (2010).

27. Whelan, S. A., Lane, M. D. & Hart, G. W. Regulation of the O-Linked ß-N-Acetylglucosamine Transferase by Insulin Signaling. *Journal of Biological Chemistry* **283**, 21411-21417 (2008).

28. Song, M. *et al*. o-GlcNAc transferase is activated by CaMKIV-dependent phosphorylation under potassium chloride-induced depolarization in NG-108-15 cells. *Cell. Signal.* **20**, 94-104 (2008).

29. Wang, Z., Pandey, A. & Hart, G. W. Dynamic Interplay between O-Linked N-Acetylglucosaminylation and Glycogen Synthase Kinase-3-dependent Phosphorylation. *Molecular & Cellular Proteomics* **6**, 1365-1379 (2007).

30. Wang, Z., Gucek, M. & Hart, G. W. Cross-talk between GlcNAcylation and phosphorylation: Site-specific phosphorylation dynamics in response to globally elevated O-GlcNAc. *Proceedings of the National Academy of Sciences* **105**, 13793-13798 (2008).

31. Wang, Z. *et al*. Extensive crosstalk between O-GlcNAcylation and phosphorylation regulates cytokinesis. *Sci. Signal.* **3**, ra2 (2010).

32. Hartweck, L. M., Scott, C. L. & Olszewski, N. E. Two O-Linked N-Acetylglucosamine Transferase Genes of Arabidopsis thaliana L. Heynh. Have Overlapping Functions Necessary for Gamete and Seed Development. *Genetics* **161**, 1279-1291 (2002).

33. Jacobsen, S. E. & Olszewski, N. E. Mutations at the SPINDLY locus of Arabidopsis alter gibberellin signal transduction. *Plant Cell* **5**, 887-896 (1993).

34. Hartweck, L. M. Gibberellin signaling. *Planta* **229**, 1-13 (2008).

35. Thornton, T. M., Swain, S. M. & Olszewski, N. E. Gibberellin signal transduction presents …the SPY who O-GlcNAc'd me. *Trends Plant Sci.* **4**, 424-428 (1999).

36. Olszewski, N. E., West, C. M., Sassi, S. O. & Hartweck, L. M. O-GlcNAc protein modification in plants: Evolution and function. *Biochimica et Biophysica Acta (BBA) - General Subjects* **1800**, 49-56 (2010).

37. Kim, Y. C. *et al*. O-GlcNAcylation of the Plum pox virus capsid protein catalyzed by SECRET AGENT: characterization of O-GlcNAc sites by electron transfer dissociation mass spectrometry. *Amino Acids* **40**, 869-876 (2011).

38. Perez Jde, J. *et al*. O-GlcNAc modification of the coat protein of the potyvirus Plum pox virus enhances viral infection. *Virology* **442**, 122-131 (2013).

39. Hartweck, L. M., Genger, R. K., Grey, W. M. & Olszewski, N. E. SECRET AGENT and SPINDLY have overlapping roles in the development of Arabidopsis thaliana L. Heyn. *Journal of Experimental Botany* **57**, 865-875 (2006).

40. Koornneef, M. & Veen, J. H. Induction and analysis of gibberellin sensitive mutants in Arabidopsis thaliana (L.) heynh. *Theor. Appl. Genet.* **58**, 257-263 (1980).

41. Richards, D. E., King, K. E., Ait-Ali, T. & Harberd, N. P. HOW GIBBERELLIN REGULATES PLANT GROWTH AND DEVELOPMENT: A Molecular Genetic Analysis of Gibberellin Signaling. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **52**, 67-88 (2001).

42. Silverstone, A. L., Mak, P. Y. A., Martinez, E. C. & Sun, T. The New RGA Locus Encodes a Negative Regulator of Gibberellin Response in Arabidopsis thaliana. *Genetics* **146**, 1087-1099 (1997).

43. http://5e.plantphys.net/article.php?ch=&id=382.

44. Daviere, J. M., de Lucas, M. & Prat, S. Transcriptional factor interaction: a central step in DELLA function. *Curr. Opin. Genet. Dev***. 1**8, 295-303 (2008).

45. English, A. M. Development and Application of Methodologies to Facilitate Peptide Sequencing with Mass Spectrometry. *ProQuest Dissertations and These*s (2011).

46. Hochuli, E., Döbeli, H. & Schacher, A. New metal chelate adsorbent selective for proteins and peptides containing neighbouring histidine residues. *Journal of Chromatography* **A 41**1, 177-184 (1987).

47. Einhauer, A. & Jungbauer, A. The FLAG™ peptide, a versatile fusion tag for the purification of recombinant proteins. *J. Biochem. Biophys. Method***s 4**9, 455-465 (2001).

48. Olsen, J. V., Ong, S. E. & Mann, M. Trypsin cleaves exclusively C-terminal to arginine and lysine residues. *Mol. Cell. Proteomic***s** 3, 608-614 (2004).

49. Udeshi, N. D., Compton, P. D., Shabanowitz, J., Hunt, D. F. & Rose, K. L. Methods for analyzing peptides and proteins on a chromatographic timescale by electron-transfer dissociation mass spectrometry. *Nat. Protoc***.** 3, 1709-1717 (2008).

50. Sleno, L. The use of mass defect in modern mass spectrometry. *J. Mass Spectrom***. 4**7, 226-236 (2012).

51. Gross, J. in 67-116 (Springer Berlin Heidelberg, 2011).

52. Zhang, G*. et a*l. in (ed FenyÃ¶, D.) 211-222 (Humana Press, 2010).

53. Iribarne, J. V. & Thomson, B. A. On the evaporation of small ions from charged droplets. *J. Chem. Phys***. 6**4, 2287-2294 (1976).

54. Wang, Z*. et a*l. Enrichment and Site Mapping of O-Linked N-Acetylglucosamine by a Combination of Chemical/Enzymatic Tagging, Photochemical Cleavage, and Electron Transfer Dissociation Mass Spectrometry. *Molecular & Cellular Proteomic***s** 9, 153-160 (2010).

55. Comer, F. I., Vosseller, K., Wells, L., Accavitti, M. A. & Hart, G. W. Characterization of a Mouse Monoclonal Antibody Specific for O-Linked N-Acetylglucosamine. *Anal. Biochem***. 29**3, 169-177 (2001).

56. Vosseller, K*. et a*l. O-Linked N-Acetylglucosamine Proteomics of Postsynaptic Density Preparations Using Lectin Weak Affinity Chromatography and Mass Spectrometry. *Molecular & Cellular Proteomic***s** 5, 923-934 (2006).

57. Syka, J. E., Coon, J. J., Schroeder, M. J., Shabanowitz, J. & Hunt, D. F. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A***. 10**1, 9528-9533 (2004).

58. Zubarev, R. A. Reactions of polypeptide ions with electrons in the gas phase. *Mass Spectrom. Rev.* **22**, 57-77 (2003).

59. Mohr, J. H., Swart, R. & Huber, C. G. Morphology and efficiency of poly(styrene-co-divinylbenzene)-based monolithic capillary columns for the separation of small and large molecules. *Anal. Bioanal Chem.* **400**, 2391-2402 (2011).

60. Schulz, B. L., Packer, N. H. & Karlsson, N. G. Small-scale analysis of O-linked oligosaccharides from glycoproteins and mucins separated by gel electrophoresis. *Anal. Chem.* **74**, 6088-6097 (2002).

61. Simon, R., Enjalbert, Q., Biarc, J., Lemoine, J. & Salvador, A. Evaluation of hydrophilic interaction chromatography (HILIC) versus C18 reversed-phase chromatography for targeted quantification of peptides by mass spectrometry. *Journal of Chromatography A* **1264**, 31-39 (2012).

62. Gilar, M. & Jaworski, A. Retention behavior of peptides in hydrophilic-interaction chromatography. *Journal of Chromatography A* **1218**, 8890-8896 (2011).

63. Krusemark, C. J., Ferguson, J. T., Wenger, C. D., Kelleher, N. L. & Belshaw, P. J. Global amine and acid functional group modification of proteins. *Anal. Chem.* **80**, 713-720 (2008).

64. Krusemark, C. J., Frey, B. L., Belshaw, P. J. & Smith, L. M. Modifying the charge state distribution of proteins in electrospray ionization mass spectrometry by chemical derivatization. *J. Am. Soc. Mass Spectrom.* **20**, 1617-1625 (2009).

65. Lundblad, R. L. in *Protein Modification* (C R C Press LLC, 1995).

66. Chou, P. Y. & Fasman, G. D. Empirical Predictions of Protein Conformation. *Annu. Rev. Biochem.* **47**, 251-276 (1978).

67. Shortreed, M. R. *et al*. Ionizable isotopic labeling reagent for relative quantification of amine metabolites by mass spectrometry. *Anal. Chem.* **78**, 6398-6403 (2006).

68. Hart, G. W. & West, C. M. in *Essentials of Glycobiology* (eds Varki, A. et al.) (The Consortium of Glycobiology Editors, La Jolla, California, Cold Spring Harbor (NY), 2009).

69. Bork, P., Downing, A. K., Kieffer, B. & Campbell, I. D. Structure and distribution of modules in extracellular proteins. *Q. Rev. Biophys.* **29**, 119-167 (1996).

70. Hase, S. *et al*. A new trisaccharide sugar chain linked to a serine residue in bovine blood coagulation factors VII and IX. *J. Biochem.* **104**, 867-868 (1988).

71. Bjoern, S. *et al*. Human plasma and recombinant factor VII. Characterization of O-glycosylations at serine residues 52 and 60 and effects of site-directed mutagenesis of serine 52 to alanine. *J. Biol. Chem.* **266**, 11051-11057 (1991).

72. Harris, R. J., Ling, V. T. & Spellman, M. W. O-linked fucose is present in the first epidermal growth factor domain of factor XII but not protein C. *J. Biol. Chem.* **267**, 5102-5107 (1992).

73. Buko, A. M. *et al*. Characterization of a posttranslational fucosylation in the growth factor domain of urinary plasminogen activator. *Proc. Natl. Acad. Sci. U. S. A.* **88**, 3992-3996 (1991).

74. Moloney, D. J. *et al*. Mammalian Notch1 is modified with two unusual forms of O-linked glycosylation found on epidermal growth factor-like modules. *J. Biol. Chem.* **275**, 9604-9611 (2000).

75. Takeuchi, H. & Haltiwanger, R. S. Role of glycosylation of Notch in development. *Semin. Cell Dev. Biol.* **21**, 638-645 (2010).

76. Stanley, P. & Okajima, T. in *Current Topics in Developmental Biology* 131-164 (Academic Press.

77. Matsuura, A. *et al*. O-linked N-acetylglucosamine is present on the extracellular domain of notch receptors. *J. Biol. Chem.* **283**, 35486-35495 (2008).

78. Sakaidani, Y. *et al*. O-linked-N-acetylglucosamine modification of mammalian Notch receptors by an atypical O-GlcNAc transferase Eogt1. *Biochem. Biophys. Res. Commun.* **419**, 14-19 (2012).

79. Li, W. *et al*. Statistical analysis of electron transfer dissociation pairwise fragmentation patterns. *Anal. Chem.* **83**, 9540-9545 (2011).

80. Good, D. M., Wirtala, M., McAlister, G. C. & Coon, J. J. Performance characteristics of electron transfer dissociation mass spectrometry. *Mol. Cell. Proteomics* **6**, 1942-1951 (2007).

81. - SELECTIVE CLEAVAGE OF THE METHIONYL PEPTIDE BONDS IN RIBONUCLEASE WITH CYANOGEN BROMIDE1. - *J. Am. Chem. Soc.*, - 1510.

82. Andreev, Y. A., Kozlov, S. A., Vassilevski, A. A. & Grishin, E. V. Cyanogen bromide cleavage of proteins in salt and buffer solutions. *Anal. Biochem.* **407**, 144-146 (2010).

83. - Capasso, S., - Mazzarella, L., - Sica, F., - Zagari, A. & - Salvadori, S. - Spontaneous cyclization of the aspartic acid side chain to the succinimide derivative. - *J. Chem. Soc. , Chem. Commun.*, - 919.

84. Tetaz, T., Morrison, J. R., Andreou, J. & Fidge, N. H. Relaxed specificity of endoproteinase Asp-N: this enzyme cleaves at peptide bonds N-terminal to glutamate as well as aspartate and cysteic acid residues. *Biochem. Int.* **22**, 561-566 (1990).

85. Nomura, T. *et al*. Purification, cDNA cloning, and expression of UDP-Gal: glucosylceramide beta-1,4-galactosyltransferase from rat brain. *J. Biol. Chem.* **273**, 13570-13577 (1998).

Chapter 3: Site-specific phosphorylation of the DNA damage response mediator rad9 by cyclin-dependent kinases regulates activation of checkpoint kinase 1

3.1 Introduction

3.1.1 DNA Damage Response

Eukaryotes have the ability to monitor and react to DNA damage including nicks, gaps, double-strand breaks, and any alterations that block DNA replication. This is known as the DNA damage response (DDR) whereby cells regulate DNA replication, transcription, and repair in response to DNA damage. [1]. Deficiencies in the DDR have negative implications for pathologies like cancer, because cells fail to induce senescence and apoptosis [1, 2].

The DDR is a signal transduction pathway regulated in large part by the post-translational modification, phosphorylation. Protein kinases activated by DNA damage are therefore key components of the DDR. In all organisms studied, these DNA-damage-sensing kinases include members from the phosphatidylinositol 3-kinase-like (PIKK) family. In humans, two PIKKs critical to DDR are ATM and ATR. Much of what is currently known about the DDR is due to studies in *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe.* ATM and ATR respectively correspond to Mec1p and Tel1p in *S. cerevisiae*, and Rad3 and Tel1 in *S. pombe* [3]. After PIKKs are activated by DNA-protein complexes formed in response to DNA-damage, they regulate various downstream proteins involved in the DDR including two important effector "checkpoint" kinases. These are Chk1 and Chk2 in humans or the orthologous spChk1 and spCds1 in *S. pombe*, and scChk1 and scRad53 in *S. cerevisiae* [4]. They contain highly conserved kinase domains but are structurally and functionally distinct with little overlap in their

downstream targets. Chk1 is primarily activated downstream of ATR in response to DNA comprises that include stalled replication forks, DNA crosslinks, ultraviolet radiation, and ionizing radiation. Chk2 is primarily activated downstream of ATM as a result of DNA double-strand breaks [4].

Another key component to the DDR are DDR mediator/adaptor proteins that localize to sites of DNA damage and are typically phosphorylated by multiple kinases including PIKKs in response to it [5]. DDR mediator proteins aid PIKK-dependent activation of checkpoint kinases by participating in the assembly of DNA damage foci and facilitating protein-protein interactions at sites of DNA damage [6, 7]. The prototypical DDR mediator and first checkpoint protein discovered was Rad9 in *S. cerevisiae* [8].

3.1.2 Rad9

Rad9 is a 148kDa protein essential for the arrest of cell division triggered by DNA damage. It is related to *S. pombe* mediator Crb2, and human mediators 53BP1, MDC1, and BRCA1. This central DDR mediator is required in late G1, intra-S, and at the G2/M transition for checkpoint delays [9]. It detects DNA damage and arrests cells in G2 until the damage is repaired, and it also plays additional key roles in DNA damage repair. [8]. Rad9 is recruited to sites of DNA damage either by interaction with two histone modifications or with mediator protein Dpb11, also an activator of Mec1 [10]. Once recruited, Rad9's SQ/TQ cluster domain (SCD) becomes phosphorylated by PIKKs to induce its oligomerization and activation [11]. This hyperphosphorylated Rad9 forms a 560kDa complex that recruits mediator Rad53 so it too may be phosphorylated by PIKKs. This in turn promotes *trans* autophosphorylation of Rad53 to fully activate itself [12, 13]. Activated Rad53 is released by Rad9, so it may transmit the checkpoint signal throughout the nucleus.

3.1.3 Checkpoint Kinase 1 (Chk1) Activation by Rad9

Chk1 is PIKK-dependently phosphorylated on its C-terminus, leading to *cis* autophosphorylation in the same region and activation [14]. The mechanism by which Rad9 interacts with and activates Chk1 is much less understood than its role in Rad53 activation. Rad9 has mainly been connected to DNA damage-induced Chk1 activation due to its role in providing a protein-binding scaffold at sites of DNA damage. The N-terminus of Rad9 has been identified as critical for phosphorylation and activation of Chk1 and has been appropriately termed the Chk1 activation domain (CAD), although the details of how the CAD works are poorly understood [15].

DDR mediators are often phosphorylated during the cell cycle even in the absence of DNA damage [16, 17]. Rad9 contains 20 Ser/Thr-Pro cyclin-dependent kinase (CDK) consensus motifs (9 are the full Ser/Thr-Pro-X-Lys/Arg motif) [18]. CDKs regulate and time events in the eukaryotic cell cycle and their activation requires their association with cyclin subunits whose concentrations fluctuate according to a variety of cellular factors [19]. One CDK, cdc28, regulates cell-cycle transitions in *S. cerevisiae*. Three cyclins, Cln1–3, operate in conjunction with cdc28 in G1. Entry into S phase and mitosis is controlled by cdc28's association with cyclins Clbs5-6 and Clbs1-4, respectively [20]. Evidence exists for the phosphorylation of Rad9 by cdc28 *in vitro* [18], and mass spectrometric analyses have revealed 13 of the 20 CDK sites to be phosphorylated *in vivo* [21-23]. Still, the biological function of CDK-dependent Rad9 phosphorylation remains largely uncharacterized.

This study sought to characterize phosphorylation of the Rad9 CAD by Cdc28/Clb complexes in S, G2, and M phases by mass spectrometry, and in collaboration with the Lowndes Lab at NUI Galway, determine its role in Chk1 activation. We report that phosphorylation of

CDK consensus sites in the Rad9 CAD regulate a DNA damage-independent interaction between Chk1 and Rad9, and regulate activation of Chk1-dependent signaling in response to DNA damage. CDK sites T125 and T143 were shown to be especially important for Chk1 activation by Rad9. Our model suggests that CDK-dependent phosphorylation of the Rad9 CAD during S, G2, and M phases is required for interaction between Chk1 and Rad9. This enables phosphorylation of both proteins by PIKK, which leads to Chk1 activation and its subsequent exit from the site of DNA damage so it may perform its appropriate nuclear signaling [24].

3.2 Materials and Methods

3.2.1 Molecular biology methods

Materials and methods used for yeast strains, cell cycle arrest and checkpoint experiments, antibodies, *in vitro* kinase assays, yeast two-hybrid experiments, yeast native extracts and immunoprecipitation, peptide pull-downs, and SPOT synthesis of peptides and array experiments are described in [24].

3.2.2 Rad9 purification

Cell cycle-modified (C-Rad9) Rad9-GFP-FLAG was purified by double immunoaffinity purification from whole-cell extracts, prepared from 10 L of asynchronously growing, G1 or G2/M arrested cells (1-2x10^7 cells/mL). The efficiency of the G1 and G2/M arrests using 10 µg/ml α-factor and 20 µg/ml nocodazole, respectively, was verified by microscopy. Approximately 1.5 g of total protein was obtained for each crude cell extract. Extracts were pre-cleared with 200 µL of packed protein G beads (GE Healthcare) prior to incubation with 200 µL of packed FLAG M2-affinity beads (Sigma-Aldrich) for 2h at 4°C. Four successive elutions were performed using 200 µL of 3xFLAG peptide (100 µg/mL, Sigma-Aldrich) and Rad9-GFP-FLAG-

containing fractions were identified by western blotting. 20 µL of GFP-Trap agarose beads (Chromotek) were added to the pooled Rad9-GFP-FLAG containing fractions. Batch binding was allowed to occur for 2 h at 4°C with gentle agitation. The beads were washed 3 times with wash buffer (150 mM KAc, 50 mM HEPES pH 7.5) and once with ultra-pure water. Washed beads were resuspended in an equal volume of 0.1% glacial acetic acid. The presence of purified Rad9-GFP-FLAG was verified by silver stain gel on which 1 µL of the final beads was loaded.

3.2.3 Mass Spectrometric Analysis of Rad9-GFP-FLAG

Rad9-GFP-FLAG purified from asynchronous, G1- or G2/M-arrested cells was reduced and carbamidomethylated on GFP beads at room temperature using dithiothreitol (Sigma Aldrich, St. Louis, MO) and iodoacetamide (Sigma Aldrich), respectively, then subjected to proteolytic digestion (1: 20 enzyme to substrate) using endoproteinase Lys-C (Roche, Penzberg, Germany) using methods similar to those previously reported [25]. Half of each sample was subjected to sub-digestion using trypsin (Promega, Madison, WI) and a quarter of the Lys-C digested asynchronous sample was subdigested with endoproteinase Asp-N (Roche, Penzberg, Germany), at the same ratios as above. Prior to mass spectrometric analysis, each digest was dried down and resuspended in 0.1% acetic acid in water. For each analysis, a 1 pmol fraction was pressure loaded onto a 360 µm o.d. × 75 µm i.d. fused silica capillary precolumn packed with 5 cm of C18 reverse-phase resin (5–20 µm diameter, 120 Å pore size, YMC Co., Ltd., Kyoto, Japan). Following a 15 min desalting rinse using 0.1 M acetic acid, the precolumn was connected to a 360 µm o.d. × 50 µm i.d. analytical column packed with 6-8 cm of C18 resin (5 µm diameter, 120 Å pore size, YMC Co., Ltd.) and equipped with an electrospray emitter tip as previously described [26]. Endoproteinase Lys-C or trypsin generated peptides were then gradient eluted into the mass spectrometer using electrospray ionization at a flow rate of 60 nL/min using an LC gradient

previously described [27]. Mass analysis was completed using an acquisition method consisting of one high resolution MS1 scan (resolving power of 50,000 (FT Ultra) or 60,000 (Orbitrap) at *m/z* 400) acquired in the FT-ICR or Orbitrap followed by 5 data dependent MS/MS scans using collision-activated dissociation (CAD) and electron transfer dissociation (ETD) acquired in the ion trap of the LTQ-FT Ultra or LTQ-Orbitrap hybrid instrument (Thermo Fisher Scientific, Bremen, Germany). Data dependent parameters included a repeat count of 3, repeat duration of 30 s, and exclusion list duration of 30 s. MS/MS parameters for ETD scans included 35 ms reaction time, 3 *m/z* precursor isolation window, charge state rejection "on" for +1 and unassigned charge state precursor ions, 5×10^5 FTMS (Orbitrap) or 1×10^6 (FT Ultra) full automated gain control target, 1×10^4 $ITMS^n$ automated gain control target, and 2×10^5 reagent target with azulene as the electron transfer reagent.

3.3 Results and Discussion

3.3.1 Rad9-GFP-FLAG Purification

Rad9-GFP-FLAG was expressed in three separate samples of *S. cerevisiae* in the absence of DNA damage. In the first, the cells were allowed to remain asynchronous. In the second and third, the cells were arrested at cell cycle stages G1 and G2, respectively. All three were tandem-immunopurified to generate highly enriched protein amenable for the detection of post-translational modifications by mass spectrometry. Samples were firstly bound to anti-FLAG M2 affinity beads, washed, and eluted with 3xFLAG peptide. These Rad9-enriched samples were then bound to GFP-trap A beads for a second round of purification, washed, and resuspended in 0.01% acetic acid (pH 4) until on-beads digestion. This yielded several pmols of purified Rad9 protein from each sample (**Figure 3.1**).

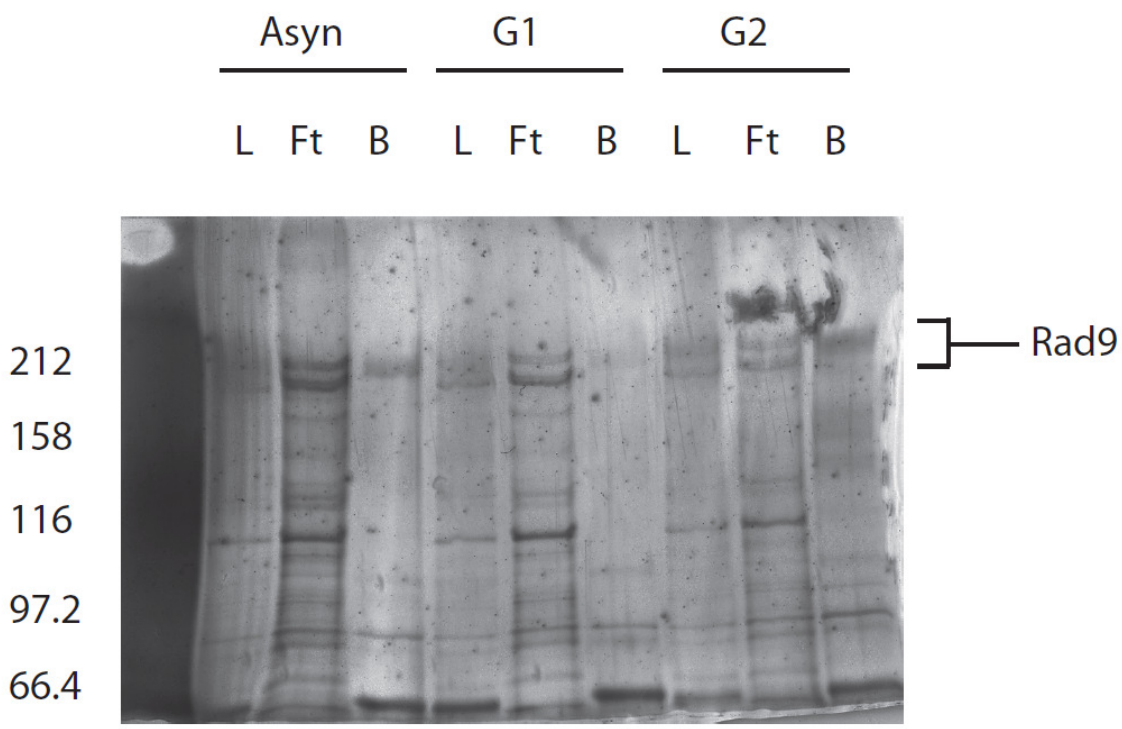


Figure 3.1 – *Tandem purified Rad9-GFP-FLAG visualized by Silver stain on a 6% acrylamide SDS-PAGE gel. L – 2.5% of Rad9 purified by FLAG IP. Ft – 2.5% of unbound flowthrough from FLAG IP. B – 2.5% of bound sample from GFP-IP.*

3.3.2 In silico Rad9 Digestion

Simulated digestions of Rad9 were performed to determine candidate peptides that would allow for the detection of all putative CDK sites in the Chk1 activation domain. Two separate theoretical digests using Endoproteinase Lys-C and trypsin generated peptides that met this goal, and covered >90% of Rad9 as a whole (**Tables 3.1 & 3.2**).

Mass	Position	Peptide	CDK Site	Comments
1075.5	1-9	MSGQLVQWK		+
1287.7	10-14	SSPDRVTQSAIK	S11	+
2734.3	22-46	EALHSPLADGDMNEMNVPVD PLENK	S26	+
1257.7	47-58	VNSTNIIEGSPK		+
738.4	59-65	ANPNPVK	S56	?
1243.6	66-75	FMNTSEIFQK		+

5283.6	76-121	SLGLLDESPRHDDELNIEVG DNDRPNANILHNERTPDLDR IANFFK	S83, T110	?
758.4	122-128	SNRTPGK		?
716.4	129-134	ENLLTK		-
1794.9	135-149	YQSSDLEDTPLMLRK	T143	+
1534.7	151-163	MTFQTPTDPLEQK	T155	+
5552.4	170-220	SDTGFCYYGEQNDGEENASL EVTEADATFVQMAERSADNY DCALEGIVTPK	T218	?
590.3	224-228	DELSK		-
1233.6	229-239	SGGMQDERVQK		+
1969.9	240-257	TQIMISAESPNSISSYDK	S248	+
1429.8	260-272	ITGNGRTTRNVNK		+
1704.8	273-287	VFNNNEDNIGAIEEK		+
3581.6	294-323	SENYSSDDLRERNNQIIQSN ESEEINELEK		+
5264.4	324-371	NLNVSGRENDVNNLDIDINS AVSGTPSRNNAEEEMYSSES VNNREPSK	T348	?
998.5	373-379	WIFRYSK		?
3826.9	382-415	TENNSNRSTQIVNNPRTQEM PLDSISIDTQPLSK		?
2937.5	416-442	SFNTETNNELETQIIVSSLS QGISAQK		+
1528.8	443-456	GPVFHSTGQTEEIK		+
4511.3	457-497	TQIINSPEQNALNATFETPV TLSRINFEPILEVPETSSPS K	S462, T474, S494	?
579.3	498-502	NTMSK		-

925.5	503-511	PSNSSPIPK	S507	?
4954.3	514-557	DTFNIHEREVETNNVFSNDI QNSSNAATRDDIIIAGSSDF NEQK		?
1534.8	558-570	EITDRIYLQLSGK		+
2338.0	571-591	QISDSGSDETERMSPNELDT K		+
3216.5	593-620	ESTIMSEVELTQELPEVEEQ QDLQTSPK		+
1431.8	622-633	LVVEEETLMEIK		+
1239.6	637-647	GNSLQLHDDNK		+
694.3	648-653	ECNSDK		-
1869.9	654-670	QDGTESLDVALIEHESK		+
875.4	671-678	GQSSELQK		?
2544.3	679-699	NLMQLFPSESQEIIQNRRTI K		+
586.4	700-703	RRQK		-
1748.8	704-718	DTIEIGEEEENRSTK		+
532.3	719-723	TSPTK	S720	-
1188.6	727-737	RNSDLDAASIK		+
1849.9	738-755	REPSCSITIQTGETGSGK		+
2287.1	759-778	EQSYVFPEGIRTADNSFLSK	S584	+
2235.0	779-796	DDIIFGNAVWCQYTWNYK	S618	+
3264.6	797-824	FYPGILLEVDTNQDGCWIYF ETGRSLTK		+
5684.8	825-873	DEDIYYLDIRIGDAVTFDGN EYVVVGLECRSHDLNIIRCI RGYDTVHLK		?
758.4	876-883	NASGLLGK		?
629.4	884-888	RTLIK		-
1518.8	889-902	ALSSISLDLSEWAK		+
972.5	906-913	IILEDNEK		?
1857.0	916-930	GDAYRYLRHPIRGRK		+

975.5	931-939	SMTNVLSPK	S937	+
743.3	941-946	HTDDEK		-
2208.0	947-965	DINTHTEVYNNEIESSSEK		+
3159.6	972-1001	DSRDALAEHAGAPSLLFSSG EIRTGNVFDK		?
4858.5	1002-1044	CIFVLTSLFENREELRQTIE SQGGTVIESGFSTLFNFTHP LAK		?
559.3	1045-1049	SLVNK		-
1342.7	1050-1061	GNTDNIRELALK		+
516.3	1062-1065	LAWK		-
1820.9	1066-1081	PHSLFADCRFACLITK		+
908.6	1082-1088	RHLRSLK		?
1827.0	1089-1103	YLETLALGWPTLHWK		+
909.5	1104-1111	FISACIEK		?
3114.7	1113-1139	RIVPHLIYQYLLPSGESFRL SLDSPSK		+
2906.4	1145-1169	SNNIFSFYTQFLRGSNLRDQ ICGVK		+
2383.2	1171-1190	MLNDYIVIVWGRSELDSFVK		+
4999.6	1191-1236	FAFACLSAGRMLTIDLPNID VDDTEPLLNALDSLVPRIGS ELSNRK		?
1281.6	1239-1249	FLIYANENNGK		+
1454.9	1254-1265	LLERLRSQISLK		+
1284.6	1269-1278	FNYIFHTESK		+

4657.3	1279-1318	EWLIQTIINEDTGFHDDITD NDIYNTISEVR(RIPGLINSK*)*		?

Table 3.1 – In silico Lys-C digestion of Rad9 displaying all >500 Da peptides. Residues belonging to the C-terminal tag linker are indicated in parentheses. Peptide masses (Da), position numbers, and putative CDK sites are indicated. The comments section refers to peptides predicted to be detected (+) or not detected (-), with (?) indicating uncertainty of detection.

Mass	Position	Peptide	CDK Site	Comments
1075.5	1-9	MSGQLVQWK		+
560.3	10-14	SSPDR	S11	-
745.4	15-21	VTQSAIK		?
2734.3	22-46	EALHSPLADGDMNEMNVPVD PLENK	S26	+
1257.7	47-58	VNSTNIIEGSPK	S56	+
738.4	59-65	ANPNPVK		+
1243.6	66-75	FMNTSEIFQK		+
1085.6	76-85	SLGLLDESPR	S83	+
2798.3	86-109	HDDELNIEVGDNDRPNANIL HNER		+
715.4	110-115	TPDLDR	T110	?
738.4	116-121	IANFFK		?
716.4	129-134	ENLLTK		?
1666.8	135-148	YQSSDLEDTPLMLR	T143	+
1534.7	151-163	MTFQTPTDPLEQK	T155	+
3875.6	170-204	SDTGFCYYGEQNDGEENASL EVTEADATFVQMAER		?
1694.8	205-220	SADNYDCALEGIVTPK	T218	+
590.3	224-228	DELSK		-
878.4	229-236	SGGMQDER		+
1969.9	240-257	TQIMISAESPNSISSYDK	S248	+
616.3	260-265	ITGNGR		?
1704.8	273-287	VFNNNEDNIGAIEEK		+
1184.5	294-303	SENYSSDDLR		+

2130.0	306-323	NNQIIQSNESEEINELEK		+
758.4	324-330	NLNVSGR		?
2229.1	331-351	ENDVNNLDIDINSAVSGTPS R	T348	+
1871.8	352-367	NNAEEEMYSSESVNNR		+
620.3	373-376	WIFR		-
833.4	382-388	TENNSNR		?
1027.5	389-397	STQIVNNPR		+
2002.0	398-415	TQEMPLDSISIDTQPLSK		+
2937.5	416-442	SFNTETNNELETQIIVSSLS QGISAQK		+
1528.8	443-456	GPVFHSTGQTEEIK		+
2643.4	457-480	TQIINSPEQNALNATFETPV TLSR	S462, T474	+
1886.0	481-497	INFEPILEVPETSSPSK	S494	+
1486.7	498-511	NTMSKPSNSSPIPK	S507	+
1030.5	514-521	DTFNIHER		?
2309.1	522-542	EVETNNVFSNDIQNSSNAAT R		+
1650.8	543-557	DDIIIAGSSDFNEQK		+
632.3	558-562	EITDR		-
920.5	563-570	IYLQLSGK		?
1322.6	571-582	QISDSGSDETER		+
1033.5	583-591	MSPNELDTK	S584	+
3216.5	593-620	ESTIMSEVELTQELPEVEEQ QDLQTSPK	S618	+
1431.8	622-633	LVVEEETLMEIK		+
1239.6	637-647	GNSLQLHDDNK		+
694.3	648-653	ECNSDK		-
1869.9	654-670	QDGTESLDVALIEHESK		+
875.4	671-678	GQSSELQK		?
2046.0	679-695	NLMQLFPSESQEIIQNR		+
1432.6	704-715	DTIEIGEEEENR		+
532.3	719-723	TSPTK	S720	-
1032.5	728-737	NSDLDAASIK		+

1693.8	739-755	EPSCSITIQTGETGSGK		+
1323.6	759-769	EQSYVFPEGIR		+
981.5	770-778	TADNSFLSK		+
2235.0	779-796	DDIIFGNAVWCQYTWNYK		+
2835.3	797-820	FYPGILLEVDTNQDGCWIYF ETGR		+
1313.6	825-834	DEDIYYLDIR		+
2155.0	835-854	IGDAVTFDGNEYVVVGLECR		+
966.5	855-862	SHDLNIIR		?
931.5	866-873	GYDTVHLK		?
758.4	876-883	NASGLLGK		?
1518.8	889-902	ALSSISLDLSEWAK		+
972.5	906-913	IILEDNEK		?
580.3	916-920	GDAYR		-
521.3	924-927	HPIR		-
975.5	931-939	SMTNVLSPK	S937	+
743.3	941-946	HTDDEK		-
2208.0	947-965	DINTHTEVYNNEIESSSEK		+
2040.0	975-994	DALAEHAGAPSLLFSSGEIR		+
779.4	995-1001	TGNVFDK		?
1440.7	1002-1013	CIFVLTSLFENR		+
545.3	1014-1017	EELR		-
2908.5	1018-1044	QTIESQGGTVIESGFSTLFN FTHPLAK		+
559.3	1045-1049	SLVNK		-
788.4	1050-1056	GNTDNIR		?
572.4	1057-1061	ELALK		-
1542.8	1062-1074	LAWKPHSLFADCR		+
794.4	1075-1081	FACLITK		?

1827.0	1089-1103	YLETLALGWPTLHWK		+
909.5	1104-1111	FISACIEK		?
2131.1	1114-1131	IVPHLIYQYLLPSGESFR		+
845.4	1132-1139	LSLDSPSK	S1136	?
1635.8	1145-1157	SNNIFSFYTQFLR		+
545.3	1158-1162	GSNLR		-
761.4	1163-1169	DQICGVK		?
1477.8	1171-1182	MLNDYIVIVWGR		+
923.5	1183-1190	SELDSFVK		?
1041.5	1191-1200	FAFACLSAGR		+
2991.5	1201-1227	MLTIDLPNIDVDDTEPLLNA LDSLVPR		+
874.5	1228-1235	IGSELSNR		?
1281.6	1239-1249	FLIYANENNGK		+
529.3	1254-1257	LLER		-
674.4	1260-1265	SQISLK		-
1284.6	1269-1278	FNYIFHTESK		+
3678.7	1279-1309	EWLIQTIINEDTGFHDDITDNDIYNTISEVR		+

Table 3.2 – In silico Trypsin digestion of Rad9 displaying all >500 Da peptides. Residues belonging to the C-terminal tag linker are indicated in parentheses. Peptide masses (Da), position numbers, and putative CDK sites are indicated. The comments section refers to peptides predicted to be detected (+) or not detected (-), with (?) indicating uncertainty of detection.

3.3.3. Analysis of Rad9 by LCMS

Using Lys-C and tryptic Rad9 digests, ~70% of the Rad9's 1309 residues were detected in each cell cycle type, including 19 of the 20 putative CDK phosphorylation sites (**Figure 3.2**).

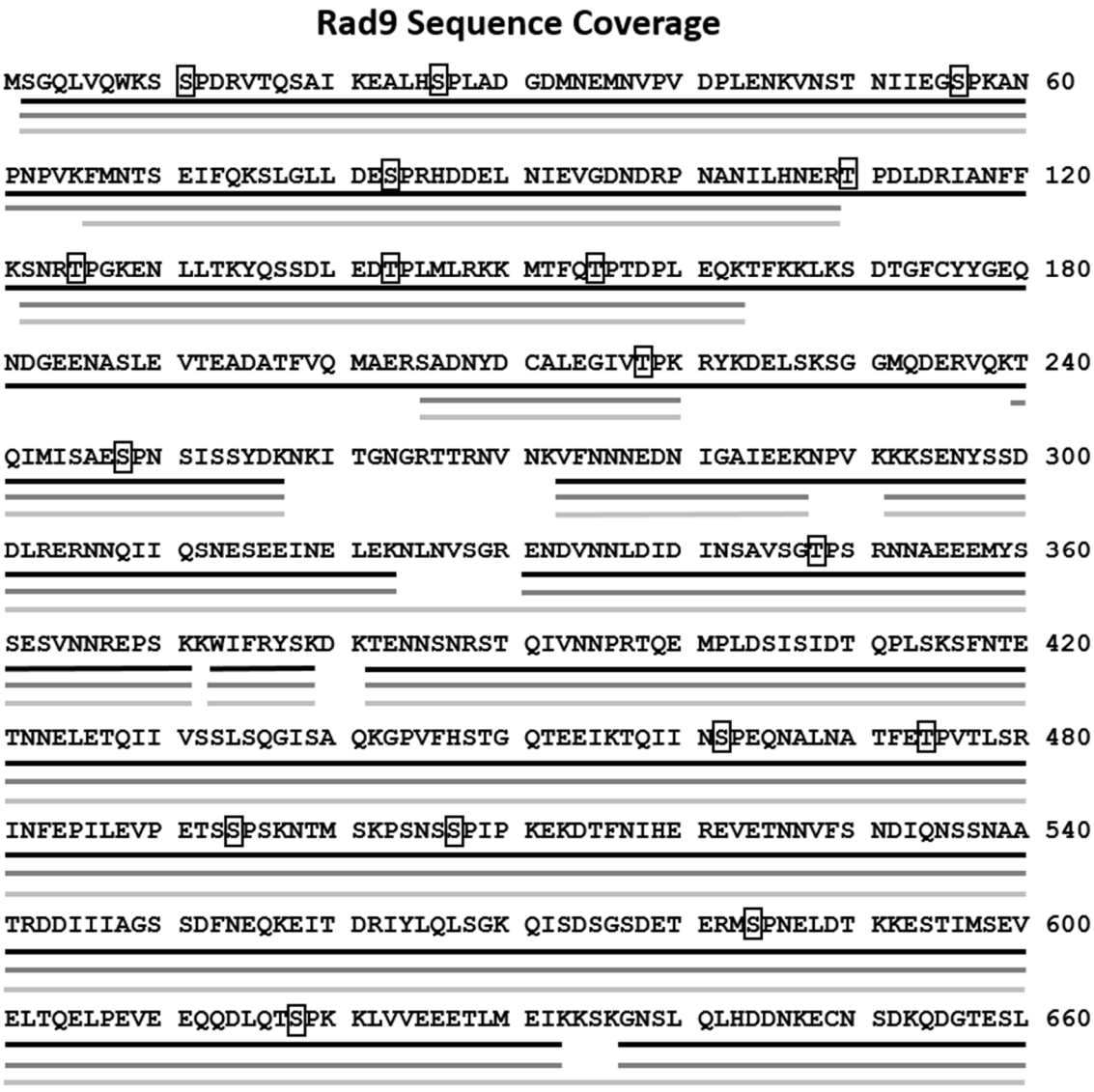


Figure 3.2 – *Sequence coverage of trypsin and Lys-C-digested Rad9 from asynchronous (black), G1- (dark gray), and G2-arrested (light gray) cells. Asynchronous coverage includes subdigestion of Lys-C with Asp-N. Putative CDK phosphorylation sites are boxed.*

The remaining CDK site, T110 located in the CAD, was detected by subdigestion of Lys-C-generated peptides with Asp-N. CDK site T110 was not detectable in trypsin or Lys-C digests, because the peptides containing it did not sufficiently elute (Lys-C) or bind to the C18 (trypsin). The asynchronous Rad9 peptide SLGLLDESPRHDDELNIEVGDNDRPNANILHNERTPDLDRIANFFK, generated from Lys-C, was subdigested with Asp-N to produce the detectable peptide, DRPNANILHNERTP. All 20 CDK-site-containing peptides were found to be phosphorylated in one

or more cell cycle forms of Rad9 and phosphorylation was unambiguously localized to 19 of these 20 sites. It is likely that the undetermined site, T348, was also phosphorylated because a phosphate was detected on the C-terminal half of peptide ENDVNNLDIDIN(SAVSGTPS)R as indicated in parentheses, and T348 is located in this region. With the combination of Lys-C, Lys-C-Asp-N, and tryptic digests, peptides containing all 20 CDK sites were detected and found to be phosphorylated (**Table 3.3**).

Residues	Peptide sequence	# Phosphorylation Sites	Phosphorylation Site(s)	Relative Abundance[a]		
				Asynch	G1	G2
10-21	S**S**PDRVTQSAIK	1	S11	++	++	++++
2-21	SGQLVQWKS**S**PDRVTQSAIK	1	S11	++	++	++++
2-21	SGQLVQWK**SS**PDRVTQSAIK	2	S10 & S11	+	+	++
2-21	SGQLVQWKS**S**PDRV**T**QSAIK	2	S11 & T16	-	-	+
22-46	EALH**S**PLADGDMNEMNVPVDPLENK	1	S26	+++	++	+++
47-58	VNSTNIIEG**S**PK	1	S56	++++	++++	++++
47-65	VNSTNIIEG**S**PKANPNPVK	1	S56	++++	++++	++++
76-85	SLGLLDE**S**PR	1	S83	++	++	++
76-109	SLGLLDE**S**PRHDDELNIEVGDNDRPNANILHNERT	1	S83	++	++	++
76-109	DRPNANILHNER**T**P	1	T110	++	ND	ND
76-109	DRPNANILHNER**T**PDL	1	T110	++	ND	ND
122-134	SNR**T**PGKENLLTK	1	T125	++++	++++	++++
135-149	YQ**S**SDLEDTPLMLRK	1	T137	+[b]	+[b]	++
135-149	YQSSDLED**T**PLMLRK	1	T143	+[b]	+[b]	+
135-149	YQ(SSDLEDT)PLMLRK	2	[d]	+[b]	-	+[b]
150-163	KM**T**FQTPTDPLEQK	1	T152	+	+[b]	+[b]
150-163	KMTFQ**T**PTDPLEQK	1	T155	+	+[b]	++
205-220	**S**ADNYDCALEGIVTPK	1	S205	+[b]	+	+
205-220	SADNYDCALEGIV**T**PK	1	T218	++	++	++

205-220	**S**ADNYDCALEGIV**T**PK	2	T218 & S205	+[b]	+[b]	+
240-257	TQIMISAE**S**PNSISSYDK	1	S248	+++	+++	++++
240-257	TQIMISAE**S**PN**S**ISSYDK	2	S248 & S251	++[b]	+	+[b]
240-257	(TQIMISAESPNSISSY)DK	3	[d]	+	+[b]	+[b]
294-305	SENY**S**SDDLRER	1	S298	++	++	++
294-305	SENYS**S**DDLRER	1	S299	++	+	+
294-305	SENY**SS**DDLRER	2	S298 & 299	++	+	+
306-323	NNQIIQ**S**NESEEINELEK	1	S312	++	-	++
306-323	NNQIIQSNE**S**EEINELEK	1	S315	+	-	+
306-323	NNQIIQ**S**NE**S**EEINELEK	2	S312 & S315	+	-	++
331-351	ENDVNNLDIDIN(SAVSGTPS)R	1	[d]	+[b]	-	+[b]
352-371	NNAEEEM(YSSESVNNREPS)K	1	[d]	+[b]	+[b]	+[b]
352-371	NNAEEEMY**SS**ESVNNREPSK	2	S360 & S361	+[b]	+[b]	+
382-415	(TENNSNRSTQIVNNPRTQEMPLDSISIDTQPLS)K	1	[d]	+[b]	-	+[b]
443-456	GPVFH(STGQT)EEIK	1	[d]	+[b]	+[b]	+[b]
457-480	TQIIN**S**PEQNALNATFETPVTLSR	1	S462	++	++	++
457-480	TQIINSPEQNALNA**T**FE**T**PVTLSR	1	T471 & 474	+[b]	+[b]	+
457-480	TQIIN**S**PEQNALNATFE**T**PVTLSR	2	S462 & T474	-	+	+
481-497	INFEPILEVPET**S**SPSK	1	S493	++	+	++
481-497	INFEPILEVPETS**S**PSK	1	S494	++	+++	++
481-497	INFEPILEVPET**SS**PSK	2	S493 & S494	+	-	-
481-497	INFEPILEVPE(TS)**S**PSK	2	T492/ S493[d], & S494	-	+	-
481-497	INFEPILEVPE**TSS**PSK	2	T492, S493, & S494	-	-	++
498-511	NTMSKPSN**S**SPIPK	1	S506	+	+	+
498-511	NTMSKPSNS**S**PIPK	1	S507	+++	+++	++
498-511	NTMSKPSN**SS**PIPK	2	S506 & S507	++	+[b]	+
512-521	EKD**T**FNIHER	1	T515	-	+[b]	+
543-557	DDIIIAG(SS)DFNNEQK	1	[d]	+[b]	+[b]	+[b]

558-570	EI(TDRIYLQLS)GK	1	[d]	-	-	+[b]
571-582	QISD**S**GSDETER	1	S575	+[e]	+[b]	+
571-591	QISDSGSDETERM**S**PNELDTK	1	S584	++	++	+
571-591	QISD(SGS)DETERM**S**PNELDTK	2	S575/ S577[d] & S584	+	+	+
571-591	QI(SDSGSDETERMSPNELDT)K	3	[d]	+[b]	+[b]	+[b]
593-620	ESTIMSEVELTQELPEVEEQQDLQT**S**PK	1	S618	++	+++	++++
635-647	(SKGNS)LQLHDDNK	1	[d]	-	-	+[b]
637-670	ECN**S**DKQDGTESLDVALIEHESK/ GNSLQLHDDNKECN**S**DK	1	S651	+	+	+
654-670	QDG(TESLDVALIEHES)K	1	[d]	+[b]	+[b]	+[b]
679-699	NLMQLFPSE**S**QEIIQNRRTIK	1	S688	+	+	+
704-718	DTIEIGEEEENR**S**TK	1	S716	+	+	++
704-723	DTIEIGEEEENRSTKT(SPT)K	1	[d]	+[be]	+[e]	+[be]
704-723	DTIEIGEEEENR**S**TKT**S**PTK	2	S716 & 720	+[be]	+[e]	+[be]
704-723	D(TIEIGEEEENRSTKTSPT)K	3	[d]	+[be]	+[be]	-
727-738	RN**S**DLDDAASIKR	1	S729	++	++	++
727-738	RNSDLDDAA**S**IKR	1	S735	-	-	+
738-755	REP(SCSITIQTGETGS)GK	1	[d]	+[b]	+[b]	+[b]
866-873	G(YDT)VHLK	1	[d]	-	-	+[b]
931-940	**S**MTNVLSPKK	1	S931	+	++	++
931-940	SMTNVL**S**PKK	1	S937	++	++	++
931-940	**S**MTNVL**S**PKK	2	S931 & S937	+[b]	+[b]	+
931-946	SMTNVLSPKKH**T**DDEK	1	T943	+[e]	+[e]	+[e]
931-946	SMTNVL**S**PKKH**T**DDEK	2	S937 & T943	+[e]	+[e]	+[e]
931-946	(SMTNVLS)PKKH**T**DDEK	2	S931/ T933/ S937[d] & T943	+[be]	+[be]	+[e]
947-966	DINTHTEVYNNEIE**S**SSEKK	1	S961	+	+	+[b]
947-966	DINTHTEVYNNEIE(**SS**S)EKK	1	S961 & 963	-	+	+
947-966	DINTHTEVYNNEIE**SSS**EKK	2	S961, 962, & 963	+[b]	+[b]	+
975-994	DALAEHAGAPSLLF**SS**GEIR	1	S989 & S990	+[b]	-	+

975-994	DALAEHAGAP**S**LLF**SS**GEIR	1	S985, S989, & S990		+	-
1132-1139	LSLD**S**P**S**K	1	S1136 & S1138	ND[c]	+	+[b]

Table 3.3 - Rad9 phosphorylation sites assigned by manual interpretation of collision-activated dissociation and/or electron transfer dissociation mass spectra. [a]Phosphorylated peptide levels are reported as % total peptide abundance detected, (++++) = 70–100%, (+++) 35–69%, (++) = 6–34%, (+) = 1–5%, (−) no phosphorylated peptide detected. % Total abundances were calculated from ion currents observed in the main beam mass spectra [(phosphorylated peptide ion current)/ (phosphorylated peptide + non-phosphorylated peptide ion current) × (100)]. [b]Phosphorylated peptide containing this phosphorylation site was assigned by HPLC retention time and correct accurate mass, only. [c]Non-phosphorylated and phosphorylated peptides were not detected (ND). [d]PTM site is located on S/T/Y within parentheses. Putative CDK sites are highlighted.

A total of 63 unique Rad9 phosphorylation sites were detected amongst the three cell cycle samples and the majority were detected in all three. The collective trend mostly showed phosphorylation levels in equal abundances throughout the cell cycle or increasing in G2. Over half of the phosphorylation sites detected were previously unknown, including 7 CDK. Of the 63 phosphorylation sites detected, 50 were unambiguously site-mapped (**Table 3.4**).

Phosphorylation Site(s)	Literature Sites	This Analysis
S10	**X**	**X**
S11	**X**	**X**
T16		**X**
S26	**X**	**X**
T50	**X**	
S56	**X**	**X**
S70	**X**	
S76	**X**	
S83	**X**	**X**
T110		**X**
T125		**X**
S137	**X**	**X**
S138	**X**	**X**
T143		**X**

T152		X
T155	X	X
S188	X	
S205	X	X
T218	X	X
S248	X	X
S251		X
S312	X	X
S315	X	X
S328	X	
T348	X	?
S360		X
S361		X
T398	X	
S414	X	
S416	X	
S435	X	
T457	X	
S462	X	X
T471	X	X
T474		X
S479	X	
T492		X
S493		X
S494		X
S506		X
S507	X	X
T515		X
S537	X	
T541	X	
S550	X	
S568	X	
S575		X
S584		X
S598	X	
T603	X	
S618	X	X

S651	**X**		
S688	**X**	**X**	
S716		**X**	
S720		**X**	
S729	**X**	**X**	
S735		**X**	
S741	**X**		
T807	**X**		
S931	**X**	**X**	
S937	**X**	**X**	
T943		**X**	
S961		**X**	
S962		**X**	
S963		**X**	
S985		**X**	
S989	**X**	**X**	
S990	**X**	**X**	
S1136	**X**	**X**	
S1138		**X**	
	Total	**Total**	
	45	**50**	
		Known	**Novel**
		23	**27**

Table 3.4 – Previously known phosphorylation sites [28, 29] and those detected in this analysis are indicated by X. This analysis revealed 27 previously unknown phosphorylation sites on Rad9, including 7 CDK sites. All CDK sites are highlighted.

All 9 CAD CDK sites were observed to be phosphorylated. Phosphorylation of S11 appeared to act constitutively with S10 or T16 phosphorylation. CDK-dependent phosphorylation of S11 is critical for the interaction between Rad9 and Dpb11, a scaffold/mediator protein required for Rad9's checkpoint activity in mitosis [30]. CDK phosphorylation of Rad9 S11 may enhance the interaction of Rad9 with kinases responsible for phosphorylating S10 and/or T16. This indicates that Dpb11-dependant Rad9 checkpoint activity requires more kinases than CDK alone. Phosphorylation on CDK sites T110, T125, and T143 was previously unreported, but detected on each in this study. There was a sharp contrast in the

abundance of phosphorylation between T125 and T143 as the levels were nearly stoichiometric in all samples for the former and <5% in all samples for the latter. Phosphorylation was also localized on S137 in the absence of T143. A di-phosphorylated form of peptide YQSSDLEDTPLMLRK, containing both modifications, was detected in similar abundance. It is likely that both modifications are present together, but the sites could not be unambiguously localized in the di-phosphorylated form. Another formerly unreported phosphorylation site, T152, was detected on the same peptide containing CDK site T155, though phosphorylation of one was only detected in the absence of the other. Modification of these sites may be mutual exclusive of one another.

Novel phosphorylation was detected within residues 448-452 in the SQ/TQ cluster domain (SCD) of Rad9. SCD domains are recognized as a core characteristic of DNA-damage-response proteins and serve as templates for DNA-damage-induced phosphorylation by ATM/ATR kinases [31]. This site does not fit the SQ/TQ motif, however. Another SCD phosphorylation site was detected amongst residues 382-414, which contains the TQ-consensus site T398, previously reported to be phosphorylated.

Two novel phosphorylation sites were detected on Rad9's Tudor domain. This domain is essential for Rad9's ability to interacting with chromatin, and both single- and double-stranded DNA during checkpoint activation and DNA repair [32, 33]. Phosphorylation sites T942 and Y867/T69 were previously unknown.

The C-terminus of Rad9 contains a BRCA1 carboxyl terminus (BRCT) domain which is essential for its function at least in part by facilitating Rad9 oligomerization [34]. Novel phosphorylation was detected on S1138 and on the known site, S1136, in the absence of the former.

Rad9 is heavily phosphorylated throughout the length of its sequence and on all four defined functional domains (**Figure 3.3**). Novel sites were also detected throughout Rad9 and on all domains including the most heavily modified, CAD, critical for Chk1 activation.

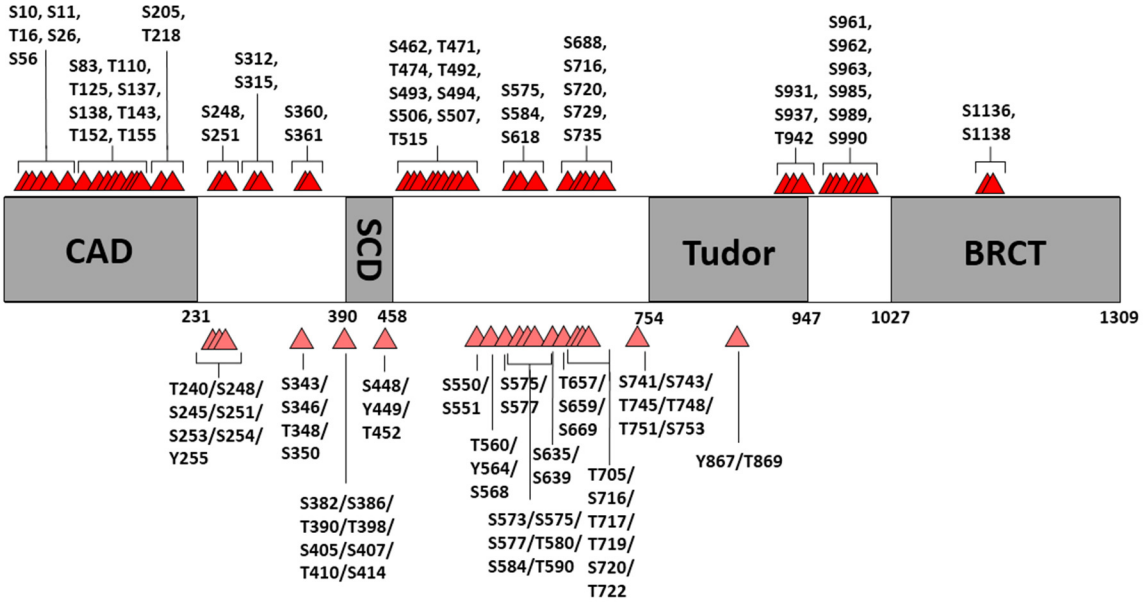


Figure 3.3 – *Schematic of all Rad9 phosphorylation sites detected in this study. Dark red triangles (top) indicate all unambiguously localized sites, and the light red triangles (bottom) indicate the regional localization of all additional phosphorylation sites detected.*

3.3.4 Biological role of CAD phosphorylation in Chk1 activation

Molecular biology experiments were conducted by Lowndes and coworkers to investigate the function of CDK and its phosphorylation sites on Rad9 CAD. CDK sites 1-9 located in the CAD (**Figure 3.4A**) were confirmed to be phosphorylated by cdc28 by *in vitro* phosphorylation by four different cdc28 cyclin complexes (**Figure 3.4B**) [24]. Complex cdc28/clb2 exhibited the highest activity.

Next, the effect of DNA damage-induced phosphorylation of CAD CDK sites on Chk1 activation was examined *in vivo*. Tagged Chk1-3HA was detected by western blot from

asynchronous, G1-, and G2/M-arrested cells treated with bleocin. Rad9 CDK sites 1-9 were mutated to alanine and Chk1 activation was compared to cells containing wild-type Rad9. While some residual Chk1 activity remained in G1-arrested cells, Chk1 activity was eliminated in G2/M cells following DNA-damage as observed by its decrease in phosphorylation (**Figure 3.4C-E**) [24].

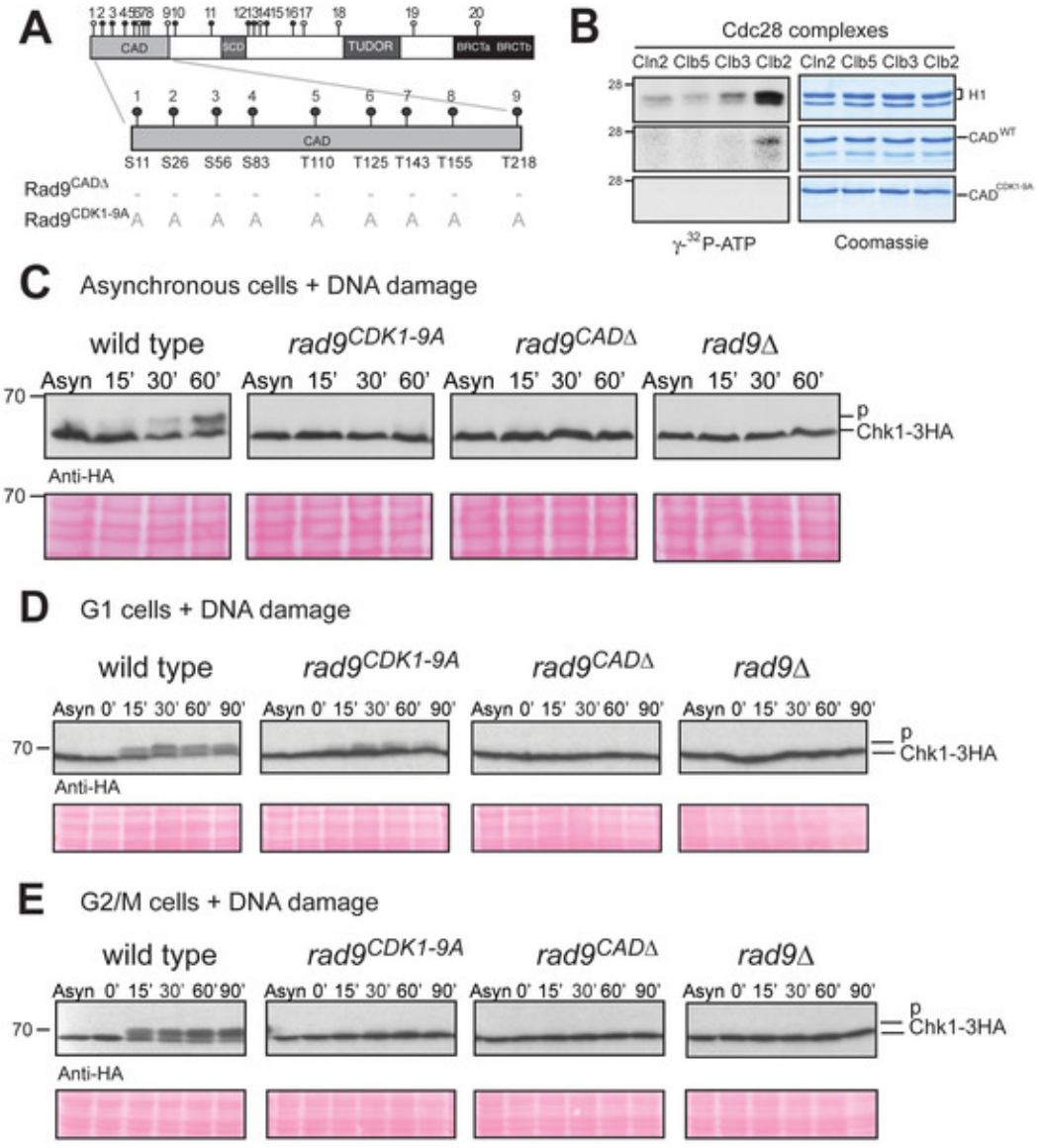


Figure 3.4 – *(A) Schematic of Rad9 CAD highlighting Cdc28 consensus sites 1-9 mutated to Ala in Rad9$^{CDK1-9A}$ protein. (B) Cdc28/Clb2 produces the highest level of in vitro phosphorylation of Rad9 CAD^{WT}. (C) DNA-damage-dependent phosphorylation of Chk1 is reduced or eliminated in Rad9$^{CDK1-9A}$, Rad9^{CADΔ}, Rad9Δ cells. (D&E) – Chk1 activation is principally dependent on CDK consensus sites 1-9 in G2/M cells. Asynchronously-growing cells were treated with bleocin for the indicated times and Chk1 phosphorylation was indicated by anti-HA western blot. G1 and G2/M cells were arrested by α–factor and nocodazole, respectively. Figure reproduced from [24].*

Furthermore, CDK-dependent activation of Chk1 was found to control Rad9/Chk1 interaction. It was previously known that Chk1 activation required Cdc28 activity in response to double-strand breaks [35, 36]. In this study, *Cdc28-as1* cells which can be rapidly inhibited, arrested, and their DNA damaged, were used to investigate whether Cdc28 can activate Chk1 independently of CDC28-dependent regulation of double-strand breaks [24]. Cells were arrested in G2/M, DNA-damaged, and several components of the DDR were examined by western blot for phosphorylation in the presence and absence of Cdc28 activity. 4-NQO was used to induce single-strand breaks and DNA adducts while bleocin induced double-strand breaks. Cdc28 activity was required for Chk1 activation in response to either agent. Rad9 and Rad53 were activated independently of Cdc28 activity in response to 4-NQO, and consistent with previous studies, Rad53 activation was dependent on Cdc28 in response to bleocin-induced double-strand breaks (**Figure 3.5**) [36]. These results provided further evidence that Cdc28-dependent phosphorylation of Rad9 is required for Chk1 activation.

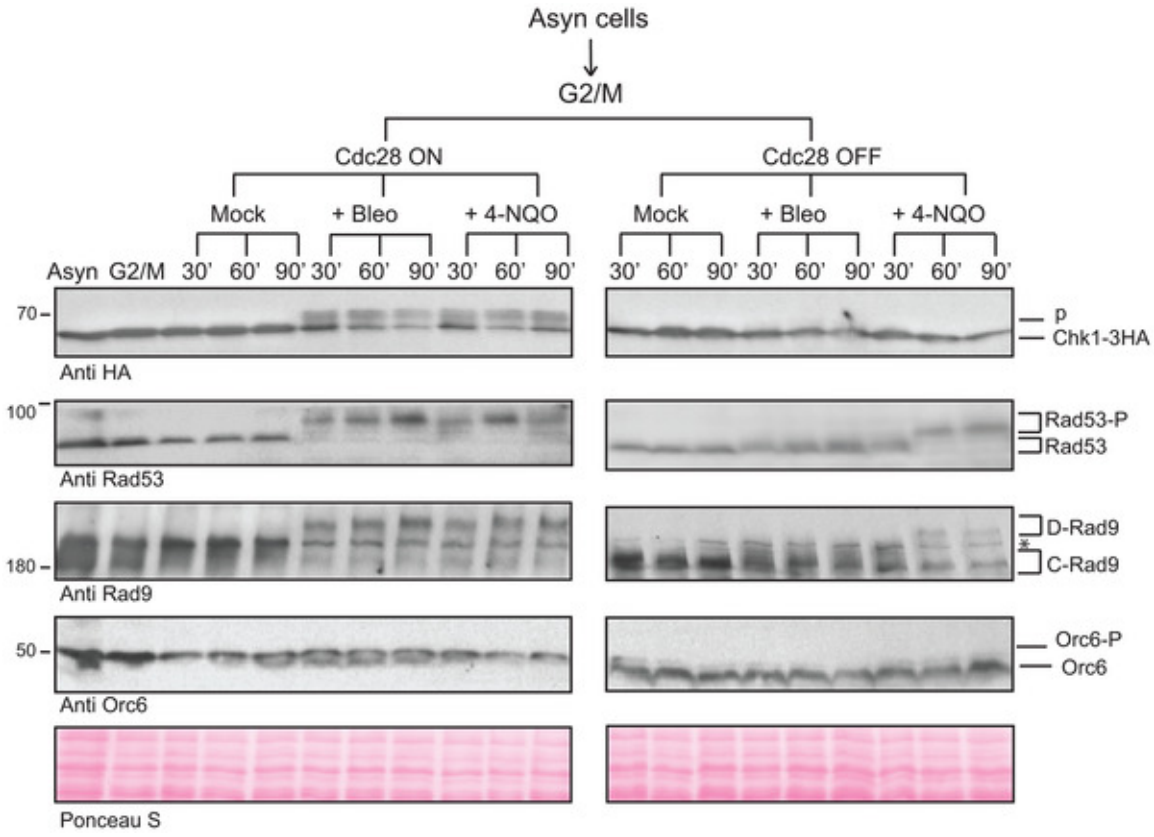


Figure 3.5 – *Chk1 activity was dependent on Cdc28 upon DNA damage separately induced by bleocin or 4-NQO. Otc6 phosphorylation marked Cdc28 inactivation. Figure reproduced from [24].*

The mechanism of Rad9-dependent activation of Chk1 was further investigated by yeast two-hybrid assays [24]. The necessity of CDK sites 1-9 for Rad9/Chk1 interaction was tested in G1- and G2/M-arrested cells (**Figure 3.6A**), and they were required in both cell stages. This interaction was also Cdc28-dependent in G2/M yet was Cdc28-independent in G1 (**Figure 3.6B**). A reciprocal co-immunoprecipitation experiment validated these results by showing that a portion of Rad9 co-immunoprecipitated with Chk1 and *vice versa*, and this was abolished by Rad9$^{CDK1\text{-}9A}$ (data not shown).

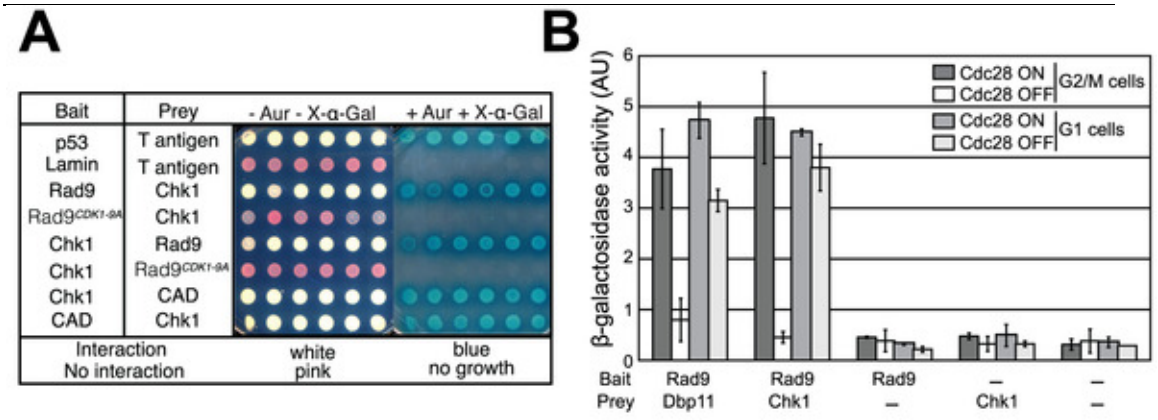


Figure 3.6 – *(A) Rad9/Chk1 interaction was dependent on CDK1-9 based on yeast two-hybrid assay in vivo. (B) Rad9/Chk1 interaction was dependent on Cdc28 in G2/M and was Cdc28-independent in G1. Figure reproduced from [24].*

Yeast two-hybrid and peptide pull-down experiments were lastly used to determine which specific CAD CDK sites facilitated Rad9/Chk1 interaction. CAD 1-231 mutant constructs were generated in which CDK1-9 sites were mutated to Ala and individually mutated back to the wild type residue. The ability of each 'reversion' CAD mutant to interact with Chk1 was assessed by yeast two-hybrid assay (**Figure 3.7A**). The most prominent rescues of CAD/Chk1 interaction were made by reversions of T125A and T143A to wild-type. Phosphorylated and non-phosphorylated 35-amino acid peptides containing T125 and T143 were constructed to test their ability to pull-down Chk1. The pT143 peptides pulled Chk1 down most efficiently while peptides containing pT125 demonstrated much reduced interaction independently of T143 phosphorylation (**Figure 3.7B**). This suggests that phosphorylated of T143 is a stimulator of Rad9/Chk1 interaction and phosphorylation of T125 is an inhibitor [24]. Mass spectrometric analysis of Rad9 indicated that T125 was heavily phosphorylated *in vivo* and only a small fraction of Rad9 protein was phosphorylated on T143. This suggests that phosphorylation of a small population of Rad9 is responsible for the DNA-damage-independent interaction of Rad9 and Chk1, and is consistent with the co-immunoprecipitation experiments in which only a minority

of Rad9 bound to Chk1 and *vice versa*.

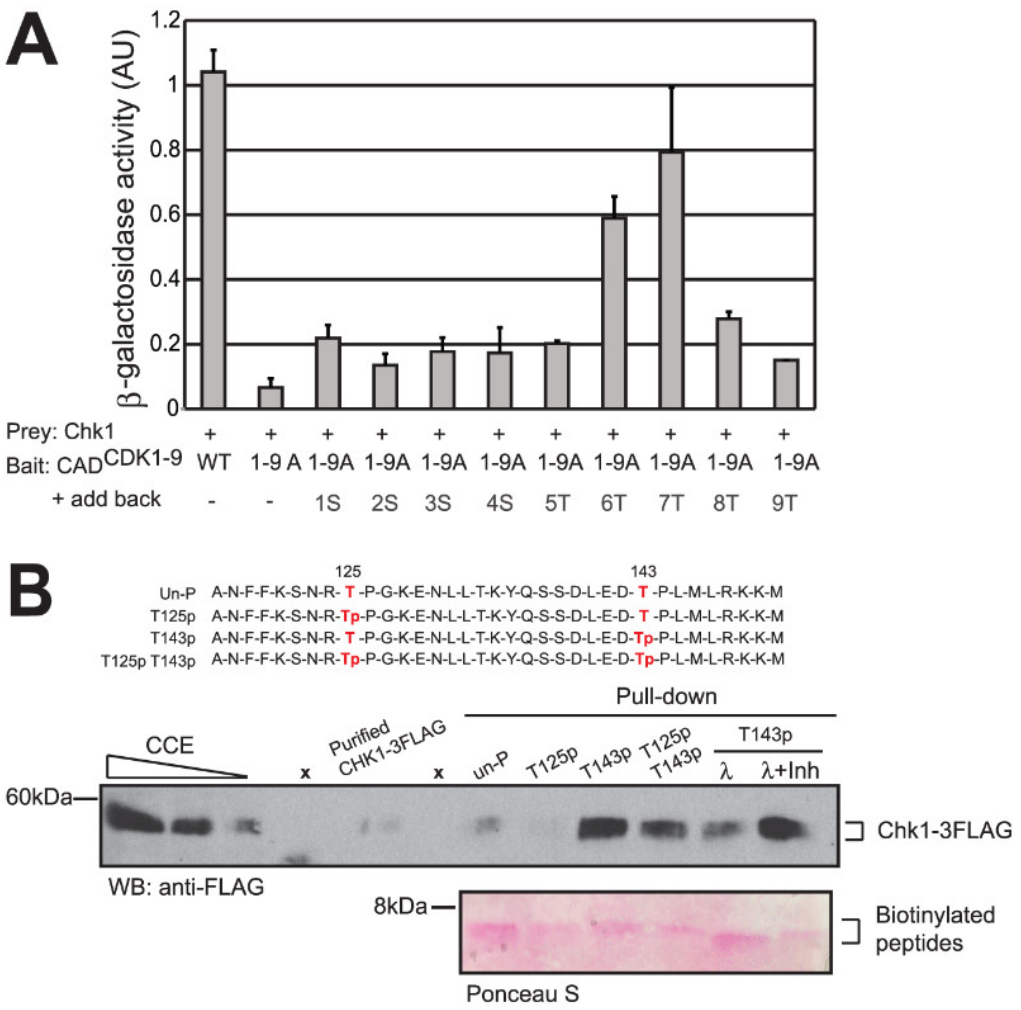


Figure 3.7 – *(A) Yeast-two hybrid assay demonstrated that T125 and T143 most contributed to CAD/Chk1 interaction of CDK1-9 sites. (B) – Peptides containing pT143 most effectively pulled down Chk1 while pT125 showed an inhibitory effect. Figure reproduced from [24].*

3.4 Conclusion

Mass spectrometric analysis of DNA damage mediator Rad9 revealed that 63 residues were *in vivo* phosphorylated independent of DNA damage (**Table 3.5**) and 34-40 of these were previously unreported.

Literature Sites	This study			
	Localized to residue		Localized to peptide	
Total	Total		Total	
45	50		13	
	Known	Novel	Novel	Undetermined
	23	27	7	6
	Summary			
	Known	Novel	Undetermined	Total
	23	34	6	63

Table 3.5 – 50 of 63 phosphorylation sites detected were localized to the Ser/Thr residue modified and 27 of these were previously unreported. In total, 34-40 unreported phosphorylation sites were detected.

This represents the largest number of phosphorylation sites reported on Rad9 in any single study. Phosphorylation was detected on CDK sites 1-9 in the Rad9 Chk1 activation domain and phosphorylation was ambiguously localized to all but one of the 10 CDK sites located outside this domain as well. Phosphorylation of CDK sites T125 and T143 was found to be especially critical for Chk1 activation by regulating the interaction between Rad9 and Chk1. This study has further elucidated the role of Rad9 and the mechanism by which its Chk1 activating domain regulates Rad/Chk1 interaction through Cdc28/Clb-dependent phosphorylation. Abreu et al.

propose a novel model of Chk1 activation in response to DNA damage based on these findings[24].

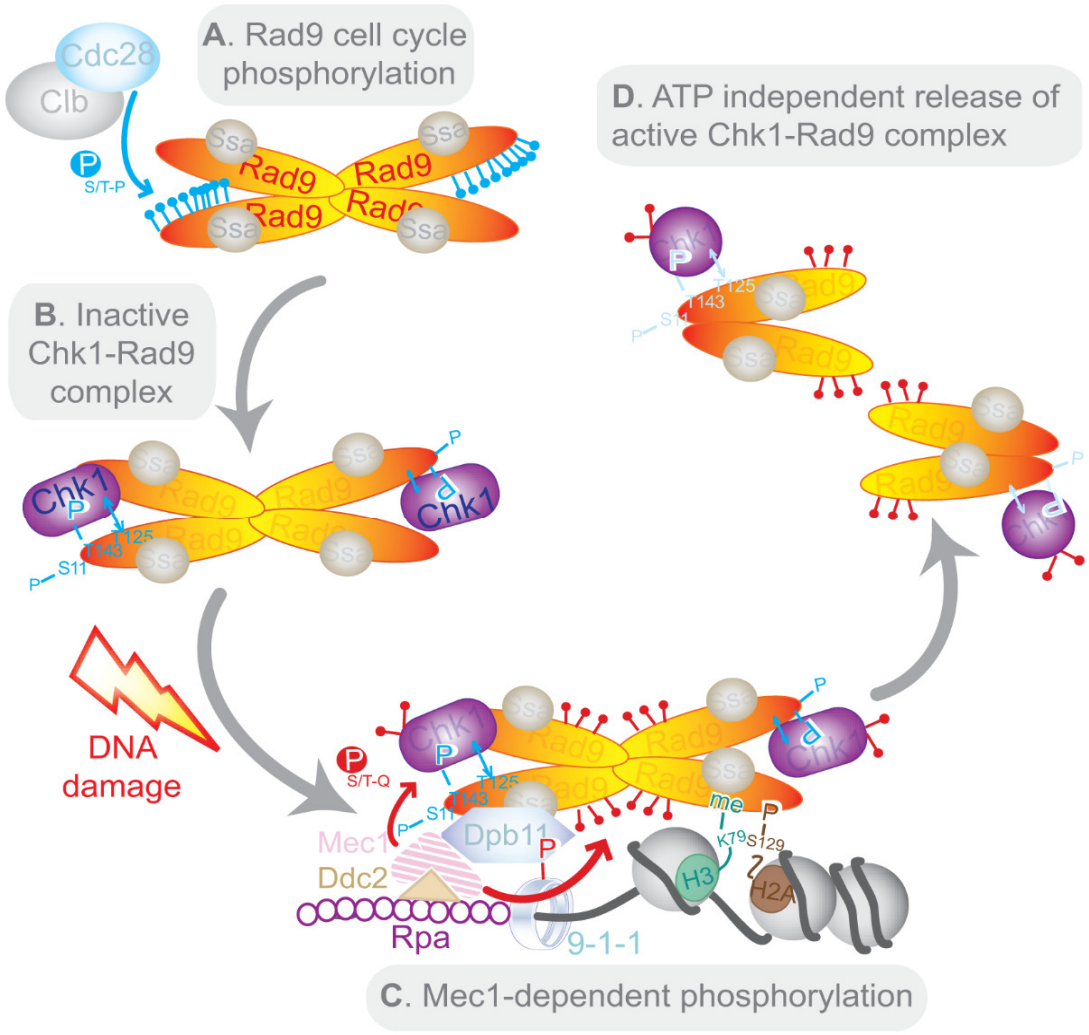


Figure 3.8 – *Model of Chk1 activation in response to DNA damage. Rad9 CAD phosphorylation by Cdc28/Clb (A) regulates Rad9/Chk1 interaction independent of DNA-damage and is most promoted by pT143 and no phosphorylation on T125 (B). Rad9 complexes are recruited to sites of DNA damage and recognize specific histone marks. Other Rad9 phosphorylation sites stimulate its interaction with Dpb11, recruited in a 9-1-1-complex-dependent manner (C). These together enhance Mec1 activity and remodeling of Rad9 and Chk1, resulting in the in cis autophosphorylation of Chk1 and release of the hyper-phosphorylated Rad9/Chk1 complex. In contrast to Rad9/Rad53 complexes, Rad9-Chk1 release was found to then proceed in an ATP-independent manner (D). Figure reproduced from [24].*

3.5 References

1. Harper, J. W. & Elledge, S. J. The DNA damage response: ten years after. *Mol. Cell* **28**, 739-745 (2007).

2. Halazonetis, T. D., Gorgoulis, V. G. & Bartek, J. An oncogene-induced DNA damage model for cancer development. *Science* **319**, 1352-1355 (2008).

3. Durocher, D. & Jackson, S. P. DNA-PK, ATM and ATR as sensors of DNA damage: variations on a theme? *Curr. Opin. Cell Biol.* **13**, 225-231 (2001).

4. Stracker, T. H., Usui, T. & Petrini, J. H. J. Taking the time to make important decisions: The checkpoint effector kinases Chk1 and Chk2 and the DNA damage response. *DNA Repair* **8**, 1047-1054 (2009).

5. Saka, Y., Esashi, F., Matsusaka, T., Mochida, S. & Yanagida, M. Damage and replication checkpoint control in fission yeast is ensured by interactions of Crb2, a protein with BRCT motif, with Cut5 and Chk1. *Genes Dev.* **11**, 3387-3400 (1997).

6. FitzGerald, J. E., Grenon, M. & Lowndes, N. F. 53BP1: function and mechanisms of focal recruitment. *Biochem. Soc. Trans.* **37**, 897-904 (2009).

7. Jungmichel, S. & Stucki, M. MDC1: The art of keeping things in focus. *Chromosoma* **119**, 337-349 (2010).

8. Weinert, T. A. & Hartwell, L. H. The RAD9 gene controls the cell cycle response to DNA damage in Saccharomyces cerevisiae. *Science* **241**, 317-322 (1988).

9. O'Shaughnessy, A. M. *et al*. in *Methods in Enzymology* 131-150 (Academic Press.

10. Finn, K., Lowndes, N. F. & Grenon, M. Eukaryotic DNA damage checkpoint activation in response to double-strand breaks. *Cell Mol. Life Sci.* **69**, 1447-1473 (2012).

11. Usui, T., Foster, S. S. & Petrini, J. H. Maintenance of the DNA-damage checkpoint requires DNA-damage-induced mediator protein oligomerization. *Mol. Cell* **33**, 147-159 (2009).

12. Gilbert, C. S., Green, C. M. & Lowndes, N. F. Budding yeast Rad9 is an ATP-dependent Rad53 activating machine. *Mol. Cell* **8**, 129-136 (2001).

13. Pellicioli, A. & Foiani, M. Signal transduction: how rad53 kinase is activated. *Curr. Biol.* **15**, R769-71 (2005).

14. Tapia-Alveal, C., Calonge, T. M. & O'Connell, M. J. Regulation of chk1. *Cell. Div.* **4**, 8-1028-4-8 (2009).

15. Blankley, R. T. & Lydall, D. A domain of Rad9 specifically required for activation of Chk1 in budding yeast. *J. Cell. Sci.* **117**, 601-608 (2004).

16. Rappold, I., Iwabuchi, K., Date, T. & Chen, J. Tumor suppressor p53 binding protein 1 (53BP1) is involved in DNA damage-signaling pathways. *J. Cell Biol.* **153**, 613-620 (2001).

17. Vialard, J. E., Gilbert, C. S., Green, C. M. & Lowndes, N. F. The budding yeast Rad9 checkpoint protein is subjected to Mec1/Tel1-dependent hyperphosphorylation and interacts with Rad53 after DNA damage. *EMBO J.* **17**, 5679-5688 (1998).

18. Ubersax, J. A. *et al*. Targets of the cyclin-dependent kinase Cdk1. *Nature* **425**, 859-864 (2003).

19. Morgan, D. O. Cyclin-dependent kinases: engines, clocks, and microprocessors. *Annu. Rev. Cell Dev. Biol.* **13**, 261-291 (1997).

20. Nasmyth, K. At the heart of the budding yeast cell cycle. *Trends in Genetics* **12**, 405-412 (1996).

21. Albuquerque, C. P. *et al*. A multidimensional chromatography technology for in-depth phosphoproteome analysis. *Mol. Cell. Proteomics* **7**, 1389-1396 (2008).

22. Holt, L. J. *et al*. Global analysis of Cdk1 substrate phosphorylation sites provides insights into evolution. *Science* **325**, 1682-1686 (2009).

23. Smolka, M. B. *et al*. Dynamic changes in protein-protein interaction and protein phosphorylation probed with amine-reactive isotope tag. *Mol. Cell. Proteomics* **4**, 1358-1369 (2005).

24. Abreu, C. M. *et al*. Site-specific phosphorylation of the DNA damage response mediator rad9 by cyclin-dependent kinases regulates activation of checkpoint kinase 1. *PLoS Genet.* **9**, e1003310 (2013).

25. Hall, E. H. *et al*. Comprehensive analysis of phosphorylation sites in Tensin1 reveals regulation by p38MAPK. *Mol. Cell. Proteomics* **9**, 2853-2863 (2010).

26. Udeshi, N. D., Compton, P. D., Shabanowitz, J., Hunt, D. F. & Rose, K. L. Methods for analyzing peptides and proteins on a chromatographic timescale by electron-transfer dissociation mass spectrometry. *Nat. Protoc.* **3**, 1709-1717 (2008).

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

28. Smolka, M. B., Albuquerque, C. P., Chen, S. H. & Zhou, H. Proteome-wide identification of in vivo targets of DNA damage checkpoint kinases. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 10364-10369 (2007).

29. Albuquerque, C. P. *et al*. A multidimensional chromatography technology for in-depth phosphoproteome analysis. *Mol. Cell. Proteomics* **7**, 1389-1396 (2008).

30. Granata, M. *et al*. Dynamics of Rad9 chromatin binding and checkpoint function are mediated by its dimerization and are cell cycle-regulated by CDK1 activity. *PLoS Genet.* **6**, 10.1371/journal.pgen.1001047 (2010).

31. Traven, A. & Heierhorst, J. SQ/TQ cluster domains: concentrated ATM/ATR kinase phosphorylation site regions in DNA-damage-response proteins. *Bioessays* **27**, 397-407 (2005).

32. Grenon, M. *et al*. Docking onto chromatin via the Saccharomyces cerevisiae Rad9 Tudor domain. *Yeast* **24**, 105-119 (2007).

33. Lancelot, N. *et al*. The checkpoint Saccharomyces cerevisiae Rad9 protein contains a tandem tudor domain that recognizes DNA. *Nucleic Acids Res.* **35**, 5898-5912 (2007).

34. Soulier, J. & Lowndes, N. F. The BRCT domain of the S. cerevisiae checkpoint protein Rad9 mediates a Rad9–Rad9 interaction after DNA damage. *Current Biology* **9**, 551-S2 (1999).

35. Enserink, J. M. & Kolodner, R. D. An overview of Cdk1-controlled targets and processes. *Cell. Div.* **5**, 11-1028-5-11 (2010).

36. Ira, G. *et al*. DNA end resection, homologous recombination and DNA damage checkpoint activation require CDK1. *Nature* **431**, 1011-1017 (2004).