

# **Characterization of Protein Sequence and Post-translational Modifications Using Mass Spectrometry**

Lichao Zhang  
Shandong, China

B.S., Chemistry, Nankai 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

May, 2014

---

This dissertation is dedicated to my Mom and Dad.

---

---

**Acknowledgments**

First of all I must thank my advisor, Professor Donald F. Hunt. Don, thank you so much for giving me the opportunity to conduct such exciting research in your group for the past few years. I have been always inspired by your enthusiasm for science. You were always kind and patient with me no matter how busy you were. You have done more than I could ask for and I feel so lucky to have you as my advisor. I also must thank Dr. Jeffery Shabanowitz. Jeff, thank you for your strong support in the lab. You have taught me how to think critically and looking at things from different perspectives. You were always there to help and encourage me. I always feel inspired by our conversations about either science or life. Also, thank you so much for making the lab a comfortable and challenging place to work. Don and Jeff, I am very proud to carry forward what I learned from both of you through my future careers.

I would also thank my collaborators. I have learned so much from each of you. I always tell people how much I enjoyed my collaboration experience. In particular, I would like to thank Dr. Kevin Pfister and Kiev Blasier for collaborating with me on the dynein project. Thank you for being so patient with me when I was first learning how to handle samples. I would also like to thank Dr. Kathleen Green and Lauren Albrecht for collaborating with me on the desmoplakin project. I really appreciate Lauren's hard work on preparing samples and performing those amazing biology experiments.

---

---

I would like to thank my committee members, Dr. David Cafiso, Dr. Cassandra Fraser, Dr. Linda Columbus and Dr. Kevin Pfister. I appreciate the time you spent reading my dissertation and your suggestions on the revisions.

I own a big thank you to all the Hunt group members. It has been a privilege to work with all of you. I feel so lucky to have the opportunity to work with everyone. Special thanks to Dr. Dina Bai for her help and support. I enjoyed chatting with you on variety of topics while you were fixing my computer. I would like to sincerely thank Dr. Michelle English. Michelle, since I first joined the group, until your last day in lab, you were always providing valuable advice and kind help. I have learned so much from you, anything related to science and beyond. Special thanks also to Dr. Weihang Wang. Weihang, I was amazed by your chemistry knowledge and endless ideas. Collaborating with you has been such a great experience. Your creativity and diligence have inspired me to work harder and think outside of the box. I must thank Drs. Jenny Cottine, Patrick James, and Jessica Chapman for taking time to teach me at the beginning of my graduate studies. Gratitude must also be expressed for the help and advice from Dr. Aaron Bailey. I have benefited a lot from your knowledge in cell biology and was inspired by your great work ethics. I also have to thank Drs. Erin Jeffery, Jeremy Balsbaugh, Andrea DeSantis, Andrew Norris, Sushmit Maitra, Josh Nicklay, Joe Strukl, and Linda Zhao. Each of you are such excellent scientist and I enjoyed being around you all while you were in the lab. Many thanks go to Dr. Andrew Dawdy, Dr. Jenn Abelin and soon-to-be Dr. Lissa Anderson. We joined the lab the same year and it has been wonderful to

---



have you all around while experiencing the same endeavors of graduate school. I

also need to thank Amanda Wriston, Stacy Malaker, Marshall Chaffee, Paisley

Trantham, Scott Ugrin and Ben Barnhill for the good times we spent together in the lab and you all have made this lab such a great place to work at every day.

I would like to express a sincere thank you to my internship supervisor Dr. Qiangwei Xia. You have generously shared with me your knowledge and experience in therapeutic drug characterization. Thank you for being so patient with me and always challenging me with interesting problems. Thank you for continuing the mentorship after I left the company and providing constant help with starting my future career.

I am very lucky to have many good friends especially those who were there to support me through my graduate studies. I need to thank Dr. Yangyang Liu, for understanding me and being there for me always. Although we are thousand miles away from each other, our friendship has been growing immensely during graduate school. I also would like to thank soon-to-be Dr. Yuanqi Tao, for always being there with me and sharing the same life experiences. We both applied to graduate school, studied mass spectrometry and are now graduating around the same time. We are always doing the same things in life, and are supporting each other along the way. I am very grateful to have two wonderful roommates, Dr. Yiwen Ouyang and soon-to-be Dr. Zhang Wang. We have shared so much during the past few years. Thank you

---

very much for always being there to help and support me. I will definitely miss the conversations we had during breakfast, lunch and dinner.

I need to acknowledge and thank my boyfriend Dr. Joe Tomlinson IV. Thank you for putting up with me the past few years and for all the support such as listening to my presentation practices again and again and again. Thank you for the happiness you brought me and being with me at UVa.

In the end, I must thank my parents, Hong Li and Dr. Shiguo Zhang, and all my other relatives as well. Mom and Dad, I can never express in words how grateful I am to be your daughter. You have provided me with more that I've ever needed and stood by me no matter what I was experiencing. Thank you for your constant endless love.

---

---

**Table of Content**

<b>1</b>	<b>Introduction .....</b>	<b>1</b>
1.1	Protein .....	1
1.2	Protein Post-translational Modifications .....	3
1.3	Protein Characterization Using Mass Spectrometry .....	4
1.4	References .....	18
<b>2</b>	<b>Identification of Phosphorylation Sites on Cytoplasmic Dynein from <i>Rattus norvegicus</i> .....</b>	<b>21</b>
2.1	Abstract .....	21
2.2	Introduction .....	23
2.2.1	Cell Signaling and Protein Phosphorylation .....	23
2.2.2	Cytoplasmic Dynein .....	26
2.2.3	Neurotrophic Signaling in Neuronal Survival and Trk Activation of Dynein IC Phosphorylation .....	28
2.3	Materials .....	29
2.3.1	Reagents .....	29
2.3.2	Equipment & Instrumentation .....	31
2.4	Methods .....	33
2.4.1	Cell Culture and Neurotrophic Stimulation .....	33
2.4.2	Affinity Purification .....	34
2.4.3	Immunoblot Analyses .....	35
2.4.4	Proteolytic Digestions and Sample Preparation for LC-MS/MS Analyses .....	35
2.4.5	HPLC Capillary Column Assembly .....	36
2.4.6	HPLC-MS/MS Analyses .....	37
2.4.7	Phosphorylation Analysis Using Immobilized Metal Affinity Chromatography for Phosphopeptide Enrichment .....	38
2.4.7.1	Cleanup of Samples to Remove Non-peptidic Impurities .....	38
2.4.7.2	Peptide Esterification with Methanol .....	39
2.4.7.3	Phosphopeptide Enrichment with Immobilized Metal Affinity Chromatography .....	40
2.4.8	Mass Spectrometric Data Analyses .....	41
2.5	Results & Discussion .....	42
2.5.1	Dynein Purification .....	42
2.5.2	<i>In silico</i> Proteolytic Digestion .....	46
2.5.3	LC-MS/MS Analyses of Dynein .....	52
2.5.3.1	Base Peak Chromatogram Analyses .....	52
2.5.3.2	Peptide Identification .....	55
2.5.3.3	Relative Peptide Quantitation .....	62
2.5.4	Targeted Phosphopeptide Analyses .....	68
2.5.5	Phosphorylation Enrichment Analyses .....	71
2.5.6	Identification of Neurotrophin-sensitive Dynein Phosphorylation .....	74

---

---

2.5.7 ERK1/2 and Phosphorylation of the Cytoplasmic Dynein Intermediate Chain .....	78
2.5.8 Intermediate Chain Phosphorylation Modulates Dynein Association with Trk-containing Organelles.....	81
2.5.9 Phosphorylation and NGF-dependent Neuronal Survival.....	84
<b>2.6 Conclusions .....</b>	<b>85</b>
<b>2.7 References.....</b>	<b>88</b>
<b>3 Identification of Post-translational Modifications of Human Desmoplakins in Regulating Interactions with Intermediate Filaments .....</b>	<b>92</b>
<b>3.1 Abstract .....</b>	<b>92</b>
<b>3.2 Introduction.....</b>	<b>93</b>
3.2.1 Desmosome and Desmoplakin .....	93
3.2.2 Phosphorylation of Desmoplakin.....	96
3.2.3 Desmoplakin Mutations in Diseases .....	99
3.2.4 Central Hypothesis .....	100
<b>3.3 Materials .....</b>	<b>101</b>
3.3.1 Reagents.....	101
3.3.2 Equipment & Instrumentation .....	102
<b>3.4 Methods.....</b>	<b>104</b>
3.4.1 Cell Culture and Affinity Purification.....	104
3.4.2 Immunoblot Analyses.....	105
3.4.3 In-solution Protein Cleavage and Sample Preparation for LC-MS/MS Analyses.....	105
3.4.4 In-gel Proteolytic Digestions and Sample Preparation for LC-MS/MS Analyses.....	106
3.4.5 HPLC Capillary Column Assembly .....	107
3.4.6 HPLC-MS/MS Analyses .....	108
3.4.7 Mass Spectrometric Data Analyses.....	109
<b>3.5 Results &amp; Discussion.....</b>	<b>110</b>
3.5.1 DP-S-tag Purification .....	110
3.5.2 Protein Cleavage and LC-MS/MS Analyses of DP-S-tag.....	112
3.5.3 PTM Analyses of DP-S-tag .....	121
3.5.3.1 Phosphorylation.....	121
3.5.3.2 Methylation .....	123
3.5.4 PTM Analyses of DP-S-tag S2849G .....	132
3.5.5 PTM Analyses of DP-S-tag R2834H .....	132
3.5.6 GSK3 Signaling Promotes Desmosome Assembly* .....	134
3.5.7 GSK3 Is Recruited to DP to Modulate DP-IF Complexes.....	135
3.5.8 GSK3 Promotes Phosphorylation Cascades of the DP C-terminal Tail.....	138
3.5.9 Sequential Phosphorylation Induces Arginine Claw Conformation of the DP C-terminal Tail.....	141
3.5.10 Crosstalk Between Arginine Methylation and Phosphorylation Contributes to the Regulatory Roles of the C-terminal Tail on DP behavior....	142

---

---

3.5.11 ARVC Point Mutation Alters Regulatory Roles of the DP C-terminal Tail .....	144
<b>3.6 Conclusions .....</b>	<b>147</b>
<b>3.7 References .....</b>	<b>149</b>
<b>4. Size-controlled Digestion for Middle-down nanoHPLC-MS/MS</b>	
<b>Characterization of the Sequence and Post-translational Modifications of</b>	
<b>Monoclonal Antibody .....</b>	<b>153</b>
<b>4.1 Abstract .....</b>	<b>153</b>
<b>4.2 Introduction .....</b>	<b>154</b>
4.2.1 Antibodies as Therapeutics .....	154
4.2.2 Structure Characterization of Therapeutic Antibodies .....	156
4.2.3 Structure Characterization Strategies of Protein Using Mass Spectrometry .....	157
4.2.4 Designing Rationales of the Size-controlled Digestion for Middle-down nanoHPLC-MS/MS Characterization of Monoclonal Antibody .....	161
<b>4.3 Materials .....</b>	<b>163</b>
4.3.1 Reagents .....	163
4.3.2 Equipment & Instrumentation .....	166
<b>4.4 Methods .....</b>	<b>168</b>
4.4.1 Covalent Linkage of Aspergillopepsin I to POROS 20 AL Beads .....	168
4.4.2 Fabrication of Enzyme-column and On-column Digestion .....	169
4.4.3 Protein Reduction, Alkylation and Size-controlled On-column Digestion .....	169
4.4.4 Protein In-solution Digestion with LysC and AspN .....	170
4.4.5 HPLC Capillary Column Assembly .....	171
4.4.6 HPLC-MS/MS Analyses .....	172
<b>4.5 Results &amp; Discussion .....</b>	<b>174</b>
4.5.1 Principles of Size-Controlled Digestion Method .....	174
4.5.2 Evaluation of Size-Controlled Digestion Method Using Apomyoglobin ..	176
4.5.3 Application of Size-Controlled Digestion for Monoclonal Antibody Sequencing and Post-translational Modifications Characterization .....	183
4.5.3.1 High Concentration of Urea Denatures Monoclonal Antibody .....	183
4.5.3.2 Optimization of Size-Controlled Digestion Time of Monoclonal Antibody .....	188
4.5.3.3 Monoclonal Antibody Sequencing and Post-translational Modifications Characterization .....	190
4.5.3.4 Charge Enhancement of Cys-containing Peptides to Improve ETD ..	194
4.5.3.5 Summary of Total Sequence Coverage and Identified PTMs of Waters Standard MAb .....	196
4.5.4 Repeatability of Size-controlled Digestion of Monoclonal Antibody .....	199
4.5.5 Comparison of Size-Control Digestion to Conventional In-solution Digest Using Other Proteases .....	200
<b>4.6 Conclusions and Future Directions .....</b>	<b>204</b>

---

Preface	Table of Content	viii
<b>4.7 References.....</b>		<b>208</b>

---

**Table of Tables**

<b>Table 2. 1</b> <i>In silico</i> tryptic digestion of dynein IC-1B. ....	48
<b>Table 2. 2</b> <i>In silico</i> AspN digestion of dynein IC 1B.....	51
<b>Table 2. 3</b> Phosphopeptides identified from dynein purified from rat brain samples. ....	67
<b>Table 2. 4</b> Phosphopeptides identified from dynein purified from PC12 samples. ....	67
<b>Table 2. 5</b> Phosphopeptides identified from dynein purified from cultured embryonic cortical neuron samples.....	68
<b>Table 2. 6</b> Phosphopeptides detected from an IMAC enrichment experiment on dynein from rat brain. ....	72
<b>Table 2. 7</b> Phosphopeptides detected from IMAC enrichment experiment on dynein from PC12 sample.....	72
<b>Table 2. 8</b> Phosphopeptides detected from IMAC enrichment experiment on dynein from cultured embryonic cortical neuron samples.....	74
 <b>Table 3. 1</b> Non-specific peptides detected by MS/MS at DP-S-tag C-terminus after AspN digestion. ....	117
<b>Table 3. 2</b> Selected non-specific peptides detected by LC-MS/MS after CNBr reaction.....	119
<b>Table 3. 3</b> Relative quantification of DP-S-tag peptides with CNBr and AspN cleavage.....	121
<b>Table 3. 4</b> A summary of methylated DP-S-tag peptide SSAPGSRSGSRSGSRSGSRSGSRSGSRSGSF. ....	131
 <b>Table 4. 1</b> ETD reaction times used for peptides with different charge states.....	180
<b>Table 4. 2</b> A summary of PTMs identified on Waters standard mAb.....	199

---

---

**Table of Figures**

<b>Figure 1. 1</b> Mechanism of CAD fragmentation. ....	11
<b>Figure 1. 2</b> A CAD spectrum of peptide ISGLIYEETR. ....	12
<b>Figure 1. 3</b> The proposed ETD radical-directed fragmentation mechanism. ....	15
<b>Figure 1. 4</b> A ETD spectrum of peptide HSDAVFTDNYTR. ....	16
 <b>Figure 2. 1</b> Phosphorylated peptide binding to Fe III imidodiacetate IMAC POROS® MC 20 resin. ....	25
<b>Figure 2. 2</b> Schematic drawing of dynein complex. ....	27
<b>Figure 2. 3</b> SDS-PAGE of immunopurified dynein from rat brain. ....	44
<b>Figure 2. 4</b> SDS-PAGE of immunopurified dynein from PC12 cells. ....	45
<b>Figure 2. 5</b> SDS-PAGE of immunopurified dynein from cultured embryonic cortical neurons. ....	46
<b>Figure 2. 6</b> Base peak chromatogram resulting from LC-MS/MS analysis of tryptic digests of dynein purified from rat brain sample. ....	54
<b>Figure 2. 7</b> Experimental isotopic distribution of phosphopeptide VGHDsELENQDK. ....	57
<b>Figure 2. 8</b> CAD spectrum of singly phosphorylated dynein IC-1 peptide VGHDsELENQDK. ....	60
<b>Figure 2. 9</b> ETD spectrum of singly phosphorylated dynein IC-1 peptide VGHDsELENQDK. ....	61
<b>Figure 2. 10</b> Sequence coverage map of dynein IC-1B from rat brain. ....	62
<b>Figure 2. 11</b> XICs of standard peptides angiotensin, vasoactive intestinal peptide, and ETQTPLATHQSEDEEDEEMVEPK. ....	64
<b>Figure 2. 12</b> XICs for dynein IC-1 peptide ETQTPLATHQSEDEEDEEMVEPK and its singly phosphorylated form. ....	66
<b>Figure 2. 13</b> Averaged CAD spectra of the phosphopeptide EAEALLQSMGLTTDSPIVPPPMsPSSK from a targeted precursor selection experiment. ....	70
<b>Figure 2. 14</b> Neurotrophin stimulation of cytoplasmic dynein IC phosphorylation. ....	77
<b>Figure 2. 15</b> IC phosphorylation requires active Trk and ERK1/2. ....	80
<b>Figure 2. 16</b> IC dephosphorylation reduces dynein binding to Trk-containing endosomes. ....	83
<b>Figure 2. 17</b> Expression of the dephospho-IC-1B S80A mutant in sympathetic neurons reduces NGF-dependent cell survival. ....	85
<b>Figure 2. 18</b> Model for ERK1/2 phosphorylation of ICs regulating cytoplasmic dynein binding to endosomes in response to neurotrophin. ....	87
 <b>Figure 3. 1</b> The desmosome and the desmoplakin. ....	94
<b>Figure 3. 2</b> Desmoplakin domain model. ....	96
<b>Figure 3. 3</b> Phospho-deficient DPS <sup>S2849G</sup> enhances DP-IF colocalization and delays DSM assembly in a Ca <sup>2+</sup> switch assay. ....	98
<b>Figure 3. 4</b> The DP structure and the proposed GSK3 phosphorylation scheme in the C-terminus. ....	100

---



<b>Figure 3. 5</b> DP-S-tag sequence. ....	110
<b>Figure 3. 6</b> SDS-PAGE and Western blot of affinity purified DP-S-tag.....	111
<b>Figure 3. 7</b> SDS-PAGE of affinity purified DP-S-tag mutates R2834H and S2849G. ....	112
<b>Figure 3. 8</b> DP-S-tag structure and protein cleavage on DP-S-tag C-terminal tail using AspN and CNBr.....	114
<b>Figure 3. 9</b> Scheme I for CNBr cleavage of Met-X bonds where X is any amino acid. ....	115
<b>Figure 3. 10</b> Scheme II for CNBr cleavage of Met-X bonds where X is a serine or threonine. ....	116
<b>Figure 3. 11</b> Base peak chromatogram resulting from LC-MS/MS analysis of CNBr and AspN cleavage of affinity purified DP-S-tag.....	120
<b>Figure 3. 12</b> Mono-phosphorylation and di-phosphorylation sites identified on DP-S-tag peptide SSAPGSRSGSRSGSRSGSRSGSRSGSRRGSF. ....	123
<b>Figure 3. 13</b> XICs of 1-6 methylation detected on DP-S-tag peptide SSAPGSRSGSRSGSRSGSRSGSRSGSRRGSF. ....	125
<b>Figure 3. 14</b> ETD spectrum of di-methylated PD-S-tag peptide SSAPGSRSGSRSGSRSGSRSGSRSGSRRGSF. ....	126
<b>Figure 3. 15</b> The mechanism of loss of small neutral molecules from symmetrically and asymmetrically methylated peptides in ETD. ....	127
<b>Figure 3. 16</b> Neutral loss analysis of ETD spectrum of penta-methylated PD-S-tag peptide SSAPGSRSGSRSGSRSGSRSGSRSGSRRGSF. ....	128
<b>Figure 3. 17</b> XICs of 1-6 methylation and phosphorylation detected on DP-S-tag peptide SSAPGSRSGSRSGSRSGSRSGSRSGSRRGSF. ....	130
<b>Figure 3. 18</b> All the phosphorylation and methylation sites identified on DP-S-tag peptide SSAPGSRSGSRSGSRSGSRSGSRSGSRRGSF. ....	130
<b>Figure 3. 19</b> GSK3 signaling promotes desmosome assembly. ....	135
<b>Figure 3. 20</b> GSK3 is recruited to DP to modulate DP-IF complexes ....	137
<b>Figure 3. 21</b> GSK3 promotes phosphorylation cascades of the DP C-terminal tail.....	140
<b>Figure 3. 22</b> Hypothesized phosphorylated-induced arginine claw conformation of the DP C-terminal tail confers the structure to this naturally unstructured region.....	142
<b>Figure 3. 23</b> Co-localization immunofluorescence microscopy of the association of DP and IF under normal condition and PRMT inhibition condition.....	144
<b>Figure 3. 24</b> ARVC point mutation alters regulatory roles of the DP C-terminal tail.....	146
 <b>Figure 4. 1</b> Schematic drawing of human IgG1.....	 156
<b>Figure 4. 2</b> A diagram of enzyme-column.....	174
<b>Figure 4. 3</b> Comparison of different times for size-controlled digestion of apomyoglobin. ....	178
<b>Figure 4. 4</b> Base peak chromatogram of apomyoglobin with 0.77 s of size-controlled digestion time.....	180
<b>Figure 4. 5</b> ETD MS2 spectrum of apomyoglobin peptide 114-153 and the sequence coverage.....	181
<b>Figure 4. 6</b> Apomyoglobin sequence coverage with ETD.....	182

<b>Figure 4. 7</b> Quantitation of apomyoglobin peptides based on normalized ion counts. ....	183
<b>Figure 4. 8</b> Comparison of the most abundant charge states of fully and partially alkylated mAb light chain and heavy chain. ....	185
<b>Figure 4. 9</b> Chromatograms of Waters standard mAb LC and HC resulting from various denaturing conditions for Cys reduction and alkylation. ....	187
<b>Figure 4. 10</b> Comparison of different side-controlled digestion times of Waters standard mAb. ....	189
<b>Figure 4. 11</b> TIC chromatogram of Waters standard mAb with 5.7 s of size-controlled digestion time. ....	190
<b>Figure 4. 12</b> Averaged MS1 spectrum of LC gradient segment 3 from Waters standard mAb with 5.7 s of size-controlled digestion. ....	192
<b>Figure 4. 13</b> ETD MS2 spectrum of Waters standard mAb HC 277-319 and Sequence Coverage. ....	193
<b>Figure 4. 14</b> The most common N-linked glycans on the Fc of mAbs. ....	194
<b>Figure 4. 15</b> Alkylation of cysteine side chain using iodoacetamide and aminoethylmaleimide. ....	196
<b>Figure 4. 16</b> Sequence coverage of Waters standard mAb peptide LC 1-52. ....	196
<b>Figure 4. 17</b> Waters standard mAb LC sequence coverage with ETD and CAD. ....	197
<b>Figure 4. 18</b> Waters standard mAb HC sequence coverage with ETD and CAD. ....	198
<b>Figure 4. 19</b> TIC and BP chromatograms of two parallel size-control digestion and HPLC-MS/MS experiments. ....	200
<b>Figure 4. 20</b> <i>In silico</i> digestion of Waters standard mAb with LysC and AspN. ....	201
<b>Figure 4. 21</b> TIC chromatogram of LysC in-solution digestion of reduced and alkylated Waters standard mAb. ....	203
<b>Figure 4. 22</b> TIC chromatogram of AspN in-solution digestion of reduced and alkylated Waters standard mAb. ....	204

---

---

**Abbreviations**

AAA	ATPases associated with diverse cellular activities
ACF-7	actin crosslinking factor-7
AGC	automated gain control
Ala, A	alanine
Arg, R	arginine
ARVC	Arrhythmogenic Right Ventricular Cardiomyopathy
Asn, N	asparagine
Asp, D	aspartic acid
ATP	adenosine triphosphate triphosphate
AV	number of spectra averaged
BDNF	brain derived neurotrophic factor
BSA	bovine serum albumin
CA	constitutively active
CAD	collision-activated dissociation
CDK	cyclin-dependent kinase
CDR	complementarity-determining region
cm	centimeter ( $10^{-2}$ meter)
cMyBP-C	cardiac myosin binding protein C
Cys, C	cysteine
Da	Dalton
DC	direct current
DM	dense midline
DMEM	dulbecco's modified eagle medium
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DP	desmoplakin
ECD	electron capture dissociation

---

---

EDTA	ethylenediaminetetraacetic acid
EM	electron multiplier
ERK	extracellular signal-regulated kinase
ESI	electrospray ionization
ETD	electron transfer dissociation
Fab	antigen binding fragment
FBS	fetal bovine serum
Fc	crystalizable fragment
FETD	front-end electron transfer dissociation
fmol	femtomole ( $10^{-15}$ mol)
FT	Fourier transform
GFP	green fluorescent protein
Gln, Q	glutamine
Glu, E	glutamic acid
Gly, G	glycine
GSK3	Glycogen synthase kinase 3
H/D	hydrogen/deuterium
HC	heavy chain
HEK	Human Embryonic Kidney
His, H	histidine
HPLC	high performance liquid chromatography
i.d.	inner diameter
IC	intermediate chain
ICR	ion cyclotron resonance
IF	intermediate filament
IgG	immunoglobulin gamma
Ile, I	isoleucine
IMAC	immobilized metal affinity chromatography

---

---

IP	immunoprecipitation
IT	ion trap
kDa	kilo Dalton ( $10^3$ Da)
LC	liquid chromatography, light chain
Leu, L	leucine
LIC	light-intermediate chain
LTQ	linear trap quadrupole
Lys, K	lysine
m	mass
M	molarity (mole/L)
m/z	mass-to-charge ratio
mAb	monoclonal antibody
MAPK	mitogen-activated protein kinase
me	methylation
Met, M	methionine
min	minute
mL	milliliter ( $10^{-3}$ L)
mM	millimolar ( $10^{-3}$ M)
mRNA	messenger RNA
MS	mass spectrometry
ms	millisecond
MS/MS	tandem mass spectrometry
MTB	microtubule-binding
MW	molecular weight
NAEM	N-(2-Aminoethyl)maleimide trifluoroacetate salt
NGF	nerve growth factor
NL	normalized ion count
nM	nanomolar ( $10^{-9}$ M)

---

---

ng	nanogram ( $10^{-9}$ g)
o.d.	outer diameter
OMSSA	open mass spectrometry search algorithm
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
Pg	plakoglobin
Phe, F	phenylalanine
PI3K	phosphatidylinositol 3-kinase
PKC	protein kinase C
PKC $\alpha$	protein kinase C $\alpha$ type
PLA	proximity ligation analysis
pmol	picomole ( $10^{-12}$ mol)
ppm	parts per million
Pro, P	proline
PTM	post-translational modification
Ras	rat sarcoma
rcf	relative centrifugal force
RF	radiofrequency
RI	relative intensity
RNA	ribonucleic acid
RNaseA	ribonuclease A
RP	reversed-phase
rpm	revolutions per minute
RT	retention time
s	second
<i>S. aureus</i>	<i>Staphylococcus aureus</i>
S/N	signal-to-noise ratio
SAM	S-adenosyl methionine

---

---

SCC9	squamous carcinoma cell line
SDS	sodium dodecyl sulfate
Ser, S	serine
TBS	tris buffered saline
TECP•HCl	tris(2-carboxyethyl)phosphine hydrochloride
TFA	trifluoroacetic acid
Thr, T	threonine
TIC	total ion count
Tris	tris(hydroxymethyl)aminomethane
Trk	tropomyosin-receptor-kinase
tRNA	transfer RNA
Trp, W	tryptophan
Tyr, Y	tyrosine
Val, V	valine
WT	wild type
XIC	extracted ion chromatogram
μg	microgram (10 <sup>-6</sup> g)
μL	microliter (10 <sup>-6</sup> L)
μm	micrometre (10 <sup>-6</sup> m)

---

---

**Abstract**

Proteins are large biological molecules consisting of one or more polypeptide chains with various amino acid sequences. The side chains of amino acids in a protein can be covalently modified by functional biochemical groups, a process known as post-translational modification (PTM). The structure and properties of a protein are essentially determined by the amino acid sequences and their PTMs. Mass spectrometry (MS) has emerged as the premier method for protein sequencing and PTM identification owing to its high sensitivity, accuracy, high analysis speed, and minimal bias. This dissertation describes the implementation of high performance liquid chromatography (HPLC) coupled with tandem MS using electron transfer dissociation (ETD) and collision-activated dissociation (CAD) as fragmentation methods to characterize sequence and PTMs of several proteins.

Chapter 1 is an overview of the application of mass spectrometry to protein characterization. This is followed by Chapter 2, which reports a comprehensive mapping of phosphorylation on cytoplasmic dynein from rat neurons that is mainly involved in retrograde axonal transport such as neurotrophin signaling. This study revealed 35 phosphorylation sites on different subunits, representing the largest number of phosphorylation sites reported on cytoplasmic dynein in any single study. Specially, the identification of two neurotrophin-sensitive dynein phosphorylation sites led to the following studies to identify a mechanism that is

---



---

used to recruit the cytoplasmic dynein to the signaling endosome upon neurotrophin stimulation for retrograde transport.

Chapter 3 reports the identification of phosphorylation and methylation on truncated desmoplakin (DP) C-terminal tail. DP is a cytoskeletal linker protein that tethers the intermediate filament (IF) network to desmosome in the heart and skin, and is required for maintaining mechanical integrity of these tissues. The subsequent studies demonstrated that the crosstalk between arginine methylation and serine phosphorylation cascades tunes DP-IF interactions and promote desmosome assembly and cell adhesion.

The last chapter reports “middle-down” MS characterization of a monoclonal antibody (mAb). This chapter involves the development of a novel “size-controlled” protein digestion enzyme reactor, which mainly generates 3-8 kDa large peptides from a protein. These large peptides yield high protein sequence coverage using ETD, facilitating the identification of sequences and PTMs of mAb.

---

---

## 1 Introduction

### 1.1 Protein

Proteins are large biological molecules consisting of one or more polypeptide chains. A polypeptide chain is a linear polymer chain of amino acid molecules bonded together. An amino acid consists of an amine group ( $-\text{NH}_2$ ), a carboxylic acid ( $-\text{COOH}$ ) group and a functional R group. There are twenty essential amino acids in most organisms differentiated by their R groups. Each essential amino acid is coded by the genetic code of deoxyribonucleic acids (DNA) (1).

To produce a polypeptide chain, genes in DNA are first transcribed into pre-messenger ribonucleic acid (mRNA) that is then processed to form the mature mRNA as the template for protein synthesis. This entire process of DNA to mRNA is known as transcription. The mRNA template is then read to synthesize a protein in a process called translation. During translation the ribosome reads along the mRNA in three nucleotide increments known as codons, matching each three nucleotide codon with its base pairing anticodon found on a transfer RNA (tRNA) molecule. Amino acids are then brought in by tRNA and bonded together to form a polypeptide chain and ultimately form a protein. The structure and property of a protein is essentially determined by all the various amino acids connected together (1).

---

Proteins function in a vast array of roles within living organisms. Some proteins have enzymatic activities and catalyze various biochemical reactions in a cell, and are essential for metabolism and survival. Other proteins are involved in the process of regulating the various expressions of genes. Proteins are also involved in cell structure, inter/intra-cellular transport or cell adhesion. An example of intra-cellular transporter is the cytoplasmic dyneins, which act as motor proteins along microtubules for retrograde transport of membrane bound organisms, and will be further discussed in Chapter 2. An example of a cell adhesion protein is the desmoplakin, which is a linker protein involved in cell adhesion, and will be discussed in Chapter 3. Moreover, proteins are also important in cell signaling and immune responses. For example, antibodies are protein components of the adaptive immune system whose main function is to bind antigens or foreign substances in the body, and target them for destruction. Characterization of antibodies will be discussed in Chapter 4.

Protein sequences determine the fundamental biological structure and function of proteins; sequencing proteins has become essential in biological science research. Although the sequences of most proteins are available in several public protein databases, advanced techniques that provide high protein sequence coverage are still in high demand. Protein sequencing is important to confident protein identification, as well as differentiation of protein variants, truncated forms and isoforms. In addition, the amino acid sequences of some proteins, e.g.

---

antibodies, are not directly inscribed in genome due to V(D)J recombination, *de novo* protein sequencing is needed to obtain the primary structure of these proteins.

## 1.2 Protein Post-translational Modifications

The entire collection of proteins produced by organisms is termed the proteome, named by an analogy to the genome which is the entirety of an organism's hereditary information (2). While the proteome is encoded by the genome, proteomes are 10-100-fold more complex than genomes (2). The protein complexity is achieved mainly by two mechanisms: splicing and PTMs. First, the alternative splicing of pre-mRNA molecules generates mature mRNAs with different subsets of the domains of the original RNA (3, 4). Secondly, shortly after or even during synthesis, the side chains of amino acids in a protein are covalently modified by functional biochemical groups, a process known as post-translational modification (PTM) (2). PTMs are one of the major subjects in this dissertation and will be further elaborated on in Chapter 2-4 when investigating the specific proteins of interest.

Protein PTMs are mediated by enzymatic activity and it is estimated that 5% of the proteome encodes enzymes that perform more than 200 types of PTMs (2). A few common examples are phosphorylation, methylation, acetylation, glycosylation, ubiquitination and methionine oxidation. Out of all the well known protein PTMs, phosphorylation is the most ubiquitous and well studied. PTMs alter the physical

---

and chemical properties of proteins; therefore affecting the folding, stability, activity, and ultimately, the function of the proteins. For example, in a phosphorylated protein, a bulky, tetrahedral phosphate is added to Ser/Thr/Tyr side chain, introducing net negative charges and therefore significantly affecting the protein structure and its interactions with other molecules. PTMs also control the catalytic activity of some proteins. For example, the autophosphorylation of protein kinases turns the active site geometry to “on” state. In addition, another consequence of PTMs on a protein, such as monoubiquitylation, labels the protein for recycling in lysosome and changes the intracellular location of the protein (1).

### **1.3 Protein Characterization Using Mass Spectrometry**

Protein sequences have historically been studied by using Edman degradation (5). Although it has been widely used for protein sequencing, Edman degradation has several disadvantages. Sample preparation for Edman degradation is exhaustive and consumes large amount of chemical reagents. Although the sequencing step is highly automated, this method is a low-throughput technology that usually can only sequence one amino acid residue per hour (6). Also, the efficiency of Edman degradation after 50 residues is very low, making it unsuitable for sequencing large protein/peptides (1).

Protein PTMs have been investigated using a variety of biochemical techniques such as radioactive labeling methods and antibody-based methods (7-9).

---

However, the localization of the PTMs is not available with radioactive labeling and it can be hazardous. Antibodies are used to probe PTMs by utilizing a western blot analysis. This technique allows for the detection of a modification at a specific site, but prior knowledge of the modification and location are required. Also, the lack of 100% specificity of the antibody can lead to ambiguity of site-specific identification (10). In addition, these techniques cannot comprehensively map all the PTMs that occur on a certain protein.

Mass spectrometry (MS) has emerged as the premier method for protein sequencing and PTM identification; because it is highly sensitive, accurate, fast, and offers minimal bias.

Mass spectrometer measures molecules' mass over charge ratio ( $m/z$ ) by transforming them into ions and then records the response of their trajectories to electric and magnetic fields or both (11, 12). Originally, it was a challenge to apply MS to large biomolecules, such as proteins, due to the difficulties of transforming those large molecules into gas-phase ions. Classical methods of ionization, such as electron ionization, photon ionization, and chemical ionization, cannot be applied to proteins since those methods are based on gas-phase encounters of the small, volatile molecules with another species to be ionized (12, 13). To overcome this limitation other ionization techniques are required for the MS analysis of proteins.

---

Electrospray ionization (ESI) was successfully applied to mass spectrometry in 1984 to ionize large molecules with no evidence of an upper limit for molecular weight (12, 13). In this technique, sample solution enters the ESI chamber through a stainless steel needle that is applied a few kilovolts relative to the surroundings. The resulting electric field at the needle tip charges the surface of the emerging liquid and disperses it into a fine spray of charged droplets by Coulomb forces. With the evaporation of the solvent, the diameter of a droplet decreases while the charge density on its surfaces increase, until the Rayleigh limit is reached. The Rayleigh limit is the point at which the Coulomb repulsion becomes the same order as the surface tension. The resulting instability tears the droplet apart, producing charged daughter droplets that evaporate. This sequence of events repeats until the daughter droplets are small enough that the field, due to the surface charge density, is strong enough to desorb ions from the droplet into ambient gas (14). Although the exact ionization mechanism is still debated, ESI has allowed for major breakthroughs in bioanalytical mass spectrometry by making the ionization of large, nonvolatile molecules such as proteins possible. Moreover, it allows the coupling of ESI with separation techniques like liquid chromatography (LC).

The coupling of ESI and LC has made it possible to fractionate a mixture of proteins or peptides, to reduce sample complexity, prior to MS analysis. Reversed-phase (RP) high performance liquid chromatography (HPLC) is commonly used in sample separation. RP-HPLC includes a polar mobile phase and nonpolar stationary

---

phase, usually consisting of silica-based C18 materials. Peptides are bound on the stationary phase through hydrophobic interactions with C18 chains in an aqueous polar mobile phase. The hydrophobicity of each peptide is determined by its amino acid composition. In general, peptides containing more hydrophobic amino acid side chains are more tightly bound to the C18 stationary phase, while peptides containing more hydrophilic amino acid side chains are more loosely bound. In a typical RP-HPLC gradient, the percentage of nonpolar component in the mobile phase is increasing linearly. As the percentage of the nonpolar component is increasing, loosely bound hydrophilic peptides will elute first, while tightly bound hydrophobic peptides will elute later in the gradient when the percentage of the nonpolar component is high. Thus, peptides are separated prior to MS analysis. This separation technique simplifies the complex samples for the following MS analysis, allowing for higher dynamic range and the possibility to identify thousands of peptides within a single experiment.

Tandem mass spectrometry (MS/MS), first used in the late 1960s, has been recognized as a powerful MS technique for identification of compounds in complex mixtures, structure elucidation of unknown compounds, as well as quantification of compounds in real samples (11). It has been widely applied to proteomics to obtain the sequences and PTMs of proteins. MS/MS refers to the coupling of two stages of mass analysis, MS1 and MS2. MS1 analysis separates peptides by  $m/z$  while MS2 analysis provides structural information for the peptides by generating fragment

---



ions. Many mass spectrometers have MS/MS functions realized in different forms, such as triple-quadrupole, linear ion trap (LIQ), time-of-flight (TOF)-based instruments, and hybrid instruments. Hybrid MS/MS instruments are constructed by coupling two different types of mass analyzers in order to take the advantages of each mass analyzer to enhance the final performance (11). One of the most popular hybrid MS/MS instruments is Linear Trap Quadrupole (LTQ) coupled with either Fourier Transform Ion Cyclotron Resonance instrument (FT-ICR) or Orbitrap (15-18).

FT-ICR and Orbitrap are both high-resolution mass analyzers with resolution over 100,000 and with mass accuracies less than 5 parts-per-million (ppm) (11, 19-21). In a FT-ICR mass spectrometer, an ICR cell is placed inside a strong magnetic field and ions are trapped laterally in the cell by a static magnetic field and axially by a static electric field. The ions are excited by a broadband radio frequency (rf) pulse to a coherent orbital motion with the frequency inversely proportional to the  $m/z$  values. The ions are detected by measuring the amplitude and frequency of the image current with the two detector plates. A time-domain free-ion decay signal is generated and Fourier-transformed to a frequency-domain signal which is then transformed to a mass spectrum (11, 19). The Orbitrap also generates image current of orbiting ions but it only uses electrostatic fields to confine and to analyze injected ions. An Orbitrap consists of an axial spindlelike central electrode and a coaxial barrel-like outer electrode. The trapped ions rotate around the central

---

electrode while harmonically oscillate in the axial direction. The frequency of axial oscillation is inversely proportional to the square root of  $m/z$ . The ions' image current that contains amplitude and axial oscillation frequency can be detected and converted into mass spectra by Fourier transform (20, 21). The accurate  $m/z$  measurement allows for the distinguishing of peptides with very close  $m/z$  and provides important information for highly accurate peptide identification. Therefore, in MS/MS analysis MS1 is usually performed in FT-ICR or Orbitrap as the first step for peptide sequencing and PTMs identification.

Ion trap mass analyzers are commonly used for MS2 experiments mainly due to its fast scan speed for unit resolution, high sensitivity (attamole level), and wide  $m/z$  range (up to 4000) (11, 22-25). These performance characteristics are due to the fact that ion trap mass spectrometers have the capability to transport, isolate, eject and analyze ions efficiently. An LTQ has the advantages over other ion trap instruments for its improved trapping efficiency, ion-storage capacity, ion-ejection efficiency, scan speed and detection sensitivity (11, 22). A typical LTQ consists of four parallel rods and each rod has three axial sections. The  $m/z$  separation of ions is accomplished by the stable vibratory motion of ions in a high-frequency oscillating electric field that is created by applying direct-current (dc) and rf potentials to these rods. In the mass-selective instability mode, with the increasing of the rf potentials, ions of successively increasing  $m/z$  become unstable and are ejected out of the trap. The ions exit through two parallel slots cut in a pair of rods

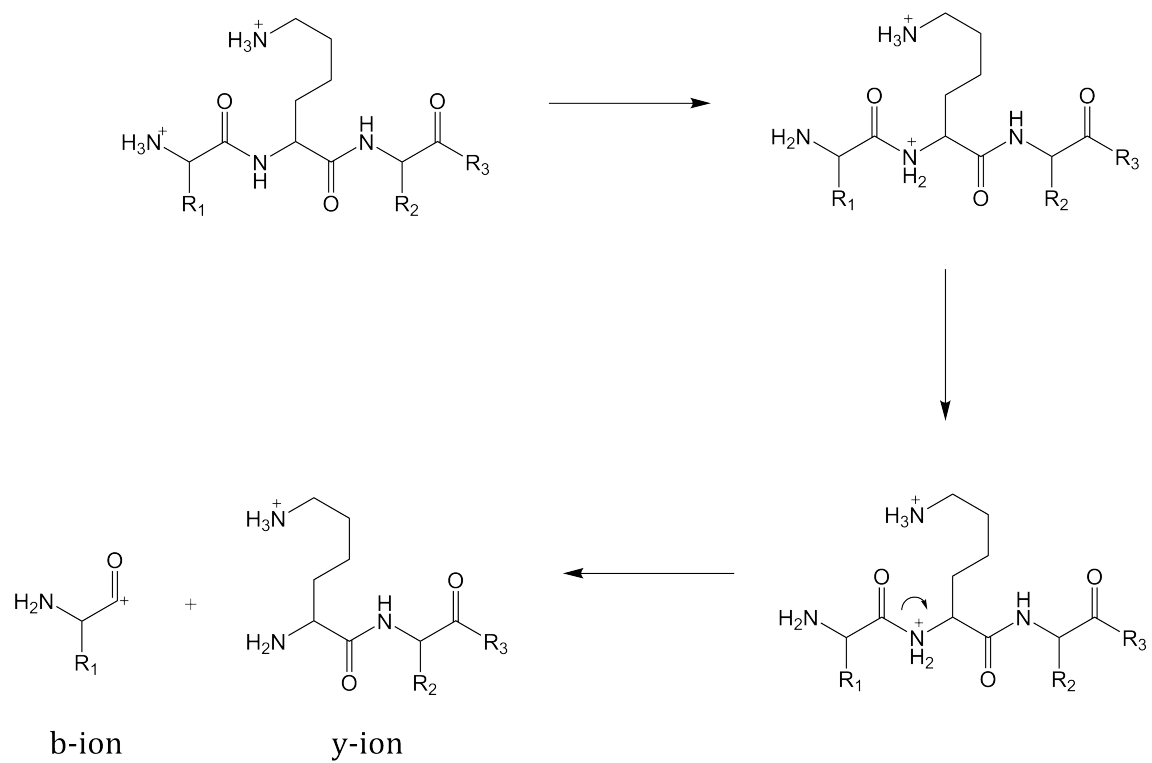
---

toward conversion dynodes where the signals are amplified and detected by electron multipliers. For ion isolation, ions with a narrow range of selected  $m/z$  remain stable in the trap while other ions are ejected (11).

To elucidate the primary structure of protein or peptide using mass spectrometry, fragmentation techniques are required in the MS2 experiments to dissociate the peptides to generate fragment ions. Two fragmentation techniques are widely applied to MS/MS in proteomics studies: collision-activated dissociation (CAD) and electron transfer dissociation (ETD) (26-31).

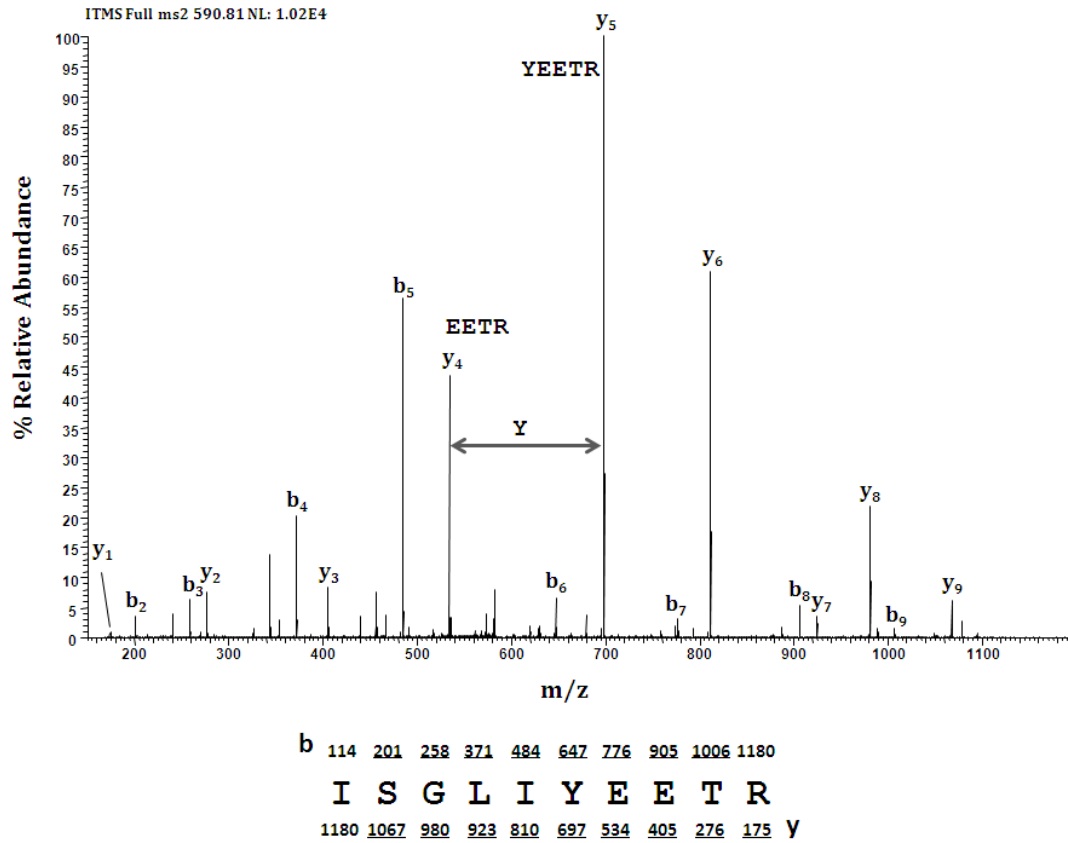
In CAD technique, the precursor ions obtain initial translational energy and are excited to higher-energy level by collisions with atoms of an inert gas such as helium, resulting in the conversion of part of the initial translational energy to the excitation energy to cause fragmentation (11, 29). Two fragment ions, b-ion and a y-ion are generated after one fragmentation of a peptide, as shown in **Figure 1.1**. These fragmentations occur in different amide bonds between two adjacent amino acid residues along the peptide backbone and generate series of b-ions and y-ions from which peptide sequences can be deduced, since the mass difference between consecutive b- or y-ions corresponds to the residue mass of an amino acid. **Figure 1.2** shows an example of a CAD spectrum and peptide sequencing using b- and y-ions.

---



**Figure 1. 1 Mechanism of CAD fragmentation.**

$\text{R}_1$  and  $\text{R}_2$  are amino acid side chains.  $\text{R}_3$  indicates the continued peptide backbone.



**Figure 1. 2** A CAD spectrum of peptide ISGLIYEETR.

The resulting b- and y-ions are labeled within the spectrum. An example of peptide sequencing using y-ions is shown here that the mass difference between  $y_5$  and  $y_4$  corresponds to the residue mass of Tyr, therefore the fifth amino acid from the C-terminus is Tyr. The sequence coverage of peptide ISGLIYEETR is shown under the spectrum where the identified fragment ions are underlined.

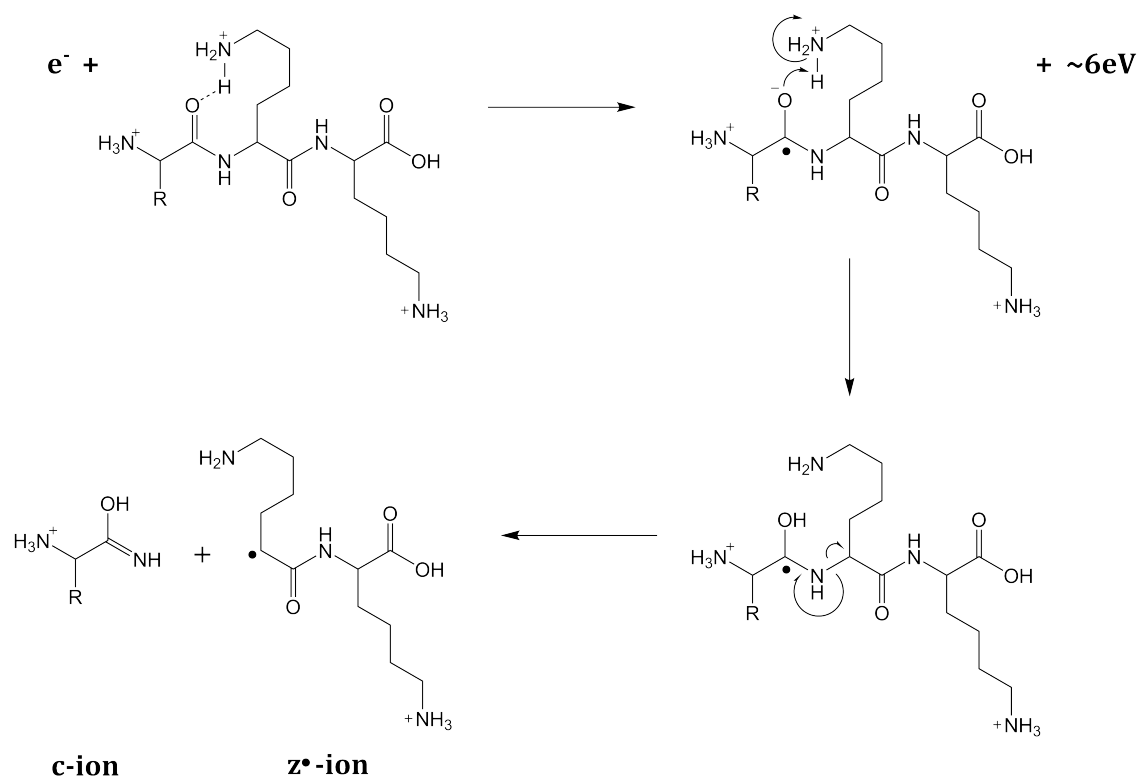
CAD favors the fragmentation of short (< 15 residues) tryptic peptides, which typically carry +2/+3 charge state. When peptides become bigger, the collision energy does not promote random fragmentation of different amide bonds along the long polypeptide backbone. Highly charged peptides cannot be completely fragmented either due to too many basic residues that block the proton transferring along the peptide backbone and therefore prevent the random fragmentation from occurring through different sites of the peptide (**Figure 1.1**) (30). Another disadvantage of CAD is that it promotes the loss of labile PTMs such as phosphate and carbohydrate. Under CAD conditions these sites are preferentially cleaved due to the lower energy requirement, resulting in a CAD spectrum displaying prominent ions corresponding to the intact peptides minus the PTM.

In 2004, Syka *et al.*, from the Hunt Group, developed another type of peptide fragmentation technique, ETD. This technique was based on the fragmentation mechanism of electron capture dissociation (ECD) (32, 33). ECD induces peptide fragmentation through the capture of near-thermal electrons in the Penning trap of an FT-ICR mass spectrometer. Capturing a thermal electron by a peptide is exothermic and leads to fragmentation along the backbone through a process independent of vibrational energy redistribution (32, 33). Therefore, PTMs are preserved upon peptide dissociation, and complete fragmentation of large peptides or intact proteins became possible. However, ECD was limited to FT-ICR instruments which are expensive to purchase and maintain. ETD has overcome this

---

issue by enabling the similar reaction in a less expensive LTQ. Rather than generating near-thermal electrons in the Penning trap, ETD utilizes radical anions (fluoranthene or azulene) generated in an external chemical ionization source to serve as electron donors to peptides (30, 31). ETD generates complementary c- and z<sup>•</sup>-ions, as shown in **Figure 1.3**, which can be used to sequence peptides and proteins similarly to CAD sequencing. **Figure 1.4** shows an example of ETD spectrum and peptide sequencing using c- and z<sup>•</sup>-ions.

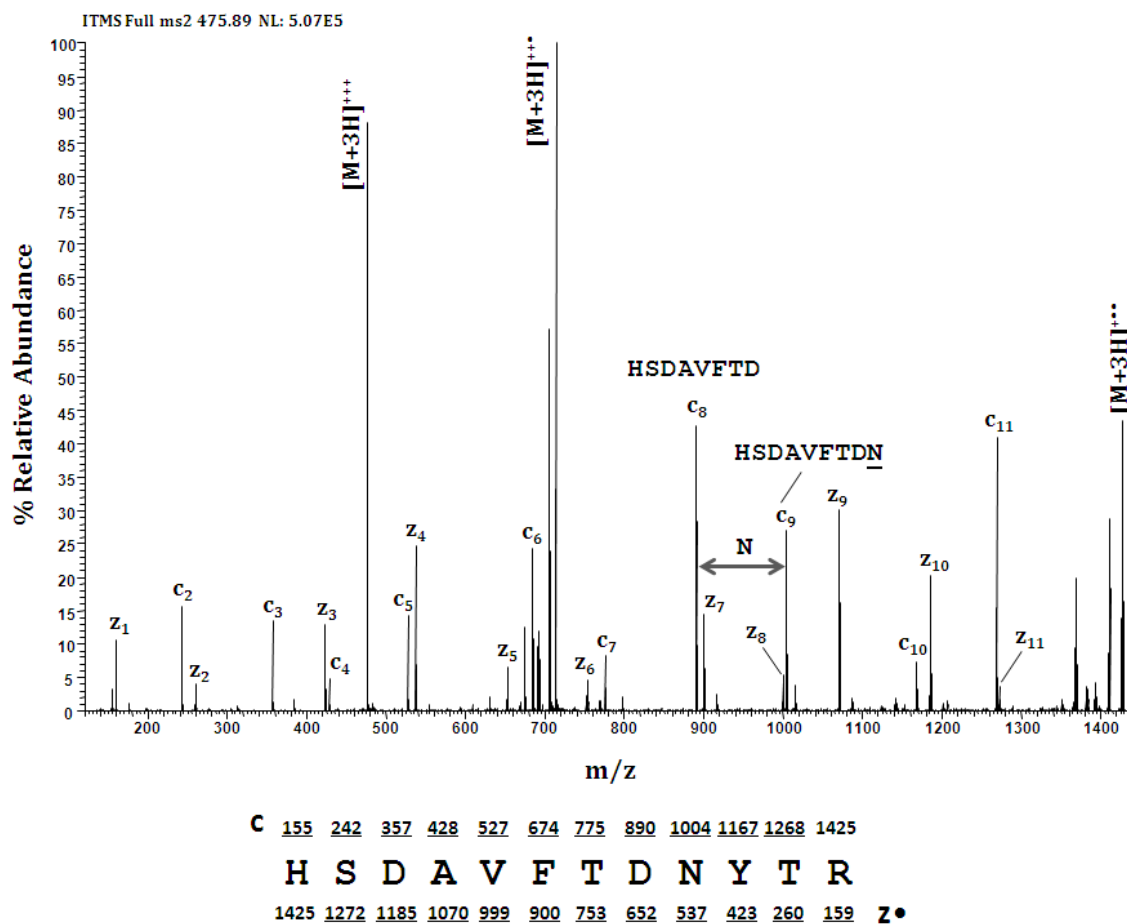
---



**Figure 1. 3 The proposed ETD radical-directed fragmentation mechanism.**

This figure is modified from (31) with permission.





**Figure 1. 4 A ETD spectrum of peptide HSDAVFTDNYTR.**

The resulting c- and z<sup>•</sup>-ions are labeled within the spectrum. Also labeled are the unreacted precursor ion and charge reduced species. An example of peptide sequencing using c-ions is shown. The mass difference between c<sub>9</sub> and c<sub>8</sub> corresponds to the residue mass of Asn, and therefore the ninth amino acid from the C-terminus is Asn. The sequence coverage of peptide HSDAVFTDNYTR is shown under the spectrum where the identified fragment ions are underlined.

Front end ETD (FETD) is a recent technique with improved robustness and reduced cost developed in the Hunt Group that allows ETD to be implemented on any instrument without modifications to footprint (34). The relocation of the reagent ion source to the front of mass spectrometer enables novel instrumental methods, such as multiple rounds of ETD to enhance signal-to-noise ratio (S/N).

This dissertation describes the implementation of LC-MS/MS with both CAD and FETD for the characterization of sequence and PTMs of proteins. Chapter 2 reports the identification of phosphorylation on rat cytoplasmic dynein 1. Chapter 3 reports the identification of phosphorylation and methylation on human desmoplakin C-terminus. In Chapter 4, a novel method is present on size-controlled digestion for middle-down characterization of the sequence and PTMs of monoclonal antibody.

---

---

**1.4 References**

1. Garrett, R. H. and Grisham, C. M. (2007) *Biochemistry*, 3rd Ed., Thomson, Belmont, CA
  2. Walsh, C. T. (2006) *Posttranslational Modification of Proteins: Expanding Nature's Inventory*, Roberts and Company Publishers, Greenwood Village, CO
  3. Black, D. L. (2003) *Mechanisms of alternative pre-messenger RNA splicing*. Annu. Rev. Biochem. 72, 291-336
  4. Maniatis, T. and Tasic, B. (2002) *Alternative pre-mRNA splicing and proteome expansion in metazoans*. Nature 418, 236-243
  5. Edman, P. (1949) *A method for the determination of amino acid sequence in peptides*. Archives of biochemistry 22, 475
  6. Henzel, W. J., Watanabe, C., and Stults, J. T. (2003) *Protein identification: The origins of peptide mass fingerprinting*. J. Am. Soc. Mass Spectrom. 14, 931-942
  7. Delom, F. and Chevet, E. (2006) *Phosphoprotein analysis: From proteins to proteomes*. Proteome Sci. 4, 15
  8. Rylatt, D. B. and Cohen, P. (1979) *Amino-acid sequence at the site on rabbit skeletal-muscle glycogen synthase phosphorylated by the endogenous glycogen synthase kinase-2 activity*. FEBS Lett. 98, 71-75
  9. Kaufmann, H., Bailey, J. E., and Fussenegger, M. (2001) *Use of antibodies for detection of phosphorylated proteins separated by two-dimensional gel electrophoresis*. Proteomics 1, 194-199
  10. Choudhary, C. and Mann, M. (2010) *Decoding signalling networks by mass spectrometry-based proteomics*. Nat. Rev. Mol. Cell Biol. 11, 427-439
  11. Dass, C. (2007) *Fundamentals of Contemporary Mass Spectrometry*, John Wiley & Sons, Hoboken, New Jersey
  12. Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F., and Whitehouse, C. (1989) *Electrospray ionization for mass-spectrometry of large biomolecules*. Science 246, 64-71
-

13. Whitehouse, C. M., Dreyer, R. N., Yamashita, M., and Fenn, J. B. (1985) *Electrospray interface for liquid chromatographs and mass spectrometers*. Anal. Chem. 57, 675-679
  14. Thmmson, B. A. and Iribarne, J. V. (1979) *Field-induced ion evaporation from liquid surfaces at atmospheric-pressure*. J. Chem. Phys. 71, 4451-4463
  15. Syka, J. E. P., Marto, J. A., Bai, D. L., Horning, S., Senko, M. W., Schwartz, J. C., Ueberheide, B., Garcia, B., Busby, S., Muratore, T., Shabanowitz, J., and Hunt, D. F. (2004) *Novel linear quadrupole ion trap/FT mass spectrometer: Performance characterization and use in the comparative analysis of histone H3 post-translational modifications*. J. Proteome Res. 3, 621-626
  16. Yates, J. R., Cociorva, D., Liao, L. J., and Zabrouskov, V. (2006) *Performance of a linear ion trap-orbitrap hybrid for peptide analysis*. Anal. Chem. 78, 493-500
  17. Makarov, A., Denisov, E., Kholomeev, A., Baischun, W., Lange, O., Strupat, K., and Horning, S. (2006) *Performance evaluation of a hybrid linear ion trap/orbitrap mass spectrometer*. Anal. Chem. 78, 2113-2120
  18. Makarov, A., Denisov, E., Lange, O., and Horning, S. (2006) *Dynamic range of mass accuracy in LTQ orbitrap hybrid mass spectrometer*. J. Am. Soc. Mass Spectrom. 17, 977-982
  19. Marshall, A. G. (1985) *Fourier-transform ion-cyclotron resonance mass-spectrometry*. Acc. Chem. Res. 18, 316-322
  20. Hu, Q. Z., Noll, R. J., Li, H. Y., Makarov, A., Hardman, M., and Cooks, R. G. (2005) *The orbitrap: A new mass spectrometer*. J. Mass Spectrom. 40, 430-443
  21. Perry, R. H., Cooks, R. G., and Noll, R. J. (2008) *Orbitrap mass spectrometry: Instrumentation, ion motion and applications*. Mass Spectrom. Rev. 27, 661-699
  22. Schwartz, J. C., Senko, M. W., and Syka, J. E. P. (2002) *A two-dimensional quadrupole ion trap mass spectrometer*. J. Am. Soc. Mass Spectrom. 13, 659-669
  23. Jonscher, K. R. and Yates, J. R. (1997) *The quadrupole ion trap mass spectrometer - A small solution to a big challenge*. Anal. Biochem. 244, 1-15
  24. Wong, P. S. H. and Cooks, R. G. (1997) *Ion trap mass spectrometry*. Current Separations 16, 85-92
-

25. March, R. E. (1997) *An introduction to quadrupole ion trap mass spectrometry*. J. Mass Spectrom. 32, 351-369
26. Jennings, K. R. (1968) *Collision-induced decompositions of aromatic molecular ions*. International Journal of Mass Spectrometry and Ion Physics 1, 227-235
27. Jennings, K. R. (2000) *The changing impact of the collision-induced decomposition of ions on mass spectrometry*. International Journal of Mass Spectrometry 200, 479-493
28. Shukla, A. K. and Futrell, J. H. (1993) *Collisional activation and dissociation of polyatomic ions*. Mass Spectrom. Rev. 12, 211-255
29. Hunt, D. F., Zhu, N., and Shabanowitz, J. (1989) *Oligopeptide sequence analysis by collision-activated dissociation of multiply charged ions*. Rapid Commun. Mass Spectrom. 3, 122-124
30. Syka, J. E. P., Coon, J. J., Schroeder, M. J., Shabanowitz, J., and Hunt, D. F. (2004) *Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry*. Proc. Natl. Acad. Sci. U. S. A. 101, 9528-9533
31. Udeshi, N. D., Shabanowitz, J., Hunt, D. F., and Rose, K. L. (2007) *Analysis of proteins and peptides on a chromatographic timescale by electron-transfer dissociation MS*. FEBS J. 274, 6269-6276
32. Zubarev, R. A., Kelleher, N. L., and McLafferty, F. W. (1998) *Electron capture dissociation of multiply charged protein cations. A nonergodic process*. J. Am. Chem. Soc. 120, 3265-3266
33. Zubarev, R. A., Horn, D. M., Fridriksson, E. K., Kelleher, N. L., Kruger, N. A., Lewis, M. A., Carpenter, B. K., and McLafferty, F. W. (2000) *Electron capture dissociation for structural characterization of multiply charged protein cations*. Anal. Chem. 72, 563-573
34. Earley, L., Anderson, L. C., Bai, D. L., Mullen, C., Syka, J. E. P., English, A. M., Dunyach, J., Stafford, G. C. J., Shabanowitz, J., Hunt, D. F., and Compton, P. D. (2013) *Front-end electron transfer dissociation: A new ionization source*. Anal. Chem. 85, 8385-8390
-

## 2 Identification of Phosphorylation Sites on Cytoplasmic Dynein from

### *Rattus norvegicus*

#### 2.1 Abstract

Dyneins are ATP-driven motor proteins that move membrane-bound organelles along microtubules. Cytoplasmic dyneins (1,500 kDa protein complex) from neurons are mainly involved in retrograde axonal transport such as neurotrophin signaling (1, 2). Several studies have supported the involvement of dynein in neurodegeneration-associated neuron diseases (3). In this project it is hypothesized that the recruitment of dynein to membranous organelles proceeds in part via phosphorylation of dynein intermediate chains (ICs, 74 kDa subunit of dynein) and dysregulated phosphorylation is responsible for the impaired function of dynein in neuron diseases.

Characterization of phosphorylation on dynein was completed using trypsin and AspN digestions of the entire 1,500 kDa protein complex followed by reversed-phase high performance liquid chromatography (HPLC) separation coupled to high resolution mass spectrometry (MS) analysis implementing both collision-activated dissociation (CAD) and electron transfer dissociation (ETD). Immobilized metal affinity chromatography (IMAC) was utilized to enrich for low- level phosphopeptides prior to MS analysis. In total, 35 phosphorylation sites were identified on cytoplasmic dynein from *Rattus norvegicus*.

---

---

Comparison analysis of phosphorylation on cultured embryonic cortical

neuron samples treated with and without neurotrophin revealed two neurotrophin-sensitive dynein IC phosphorylation sites. Pharmacological studies, overexpression of constitutively active MAP kinase and in vitro assay with recombinant proteins demonstrated that the ICs are phosphorylated by the MAP kinase, extracellular signal-regulated kinase (ERK) 1/2, a major downstream effector of tropomyosin-receptor-kinases (Trks). Trks are axonal neurotrophin receptor kinases residing in the plasma membrane. Neurotrophin binding results in Trk auto-phosphorylation and internalization as a signaling endosome with neurotrophin inside (4). Live cell imaging with fluorescently tagged IC mutants have shown that the dephosphomimic mutants had significantly reduced colocalization with Trk. The phosphorylated ICs were enriched on immunoaffinity-purified Trk-containing organelles and the inhibition of ERK reduced the amount of phosphorylated IC that co-purified with the signaling endosomes. Neurotrophin-dependent survival of sympathetic neurons was significantly reduced by the overexpression of the dephosphomimic mutant but not the wild type. These results demonstrate that neurotrophin binding to Trk initiates the phosphorylation of ICs for the recruitment of cytoplasmic dynein to signaling endosomes so that the neurotrophin can be retrograde transported in axons for axonal and neuronal survival (5).

---

---

## **2.2 Introduction**

### **2.2.1 Cell Signaling and Protein Phosphorylation**

Cell signaling is the process by which cells communicate with their environment and respond temporally to external messages received (6) . A large group of genes in all eukaryotes encode for cell membrane receptors that can bind to a variety of ligands. Then the signal can be transmitted to the cell through a series of intracellular interactions to induce internal conformational changes, chemical activation, and genetic expression (7). Cell development, tissue repair, immunity, and normal tissue homeostasis rely heavily on the ability of cells to perceive and correctly respond to the messages from their microenvironment (7). Errors in cellular information processing are responsible for a variety of diseases such as cancer, autoimmunity, and diabetes (8). Therefore, diseases may be prevented and treated effectively by understanding cell signaling pathways.

Proteins play critical roles in cell signaling. Cell signaling involves protein changes in three different ways: regulated protein PTMs; protein–protein interactions; and signal-induced protein expression changes (6, 9). Protein PTMs are closely involved in the other two levels of protein changes during cell signaling, and therefore, the study of protein PTMs is an important task in cell signaling research.

---



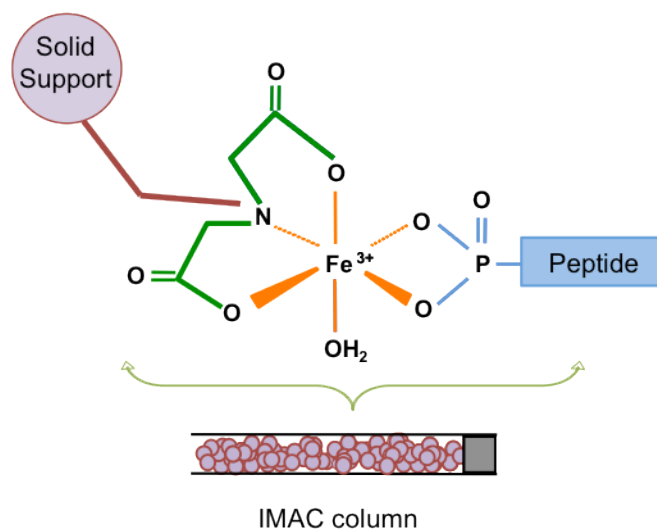
Protein PTMs are covalent additions of functional biochemical groups to amino acids. Common modifications include phosphorylation, glycosylation, ubiquitination, methylation, acetylation, and lipidation. PTMs are mediated by enzymatic activity and it is estimated that 5% of the proteome encodes enzymes that perform more than 200 types of PTMs (10). Of all known protein PTMs, phosphorylation is the most ubiquitous and well studied. The human genome encodes 1000-2000 protein kinases and 300-500 protein phosphatases, and the reversible phosphorylation of proteins regulates almost all aspects of cell life (11). It is estimated that about one-third of all cellular proteins are substrates for phosphorylation in their life cycle in eukaryotic cells (12).

Various analytical approaches have been used to detect phosphorylation. Historically, protein phosphorylation was detected by the transfer of radioactive  $^{32}\text{P}$  from  $\gamma$ - $^{32}\text{P}$  ATP to the protein substrate (13, 14). However, the localization of the phosphorylation is not available with this technique. Another approach is the recognition of the phosphorylated protein by specific antibodies (15). However, antibody recognition is not completely specific, the antibody signal is difficult to quantify, and antibodies are designed to a small number of known phosphorylation sites (9). These disadvantages make it difficult for the accurate identification and quantitation of novel phosphorylation sites. Mass spectrometry overcame the disadvantages of these approaches and became the most popular technique for

---

protein PTM detection with its high sensitivity and ability to detect the exact location.

A major obstacle in the study of phosphorylated proteins is that they are often at substoichiometric levels. Thus, many phosphoproteins cannot be detected. Several techniques have been developed to enrich for phosphopeptides from a complex mixture. IMAC is the most common technique for phosphopeptide enrichment. This method is based on the chelating affinity between  $\text{Zn}^{2+}$ ,  $\text{Fe}^{3+}$ ,  $\text{Ga}^{3+}$ ,  $\text{Al}^{3+}$  and phosphate groups (**Figure 2.1**) (16-18). Carboxylic acid-containing side chains (Asp, Glu) and the C-terminus have similar interactions with the metal ions. Converting the carboxylic acid groups to their methyl esters can prevent the non-specific binding and allow for the exclusive enrichment of phosphopeptides.



**Figure 2. 1 Phosphorylated peptide binding to Fe III imidodiacetate IMAC POROS® MC 20 resin.**

When peptide mixture is passing through the IMAC column, phosphopeptides bind on Fe III and non-phosphopeptides pass through. Phosphopeptides are eluted by reducing Fe III to Fe II using ascorbic acid.

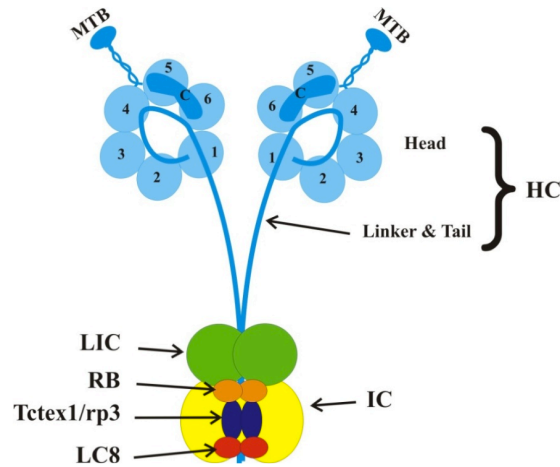
---

### 2.2.2 Cytoplasmic Dynein

Multicellular organisms need to move a variety of substances to specific cellular locations in order to maintain proper cell function. The most widely used mechanism for intracellular transport involves molecular motor proteins that carry cargo directionally along cytoskeletal tracks (19). One of the cytoskeletal tracks, microtubules, use kinesin and dynein as motor proteins. Kinesins move cargos to the polarized microtubule plus ends and dyneins move cargos to the minus ends (1, 2).

Cytoplasmic dynein (dynein) is a large protein complex composed of six subunits, the motor subunit and five smaller base subunits (**Figure 2.2**). The motor subunit is composed of a pair of heavy chains (HCs), with a molecular mass of approximately 530 kDa for each chain. Most of the base subunits have been implicated in cargo binding, and vertebrates have at least two genes for each of these subunits (2). Attached to the HC N-terminal region is a dimer of the intermediate chain (IC) and a dimer of the light-intermediate chain (LIC). Attached to the ICs are three pairs of light chains (LCs), including the Tctex1 family, the Roadblock family, and the LC8 family (20).

---



**Figure 2. 2 Schematic drawing of dynein complex.**

Different parts of dynein complex are represented by different shapes and colors, and labeled with the corresponding names. MTB: microtubules. C: HC C-terminal domain. The blue cycles for each HC are six AAA ((ATPases associated with diverse cellular activities) domains. RB, Tctex1/rp3 and LC8 are the three different families of LC. This figure is made by Dr. K. Kevin Pfister.

The HCs produce the moving force for the dynein complex. The major portion of the HC forms a head including seven domains: six AAA (ATPases associated with diverse cellular activities) domains and the HC C-terminal domain. A linker region connects the HC tail to the AAA domain that is most active as an ATPases (AAA domain 1 in **Figure 2.2**). Between the fourth and fifth AAA domains is a HC domain that forms a long anti-parallel coiled-coil stalk with an ATP hydrolysis-regulated microtubule-binding (MTB) site at its tip. Dynein's power stroke is driven by the releases of ADP and phosphate, which causes the head to rotate relative to the tail (21).

Dynein has many functions in cells. It is responsible for the transport of many different membranous organelles, chromosome movement, mitotic spindle

orientation, and nuclear migration (22). Several viruses including Adenovirus,

Herpes and HIV also utilize dynein to move along microtubules during the infection process (22, 23). Several studies have supported the involvement of dynein in neurodegeneration associated with Alzheimer's disease, Parkinson's disease, Huntington's disease and motor neuron diseases (3).

### **2.2.3 Neurotrophic Signaling in Neuronal Survival and Trk Activation of Dynein IC Phosphorylation**

The growth and development of neurons depends on the cell signaling initiated by target-derived growth factors. Neurotrophins, including nerve growth factor (NGF) and brain derived neurotrophic factor (BDNF), are the best characterized growth factors. Primary receptors for the neurotrophins are the Trk family (4, 24). Upon neurotrophin binding, the Trks are internalized in signaling endosomes. The retrograde transport of the endosomes to the cell body by dynein is essential for the signaling required for axonal and neuronal survival (5). However, the regulation of cytoplasmic dynein during retrograde axonal transport is less known. In previous studies, two of the subunits, IC and LIC have been implicated in dynein binding to membrane-bound organelles (2). Moreover, it was found that neuronal dynein was phosphorylated, and that after the addition of NGF to PC12 cells (a rat pheochromocytoma cell line), the level of dynein IC phosphorylation increased (25, 26).

---

Trks are upstream of several major kinase signaling pathways such as protein kinase C (PKC), rat sarcoma/mitogen-activated protein kinase (Ras/MAPK) and the phosphatidylinositol 3-kinase (PI3K) pathways (27). It is important to determine the Trk-dependent IC phosphorylation pathway and to investigate if IC phosphorylation regulates dynein binding to signaling endosomes. There are multiple IC isoforms in vertebrates, the products of two genes and alternative splicing. The IC-1B, IC-1C, IC-2B, and IC-2C isoforms are expressed in cultured embryonic cells, while only IC-2C and IC-2B are expressed in PC12 cells (28, 29). PC12 cells have been used as a model for growth factor induced differentiation and neurite extension. In the presence of NGF, these cells differentiate and extend axon-like processes (30). In this work, dyneins were purified from rat brain, cultured embryonic cortical neurons and PC12 cells, and were then characterized using MS. The goal was to comprehensively map all phosphorylation sites on dynein and to identify the neurotrophin-sensitive phosphorylation sites.

## **2.3 Materials**

### **2.3.1 Reagents**

#### **Alltech Associates, Inc. (Deerfield, IL)**

Acetyl chloride, anhydrous

D<sub>0</sub>-Methanol, anhydrous

---

---

**Mallinckrodt, Inc. (Paris, KY)**

Acetonitrile, HPLC grade,  $\geq 99.8\%$  purity

**Promega (Madison, WI)**

Trypsin, sequencing grade (porcine)

**PQ Corporation (Valley Forge, PA)**

Kasil<sup>®</sup> 1624

**Roche Diagnostics (Indianapolis, IN)**

Endoproteinase AspN

**Sigma Aldrich (St. Louis, MO)**

Ammonium bicarbonate

Angiotensin I acetate salt hydrate,  $\geq 90\%$  purity (human)

Ascorbic acid

Anulene

Dithiothreitol

Ethylenediaminetetraacetic acid, EDTA

Fluoranthene, 99% purity

---

Formamide

Glacial acetic acid, ≥99.9% purity, AcOH

Iodoacetamide

Iron (III) chloride, FeCl<sub>3</sub>

Trifluoroacetic acid, TFA

Vasoactive intestinal peptide fragment 1-12, ≥ 97% purity

**Thermo Scientific (Rockford, IL)**

Pierce® Water, LC-MS Grade

**Waters (Milford, MA)**

RapiGest SF Surfactant

**2.3.2 Equipment & Instrumentation**

**Agilent Technologies (Palo Alto, CA)**

HP1100 high performance liquid chromatograph

**Applied Biosystems™ (Carlsbad, CA)**

POROS® MC 20 metal chelating packing material, 20 µm diameter

**Branson (Danbury, CT)**

---



---

Branson 1200 Ultrasonic Bath

**Eppendorf (Hauppauge, NY)**

5415R Benchtop Centrifuge

**Labconco Corporation (Kansas City, MO)**

Centrivap centrifugal concentrator

**Polymicro Technologies, LLC (Pheonix, AZ)**

360  $\mu\text{m}$  o.d. x 50  $\mu\text{m}$  i.d. polyimide coated fused silica capillary

360  $\mu\text{m}$  o.d. x 75  $\mu\text{m}$  i.d. polyimide coated fused silica capillary

360  $\mu\text{m}$  o.d. x 150  $\mu\text{m}$  i.d. polyimide coated fused silica capillary

**Sutter Instrument Co. (Navato, CA)**

P-2000 microcapillary laser puller

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

LTQ-Orbitrap hybrid mass spectrometer

LTQ-FT-ICR hybrid mass spectrometer

**YMC Co., Ltd. (Kyoto, Japan)**

C18 resin (5-20  $\mu\text{m}$  diameter, 300 Å pore size)

---

---

C18 resin (5  $\mu\text{m}$  diameter, 120 Å pore size)

## **Zeus Industrial Products (Orangeburg, SC)**

Teflon tubing, 0.060" in o.d. x 0.012" in i.d.

## **2.4 Methods**

The sample preparation and cell biology-based experiments were completed by Kiev R. Blasier, former Laboratory and Research Specialist in Dr. K. Kevin Pfister's laboratory in the Department of Cell Biology at the University of Virginia, School of Medicine. The experimental methods for kinase inhibition and activation, live cell imaging, colocalization and survival assay are described in (31).

### **2.4.1 Cell Culture and Neurotrophic Stimulation**

Rat brain tissues were cut from adult rats.

PC12 cells were grown on 6 cm dishes coated with poly-L-lysine to confluency in dulbecco's modified eagle medium (DMEM) with 5% fetal bovine serum (FBS) and 10% calf serum. NGF was added to the medium at 4 nM final concentration, and the cells were incubated at 37°C. For samples analyzed by mass spectrometry NGF was added for 30min, and for samples analyzed by neurotrophic stimulation immunoblotting NGF was added for 5–60 min. The dishes were washed two times with phosphate buffered saline (PBS) before cell lysis.

---

Cultured embryonic cortical neurons were plated on poly-L-lysine-coated 6 cm dishes at  $2.5 \times 10^6$  cells/dish. For experiments with growth factor depletion, the dishes were washed three times in Neurobasal medium with Glutamax and incubated overnight. The dishes were then washed three times with Neurobasal medium only and incubated for an additional 2 hours in 3 ml/dish. For mass spectrometry analysis, BDNF at 100 ng/mL final concentration was added for 20 minutes or no BDNF was added. For neurotrophic stimulation immunoblotting analysis, BDNF was added for 5–60 minutes. The control dishes did not have medium replacement. The cells were washed the same as described for PC12 cells.

#### 2.4.2 Affinity Purification

Lysed rat brain, PC12 cells or cultured embryonic cortical neurons were immunoprecipitated on protein A-Sepharose-4B beads using 74.1 antibody. The antibody-linked beads were prepared by incubating 50-100  $\mu$ L of 10% slurry of protein A-Sepharose-4B and 10-20  $\mu$ g of 1mg/mL antibody in a microfuge tube on shaker at 4°C for 1 hour. The antibody-linked beads were centrifuged and washed with tris buffered saline (TBS) three times. Then the beads were pre-cleared with *Staphylococcus aureus* (*S. aureus*).

The cells were incubated on ice for 5 minutes in lysis buffer (50mM Tris, pH 8.1, 5mM EDTA, 150mM NaCl, 1.0% Triton X-100, 2 mM PMSF, 10 g/mL leupeptin and pepstatin, 10 mM benzamidine, 1 $\mu$ L/mL Phosphatase Inhibitor Cocktail II, 20

---

nM calyculin, and 200 nM staurosporine). Then, the samples were spun down at 4°C for 5 minutes at 10,000 rpm.

The cellular extract was incubated with antibody-linked Sepharose-4B beads on shaker at 4°C for 3 hours. Protease inhibitors were added before the incubation to avoid proteolysis. Antibody-linked Sepharose-4B beads with bound dynein were washed three times with varying salt concentrations to remove nonspecifically bound proteins prior to digestion. Phosphatase inhibitors were added to the sample to preserve phosphorylation during the sample storage process.

### **2.4.3 Immunoblot Analyses**

Dyneins immunoprecipitated from rat brain, PC12 cells, or cultured embryonic cortical neurons were resolved on a 7.5% Tris glycine gel after which proteins were transferred to a polyvinylidene fluoride membrane and immunoblotted against dynein IC or phosphorylated dynein IC. Immunoblots were then scanned and analyzed using Metamorph.

### **2.4.4 Proteolytic Digestions and Sample Preparation for LC-MS/MS Analyses**

Affinity purified dynein from different cells and conditions retained on Sepharose-4B beads was washed three times with 100 µL aliquots of 100 mM ammonium bicarbonate (pH 8.0). To improve the solubility of dynein, 0.1% RapiGest SF surfactant was added to 100 mM ammonium bicarbonate as the digestion buffer. For the dynein sample containing 5 pmols of protein, Cys residues

---

were reduced for 1 hour at room temperature using 4  $\mu$ L of 5 mM dithiothreitol in 100 mM ammonium bicarbonate and carbamidomethylated for 45 minutes in dark using 10  $\mu$ L of 5 mM iodoacetamide in 100 mM ammonium bicarbonate. The amount of dithiothreitol used was 20 times of dynein Cys residues and the amount of iodoacetamide used was 2.5 times of dithiothreitol. On-beads digestion was performed on a shaker at room temperature for 6 hours using trypsin or AspN. The ratio of enzyme-to-protein was kept at 1:20. Upon the completion of the digestion, the supernatant containing peptides was removed from the beads and the beads were washed three times with 100  $\mu$ L aliquots of 100 mM ammonium bicarbonate. Supernatants from each washing step were collected into the same tube with the first supernatant. TFA was added to the supernatant to adjust the pH below 2 to quench the digestion and cleave RapiGest SF. Then this sample was incubated at 37°C for 45 minutes and centrifuged at 13,000 rpm for 10 minutes so that any precipitates can be removed. The final sample was dried down in a centrivap concentrator and reconstituted in 0.1% acetic acid for storage at -35°C.

#### **2.4.5 HPLC Capillary Column Assembly**

An analytical column was constructed using 360  $\mu$ m o.d. x 50  $\mu$ m i.d. fused silica capillary with a laser-pulled bottleneck frit packed with 2 mm of irregular C18 resin (5-20  $\mu$ m diameter, 300 Å pore size). The analytical column was packed with 7 cm of regular C18 resin (5  $\mu$ m diameter, 120 Å pore size) and equipped with a laser-pulled electrospray emitter tip. A precolumn was constructed using 360  $\mu$ m

---

o.d. x 75  $\mu\text{m}$  i.d. fused silica capillary with a 2 mm Kasil 1624 frit and packed with 5 cm of C18 resin (15  $\mu\text{m}$  diameter, 120 Å pore size). Both columns were conditioned with several picomoles of angiotensin for three times to optimize chromatographic performance. The chromatographic performance was examined by analyzing 100 fmols of angiotensin and vasoactive intestinal peptide 1-12 (HSDAVFTDNYTR) prior to use for sample separation.

#### 2.4.6 HPLC-MS/MS Analyses

A 10% fraction of the dynein digest (corresponding to 1 pmol of ICs) was pressure-loaded onto the precolumn on a pressure bomb at a flow rate of 1  $\mu\text{L}/\text{min}$ . Internal standard peptides (250 fmols of angiotensin and vasoactive intestinal peptide) were added to the sample prior to loading for relative quantitation. The sample was desalted by rinsing the precolumn with 0.1 M acetic acid on HPLC for 10 minutes at a flow rate of 1  $\mu\text{L}/\text{min}$ . The precolumn was dried on the pressure bomb and then connected to the analytical column using a 2 cm piece of Teflon tubing (0.060" in o.d. x 0.012" in i.d.). Peptides were gradient eluted at a flow rate of 60 nL/min from the HPLC column and electrosprayed into the mass spectrometer. The LC gradient was as follows: 0-60% of solvent B in 60 min, 60-100% solvent B in 10 min, 100% of solvent B in 5 min, 100- 0% of solvent B in 5 min, 100% of solvent A for 10 min (solvent A: 0.1 M acetic acid in water; solvent B: 0.1 M acetic acid in 70% acetonitrile and 30% water).

---

Mass spectrometric analyses included a high resolution MS1 scan in the Orbitrap (resolving power of 60,000 at  $m/z$  400) or FT-ICR cell (resolving power of 50,000 at  $m/z$  400) and 10 low resolution MS2 scans in the LTQ. The top five most abundant ions were selected for MS2 analysis. For each precursor, both CAD and ETD were performed to obtain complementary mass spectra for optimal peptide identification. Parameters for data dependence MS2 data acquisition included monoisotopic precursor selection enabled, charge state rejection enabled for  $z = +1$  and unassigned species, dynamic exclusion enabled: repeat count of 3, repeat duration of 30 s, exclusion list duration of 20 s, exclusion list size of 500, and precursor isolation window of 3  $m/z$ . ETD reaction time was 30-45 ms. FTMS automated gain control (AGC) target was  $5e5$  on LTQ-Orbitrap,  $1e6$  on LTQ-FT-ICR; ITMSn AGC target was  $1e4$ ; and ETD reagent target was  $2e5$  with azulene as the electron transfer reagent.

#### **2.4.7 Phosphorylation Analysis Using Immobilized Metal Affinity Chromatography for Phosphopeptide Enrichment**

##### **2.4.7.1 Cleanup of Samples to Remove Non-peptidic Impurities**

A cleanup column was constructed using 360  $\mu\text{m}$  o.d. x 150  $\mu\text{m}$  i.d. fused silica capillary with a 2 mm Kasil 1624 frit and packed with 10 cm of C18 resin (15  $\mu\text{m}$  diameter, 120 Å pore size). Peptides in 0.1% acetic acid were combined with 100 fmols of an internal standard (angiotensin II phosphate, DRVpYIHPF) and were

---

pressure loaded at a flow rate of 1  $\mu\text{L}/\text{min}$  onto this cleanup column. The sample was washed with 0.1M acetic acid at a flow rate of 1  $\mu\text{L}/\text{min}$  for 10 minutes on HPLC and then gradient eluted into an Eppendorf tube using a 105 minutes gradient as follows: 0-80% of solvent B in 40 min, 80-100% solvent B in 30 min, 100% of solvent B in 10 min, 100- 0% of solvent B in 5 min, 100% of solvent A in 20 min. The eluate was collected for the first 70 minutes of the gradient into the Eppendorf tube. Samples were dried down in a centrifugal concentrator and stored at  $-35^{\circ}\text{C}$ .

#### **2.4.7.2 Peptide Esterification with Methanol**

Dynein peptides were esterified on Asp, Glu residues and C-terminus using  $\text{D}_0$ -methanol, in order to prevent non-specific binding of peptides containing carboxylic acids to the Fe III ions in the IMAC column. Esterification reagents were prepared with 160  $\mu\text{L}$  anhydrous acetyl chloride added dropwise to 1 mL anhydrous  $\text{D}_0$ -methanol. The mixture was stirred for 5 minutes and then 40  $\mu\text{L}$  was added to the peptide sample. The sample tube was sonicated for 1 minute followed by 1 hour reaction at room temperature. Then the sample was dried down completely. The reaction was performed twice to ensure complete conversion of carboxylic acid groups to their corresponding methyl esters.

---



### 2.4.7.3 Phosphopeptide Enrichment with Immobilized Metal Affinity

#### Chromatography

An IMAC column was constructed using 360  $\mu\text{m}$  o.d. x 150  $\mu\text{m}$  i.d. fused silica capillary with a 2 mm Kasil® 1624 frit and packed with 6 cm of POROS® MC 20 imidodiacetate functional groups packing material. The column was rinsed with water at 20  $\mu\text{L}/\text{min}$  for 10 minutes, then rinsed with 50 mM EDTA (pH 9) at 20  $\mu\text{L}/\text{min}$  for 10 minutes to remove any chelating metal ions, and finally rinsed with water at 20  $\mu\text{L}/\text{min}$  for 10 minutes to remove EDTA from the IMAC column. The IMAC column was then activated using 100 mM  $\text{FeCl}_3$  at 20  $\mu\text{L}/\text{min}$  for 3 minutes and the flow was stopped for 1 minute to incubate the  $\text{Fe}^{3+}$  and imidodiacetate functional groups on the resin. This process was repeated three times. The IMAC column was then rinsed using 0.01% acetic acid at 1  $\mu\text{L}/\text{min}$  for 25 minutes to equilibrate the pH.

A standard phosphopeptide STsLVEGR (100 fmols) was added to the esterified sample in order to evaluate IMAC enrichment efficiency. The peptide sample was reconstituted in 50  $\mu\text{L}$  1:1:1 solution (methanol: acetonitrile: 0.01% acetic acid) and loaded onto the IMAC column at a flow rate of 1  $\mu\text{L}/\text{min}$ . Then, 25  $\mu\text{L}$  of the 1:1:1 solution was added into the empty sample tube and loaded onto the IMAC column again to reduce sample loss in the sample tube. The IMAC column was rinsed with 15  $\mu\text{L}$  of 0.01% acetic acid at 1  $\mu\text{L}/\text{min}$  to remove non-phosphopeptides. Phosphopeptides were eluted onto a precolumn connected with a Teflon sleeve

---

using 10  $\mu$ L of 250 mM ascorbic acid. The precolumn was separated from the IMAC column and washed with 0.1 M acetic acid for 10 minutes on the HPLC at a flow rate of 1  $\mu$ L/min. Angiotensin and vasoactive intestinal peptide (100 fmol) were loaded onto the precolumn for relative quantitation. The precolumn was then dried and connected with the analytical column. Peptides were gradient eluted and electrosprayed into the mass spectrometer as previously described. MS analysis parameters for the phosphopeptide enrichment sample were the same as previously described.

#### **2.4.8 Mass Spectrometric Data Analyses**

MS data were searched against a database containing the sequences of all subunits of rat dynein using the Open Mass Spectrometry Search Algorithm (OMSSA) software (32). MS data were searched using either CAD spectra containing b and y fragment ions or ETD spectra containing c and z<sup>•</sup> fragment ions. Database searches used trypsin or AspN enzyme specificity, missed cleavage count of 3 and E-value cutoff of 1. Variable modifications included in the database search were carbamidomethylation of Cys, oxidation of Met, phosphorylation of Ser, Thr and Tyr. For those data from IMAC experiments, the database search also included fixed modifications including methyl esters of Asp, Glu residues and C-terminal. A precursor mass tolerance of  $\pm 0.01$  Da was used for high-resolution MS1 data and a fragment ion mass tolerance of  $\pm 0.35$  Da was used for low resolution MS2 data. The removal of reduced charge species within ETD data prior to searches was selected.

---

All peptide sequences were manually validated by inspection of the accurate precursor mass and the corresponding CAD or ETD spectra. Some peptide sequences were also determined by *de novo* interpretation of the CAD and ETD spectra with the use of Mascot sequence queries (33). Search conditions were: trypsin or AspN cleavage, precursor mass tolerance of  $\pm 0.01$  Da, fragment ion mass tolerance of  $\pm 0.30$  Da, variable carbamidomethylation on Cys, phosphorylation on Ser/Thr/Tyr, and oxidation on Met. For IMAC experiments there were also fixed methyl ester modifications of the Asp/Glu/C-terminus.

## 2.5 Results & Discussion

### 2.5.1 Dynein Purification

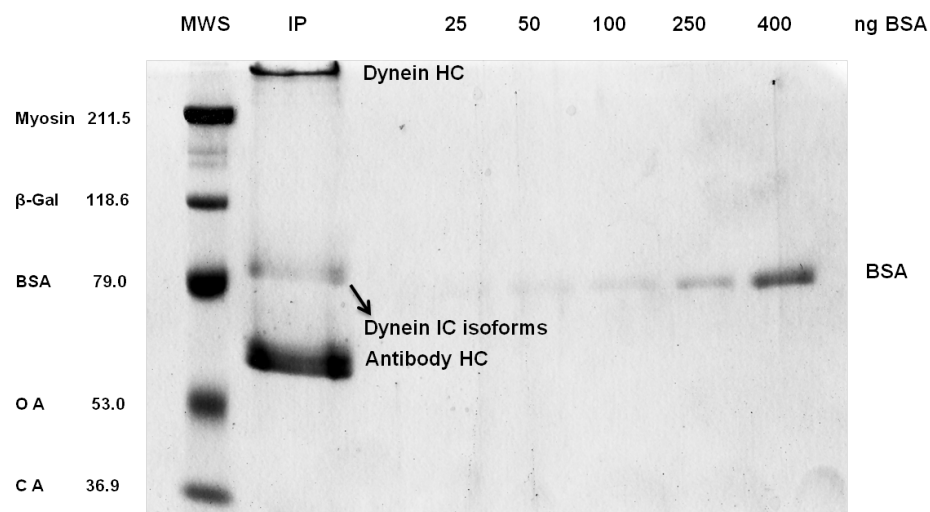
Immunoprecipitation of dynein utilized a 74.1 antibody ( $\alpha$ -pan-IC) that recognizes about 10 amino acid residues near the N-terminus of dynein IC. The 74.1 antibody was linked to Sepharose-4B beads through protein A. Protein A was covalently-linked to Sepharose-4B beads while maintaining the affinity to the 74.1 antibody. Since the Sepharose-4B beads, protein A and 74.1 antibody all have non-specific affinity to proteins, other proteins can co-precipitate as contaminants. After enzymatic digestion, the chromatography can be complicated by the peptides resulting from other proteins, and the MS signal of the dynein peptide can be suppressed if other peptides are present at higher levels. Therefore, to minimize non-specific protein binding it is important to purify dynein efficiently. Since the

---

74.1 antibody is linked to protein A-Sepharose-4B beads, it is critical to optimize the ratio of 74.1 antibody and protein A-Sepharose-4B beads. If the protein A-Sepharose-4B beads are not saturated with 74.1 antibody, the empty spots on protein A could bind to other proteins non-specifically. The pre-clearing of the 74.1 antibody linked protein A-Sepharose-4B beads is very necessary to saturate the Sepharose-4B beads with other proteins in order to minimize the empty binding spots on Sepharose-4B. Protein A is a 56 kDa surface protein originally found in the cell wall of the bacteria *S. aureus*. Therefore, *S. aureus* can be used as a protein source for the pre-clearing. Additionally, the choice of wash buffer can also affect the amount of loosely bound proteins that can be washed off the antibody-Sepharose beads. Usually the first wash buffer contains detergents to break hydrophobic interactions between proteins and the second wash buffer contains a high concentration of salt to counteract protein electrostatic interactions. Finally, the third buffer contains a low amount of salt to remove the large amount of salt from the second wash buffer. MS analysis on the first few samples showed that the samples were heavily contaminated with protein A and actins that are components of the cell cytoskeleton. After optimizing the immunoprecipitation protocol by adjusting the ratio of 74.1 antibody and protein A-Sepharose-4B beads, as well as testing different wash buffers, the amount of protein contaminants was reduced significantly. **Figures 2.3-2.5** show the SDS-PAGE results of immunopurified dyneins from different samples. The 74 kDa dynein ICs often appear to have several bands due to the slightly different molecular weights of the isoforms. Only dynein

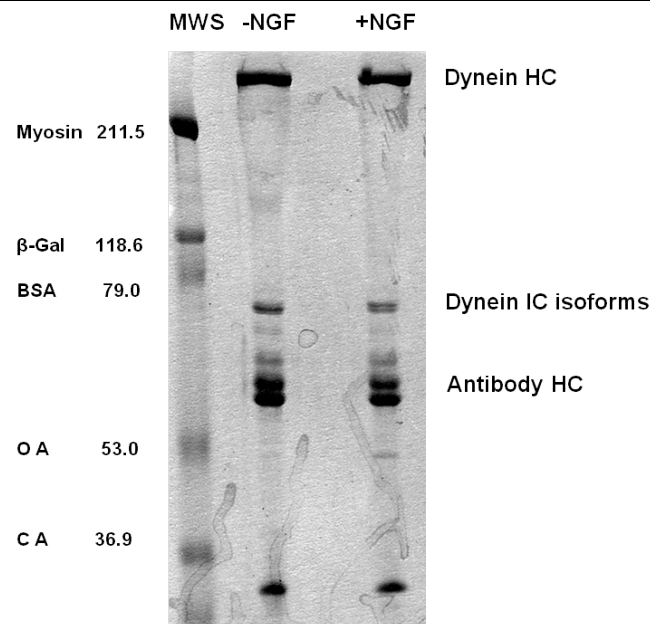
---

HC and IC appeared in the lane loaded with dynein because the dynein LICs (56 kDa) were covered by the antibody HC (about 50 kDa) and the dynein LCs (8 kDa) were beyond the molecular weight range. As shown in **Figure 2.3**, different amounts of bovine serum albumin (BSA) were loaded as a standard marker for relative quantitation. The relative amounts of dynein were estimated by comparing the darkness of the dynein gel bands to that of the BSA gel bands. **Figure 2.3** and **2.4** also show immunopurified dynein from PC12 cells or cultured embryonic cortical neurons when they were treated with or without neurotrophin.



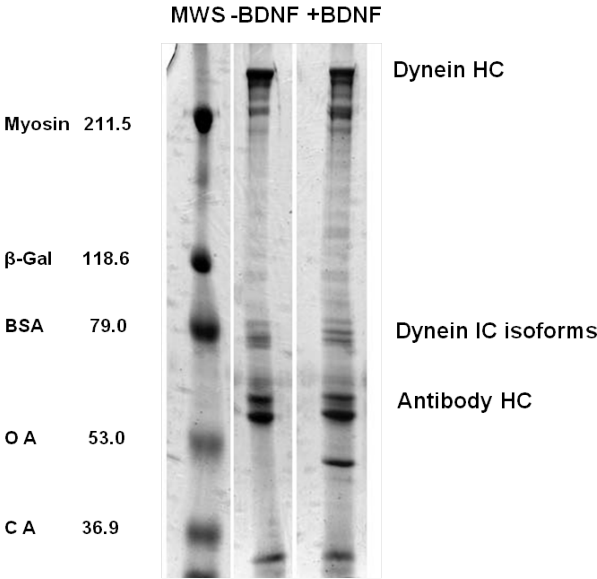
**Figure 2. 3 SDS-PAGE of immunopurified dynein from rat brain.**

The first lane was loaded with a BioRad prestained molecular weight standard (MWS) kit and the company provided the mass estimates (unit: kDa). B-Gal, beta galactosidase; BSA, bovine serum albumin; O A, ovalbumin; C A, carbonic anhydrase. The second lane was loaded with immunopurified dynein. Three major bands were labeled as dynein HC, dynein IC isoforms and antibody HC. The five lanes on the right were loaded with different amounts of BSA for relative quantitation (unit: kDa). The gel was stained with coomassie.



**Figure 2. 4 SDS-PAGE of immunopurified dynein from PC12 cells.**

The first lane was loaded with molecular weight standards (MWS, unit: kDa). The second lane was loaded with immunopurified dynein from PC12 cells with no treatment. The third lane was loaded with immunopurified dynein from PC12 cells treated with NGF. Three major bands were labeled as dynein HC, dynein IC isoforms and antibody HC. The gel was stained with coomassie.



**Figure 2. 5 SDS-PAGE of immunopurified dynein from cultured embryonic cortical neurons.**

The first lane was loaded with molecular weight standards (MWS, unit: kDa). The second lane was loaded with immunopurified dynein from cultured embryonic cortical neurons with depletion of BDNF. The third lane was loaded with immunopurified dynein from cultured embryonic cortical neurons with the addition of BDNF. Three major bands were labeled as dynein HC, dynein IC isoforms and antibody HC. The gel was stained with coomassie.

**2.5.2 *In silico* Proteolytic Digestion**

Proteolytic digestion of proteins for mass spectrometric analysis is necessary since the analysis of intact proteins using mass spectrometry is currently not fully developed. The size of peptides resulting from proteolytic digestion is very critical for the following LC-MS/MS analysis because small hydrophilic peptides may not be retained on the column by commonly used RP-HPLC, while large hydrophobic peptides are retained too tightly to the HPLC column and cannot be eluted. The molecular weight of dynein is uncommonly high at 150,000 kDa, compared to the average molecular weight of a protein is 50 kDa. Proteins with a higher molecular

weight have increased sequence complexity; therefore, in order to cover the entire protein sequence, usually more than one endoproteases are needed to generate peptides that can be detected and analyzed by mass spectrometer.

*In silico* digestions allow us to compare the theoretical peptides generated by different endoproteases in order to choose one or a combination of endoproteases that provide peptides with appropriate size and hydrophobicity to cover the entire protein sequence. *In silico* digestions were performed on the sequences of each dynein subunit using ExPASy PeptideMass digest simulator at <http://expasy.org/tools/peptide-mass.html>. Since dynein ICs were the primary focus of PTM analysis, comparison of dynein IC peptides generated by commonly used endoproteases resulted in the choice of trypsin and AspN.

As introduced in Chapter 1, trypsin specifically cleaves amide bonds C-terminal to Lys and Arg and has been highly preferable to other endoproteases for MS analysis due to its high efficiency and specificity. Also, Tryptic digestions are normally used for CAD analysis due to the production of short peptides with a charge of +2, where one charge is from the N-terminus and one charge is from Lys/Arg at the C-terminus. The endoprotease AspN cleaves amide bonds N-terminal to Asp residues, resulting in peptides with the Lys and Arg residues inside the sequences. Therefore, AspN digestion generates peptides with charge states greater than +2, rendering them suitable for ETD analyses. **Tables 2.1** and **2.2** show the *in silico* digestion results on dynein IC-1B isoform using Trypsin and AspN,

---



respectively. The combined theoretical sequence coverage by Trypsin and AspN

digestions was expected to yield detection of most of the phosphorylatable residues, Ser/Thr/Tyr. In practice, miscleavages often occur; therefore, the trypsin digestion would also generate peptides with Lys and Arg residues inside the sequences, resulting in peptides that have more than 2 charges. Moreover, AspN digestion could also generate peptides with  $z=+2$  if there are not enough residues that can carry charge present. Therefore, both CAD and ETD were utilized for the trypsin and AspN digested samples.

**Table 2. 1 *In silico* tryptic digestion of dynein IC-1B.**

Predicted unmodified masses, residue positions and number of phosphorylatable residues (STY) per peptide are listed. Peptides that were predicted to be detectable are denoted with a "+", those predicted to be not detectable with a "-" and questionable ones with a "?". The predication is based on the size of the peptides. Peptides with four or less amino acids were predicated to be not detectable. Peptides with more than five amino acids were predicated to be detectable. The detection of peptides with five amino acids was questionable.

Mass	Position	Peptide Sequence	#STY	Predicted Detection
479.205	1-4	MSDK	1	-
461.2485	5-8	SDLK	1	-
616.318	9-13	AELER	0	?
146.1055	14-14	K	0	-
146.1055	15-15	K	0	-
302.1702	16-17	QR	0	-
599.3755	18-22	LAQIR	0	?
404.1907	23-25	EEK	0	-
146.1055	26-26	K	0	-
174.1117	27-27	R	0	-
146.1055	28-28	K	0	-

561.2394	29-32	EEER	0	-
146.1055	33-33	K	0	-
146.1055	34-34	K	0	-
146.1055	35-35	K	0	-
848.3698	36-42	EADMQQK	0	+
146.1055	43-43	K	0	-
1256.552	44-54	EPVPDDSDLDR	1	+
146.1055	55-55	K	0	-
174.1117	56-56	R	0	-
174.1117	57-57	R	0	-
2804.4517	58-84	ETEALLQSIGISPEPPLVPTPMSPSSK	7	+
1915.9123	85-103	SVSTPSEAGSQDDLGLPTR	6	+
2387.1604	104-124	TLQWDTDPSVLQLQSDSELG R	5	+
174.1117	125-125	R	0	-
373.2325	126-128	LNK	0	-
502.3115	129-133	LGVSK	1	?
1073.5869	134-142	VTQVDFLPR	1	+
810.4123	143-149	EVVSYSK	3	+
2800.1868	150-173	ETQTPLATHQSEDEEDEEMVEPK	4	+
1369.611	174-185	VGHDSELENQDK	1	+
146.1055	186-186	K	0	-
504.2544	187-190	QETK	1	-
568.2969	191-195	EAPPR	0	?
876.4076	196-202	ELTEEEK	1	+
1920.9734	203-217	QQILHSEEFLLFFDR	1	+
388.2434	218-220	TIR	1	-
515.3067	221-224	VIER	0	-
1704.7631	225-239	ALAEDSDIFFDYSGR	3	+
646.3173	240-244	ELEEK	0	?
1462.68	245-258	DGDVQAGANLSFNR	1	+
1238.5356	259-267	QFYDEHWSK	2	+
311.1706	268-269	HR	0	-

4438.0054	270-308	VVTCMDWSLQYPELMVASYS NNEDAPHEPDGVALVWNMK	6	+
293.1739	309-310	FK	0	-
146.1055	311-311	K	0	-
2291.0173	312-331	TTPEYVFHCQSSVMSVCFAR	6	+
2371.2073	332-352	FHPNLVVGGTYSGQIVLWDNR	3	+
398.2026	353-355	SHR	1	-
174.1117	356-356	R	0	-
599.3391	357-361	TPVQR	1	?
3514.7626	362-395	TPLSAAAHTHPVYCVNVVGTQNAHNLITVSTDGK	8	+
2591.1416	396-417	MCSWSLDMLSTPQESMELVY NK	6	-
3402.6224	418-450	SKPVAVTGMAFPTGDVNNFV VGSEEGTVYTACR	7	+
427.2179	451-454	HGSK	1	?
4486.1514	455-496	AGIGEVFEGHQGPVTGINCH MAVGPIDFSHLFVTSSFDWTVK	6	+
647.3642	497-501	LWTTK	2	?
4142.8753	502-537	HNKPLYSFEDNADYVYDVMW SPVHPALFACVDGMGR	5	+
2855.4301	538-564	LDLWNLNSDTEVPTASVAIEGASALNR	5	+
273.1801	565-566	VR	0	+
645.3234	567-572	WAQGGK	0	+
1017.4727	573-582	EVAVGDSEGR	1	+
2312.1225	583-601	IWIYDVGELAVPHNDEWTR	2	+
392.2172	602-604	FAR	0	-
729.4385	605-610	TLVEIR	1	+
359.1917	611-613	ANR	0	-
1289.5623	614-626	ADSEEEGAVELAA	1	+

**Table 2. 2 *In silico* AspN digestion of dynein IC 1B.**

Predicted unmodified masses, residue positions and number of phosphorylatable residues (STY) per peptide are listed. Peptides that were predicted to be detectable are denoted with a "+", those predicted to be not detectable with a "-" and questionable ones with a "?". The predication is based on the size of the peptides. Peptides with four or less amino acids were predicated to be not detectable. Peptides with more than five amino acids were predicated to be detectable. The detection of peptides with five amino acids was questionable.

Mass	Position	Peptide Sequence	#STY	Predicted Detection
236.0831	1-2	MS	1	-
348.1645	3-5	DKS	1	-
4020.3031	6-37	DLKAELERKKQRLAQIREKKRKEEERKKKEA	0	+
1198.6016	38-47	DMQQKKEPVP	0	+
133.0375	48-48	D	0	-
220.0695	49-50	DS	1	-
246.1216	51-52	DL	0	-
4546.3336	53-95	DRKRRETEALLQSIGISPEPPLVPTPMSPSSKSVST PSEAGSQ	12	+
133.0375	96-96	D	0	-
1298.6983	97-107	DLGPLTRTLQW	2	+
234.0852	108-109	DT	1	-
985.508	110-118	DPSVLQLQS	2	+
2098.1858	119-137	DSELGRRLNKLGVSKVTQV	3	+
2890.3984	138-162	DFLPREVVSYSKETQTPLATHQSEE	7	+
391.1227	163-165	DEE	0	-
1268.5707	166-176	DEEMVEPKVGH	0	+
833.3403	177-183	DSELENQ	1	+
3915.9894	184-215	DKKQETKEAPPRELTEEEKQ QILHSEEFLIFF	3	-
1540.8685	216-228	DRTIRVIERALAE	1	+
220.0695	229-230	DS	1	-
540.2584	231-234	DIFF	0	-
1224.5622	235-244	DYSGRELEEK	2	+
190.059	245-246	DG	0	-
1728.822	247-261	DVQAGANLSFNRQFY	2	+
1626.7395	262-274	DEHWSKHRVVTCM	2	+
2144.9361	275-292	DWSLQYPELMVASYSNNE	5	+

664.2816	293-298	DAPHEP	0	+
5775.8655	299-349	DGVALVWNMKFKKTTPEYVFHCQSSVMSVCFAR FHPNLVV GGTYSQQIVLW	9	+
4717.412	350-392	DNRSHRRTVPVQRTPLSAAAHTHPVYCVNVVGTQ NAHNLIT VST	10	+
1025.431	393-401	DGKMCSWSL	2	+
3227.5552	402-431	DMLSTPQESMELVYNKSKPVAVTGMAFPTG	7	+
5039.3752	432-480	DVNNFVVGSEEGTVYTACRHGSKAGIGEVFEGHQ GPVTGI NCHMAVGPI	6	+
1285.5979	481-491	DFSHLFVTSSF	4	+
2392.2215	492-510	DWTVKLWTTKHNKPLYSFE	5	+
318.1175	511-513	DNA	0	-
558.2326	514-517	DYVY	2	-
1670.7949	518-532	DVMWSPVHPALFACV	1	+
647.3061	533-538	DGMGRL	0	+
860.4028	539-545	DLWNLNS	1	+
3237.6741	546-577	DTEVPTASVAIEGASALNRV RWAQGGKEVAVG	4	+
1137.5454	578-586	DSEGRIWIY	2	+
1049.5141	587-596	DVGELAVPHN	0	+
2203.161	597-614	DEWTRFARTLVEIRANRA	2	+
1218.5251	615-626	DSEEEGAVELAA	1	+

### 2.5.3 LC-MS/MS Analyses of Dynein

#### 2.5.3.1 Base Peak Chromatogram Analyses

MS analyses were conducted on dynein samples purified from rat brain, PC12 cells, and cultured embryonic cortical neurons. Tryptic or AspN digests were separated by a 90-minute HPLC gradient prior to entering the mass spectrometer. Peptides were eluted throughout the HPLC gradient due to the differences in hydrophobicity of the peptides. A base peak chromatogram was generated which

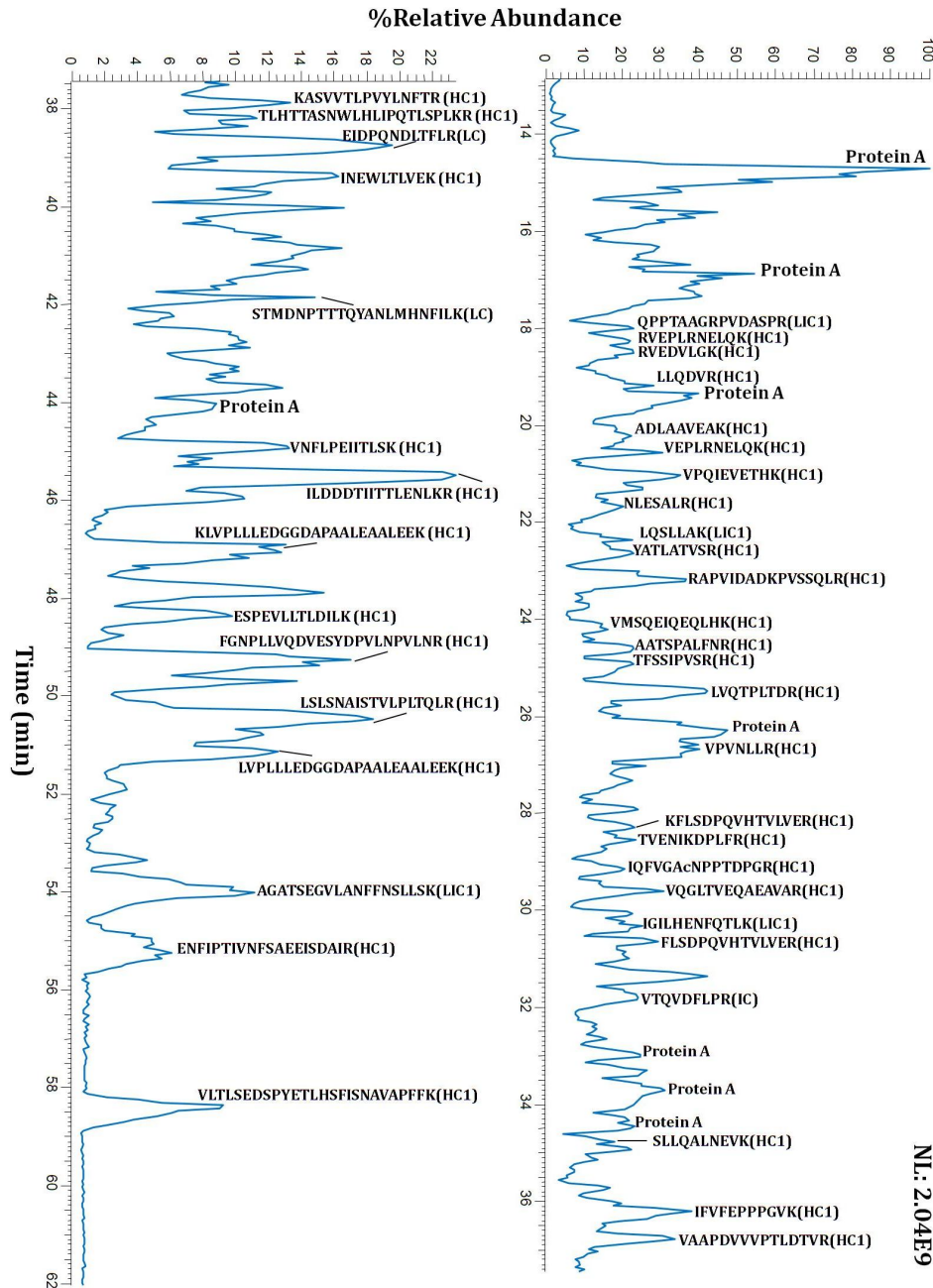
---

displays the elution profiles of the most abundant species through the gradient using MS1 data.

**Figure 2.6** shows the base peak chromatogram of a trypsin digestion of dynein from rat brain, with peaks labeled with the most abundant peptide under that peak. Most of the peaks are identified and labeled as peptides from dynein. There are a few peaks labeled as “protein A”, which are possibly from the unsaturated protein A on the Sepharose-4B beads. While unsaturated protein A can introduce protein A peptides as contamination to the digested sample, over digestion can also result in the presence of protein A peptides. However, given the amount of protein A detected from the digests compared to other dynein peptides, this contamination is not a serious issue. Overall, it demonstrates successful sample purification and high efficiency of proteolytic digestion and separation.

The use of RapiGest SF as a detergent in trypsin and AspN digestions was proved to enhance enzymatic digestion of dynein. RapiGest SF helps solubilize proteins, making them more susceptible to enzymatic cleavage without inhibiting enzyme activity. The use of RapiGest SF increased the dynein peptides generated from enzymatic digestion significantly. Since dynein HCs contain the most hydrophobic regions compared to other subunits, the amount of dynein HC peptides increased the most, up to 100 times, with the assistant of RapiGest SF. Also, RapiGest SF is acid-labile, rendering it to be easily degraded at low pH and removed from the sample; therefore, it is compatible with LC-MS/MS analysis.

---



**Figure 2. 6 Base peak chromatogram resulting from LC-MS/MS analysis of tryptic digests of dynein purified from rat brain sample.**

Most base peaks are identified and labeled as peptides from dynein (the subunits are also labeled for each peptide in parentheses). Lowercase "c" indicates a Cys residue that is carbamidomethylated. Due to the large size of the dynein complex and the complexity of the digests, it is inevitable that more than one peptide will have a similar hydrophobicity and elute around the same time and their chromatograms will overlap with each other.

### 2.5.3.2 Peptide Identification

MS/MS analyses of dynein peptides were performed using a LTQ-Orbitrap or a LTQ-FT-ICR hybrid mass spectrometer. The Orbitrap and FT-ICR mass analyzer have high resolution (exceeds 100,000), high mass accuracy, and high dynamic range; therefore are advantageous for acquiring MS1 data with monoisotopic  $m/z$  and charge state information (34, 35). Since the data acquisition time for an MS2 is longer in the high-resolution mass analyzer, MS2 fragments were analyzed in the LTQ in order to obtain MS2 spectra for as many precursors as possible in a chromatographic time scale.

The ability to measure MS1 scans in the Orbitrap or FT-ICR cell yields routine mass error less than 5 ppm (part per million). Mass error was calculated as described in **Equation 2.1**. Mass errors should be relatively consistent in each analysis. Usually the mass errors of standard peptides and sample peptides are expected to agree within a few ppm.

$$(2.1) \text{ ppm mass error} = \frac{\left(\text{experimental} \frac{m}{z}\right) - \left(\text{theoretical} \frac{m}{z}\right)}{\text{theoretical} \frac{m}{z}} \times 10^6$$

The high-resolution mass analyzers allows for the detection of distinct isotopic distributions, which can be used to calculate the charge state of each ion. Since each successive isotopic peak represents a species separated by  $\sim 1$  Da (the mass difference of  $^{12}\text{C}$  and  $^{13}\text{C}$ ) and the mass spectrometer measures  $m/z$ , the  $m/z$

---



difference between successive isotopic peaks indicates its charge state. The charge state can be solved by **Equation 2.2**:

$$(2.2) \ z = \frac{1}{\Delta m/z}$$

**Figure 2.7** shows the experimental MS1 data of the phosphopeptide VGHDsELENQDK (the lower case “s” indicates a phosphorylated Ser). On the left is the isotopic distribution of a species with monoisotopic  $m/z$  725.7978. The  $m/z$  difference between successive isotopic peaks is  $\sim 0.50$ . Therefore, the charge state of this species is:

$$z = \frac{1}{0.5} = 2$$

On the right is the isotopic distribution of a species with monoisotopic  $m/z$  484.2006. The  $m/z$  difference between successive isotopic peaks is  $\sim 0.33$ .

Therefore, the charge state of this species is:

$$z = \frac{1}{0.33} = 3$$

---

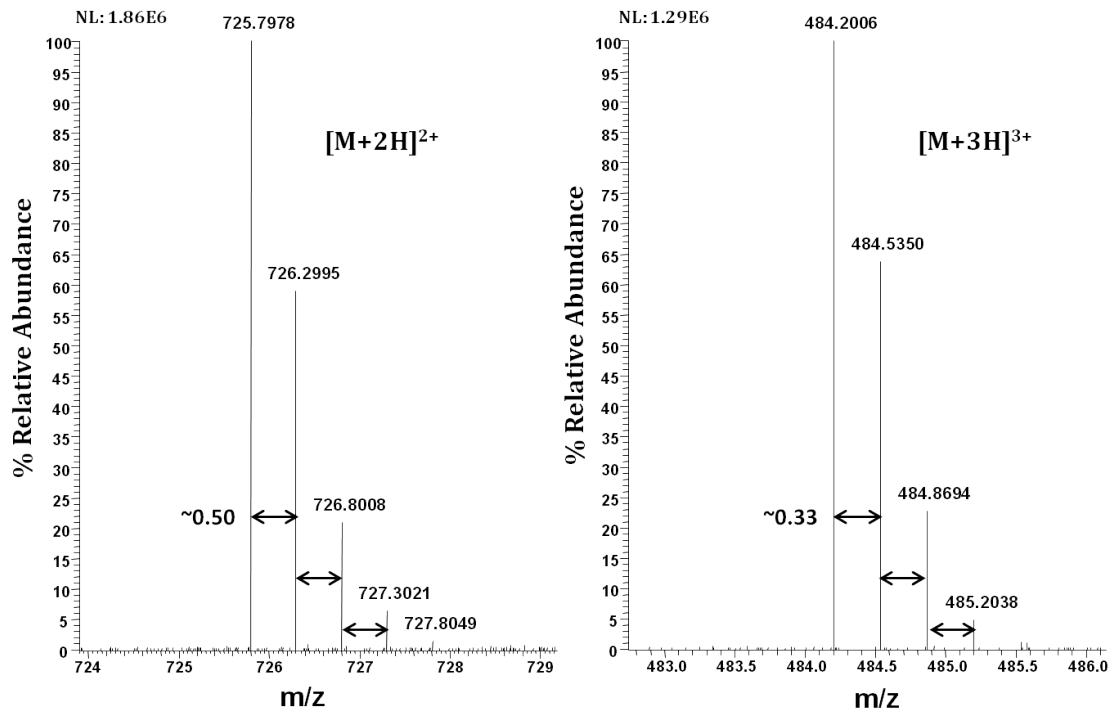


Figure 2. 7 Experimental isotopic distribution of phosphopeptide VGHDsELENQDK.

Data analysis for each dynein sample started with an OMSSA search of the MS/MS data against a dynein sequence database containing the sequences of all possible isoforms for all of the subunits. For each precursor, OMSSA determines its charge state and compares its mass with those calculated by digesting proteins *in silico* from the protein database with the enzyme of interest. If the masses match within a given tolerance, then the algorithm takes the experimental MS2 spectrum, filters noise peaks, extracts  $m/z$  values, and then compares these  $m/z$  values to calculated  $m/z$  values derived from those theoretical peptides. The resulting search hits are then statistically scored and filtered with a given score threshold (32). When searching for phosphopeptides, phosphorylation on Ser, Thr, and Tyr are selected as various modifications, then the masses of possibly modified theoretical

peptides are calculated and theoretical MS2 mass lists are generated for comparison.

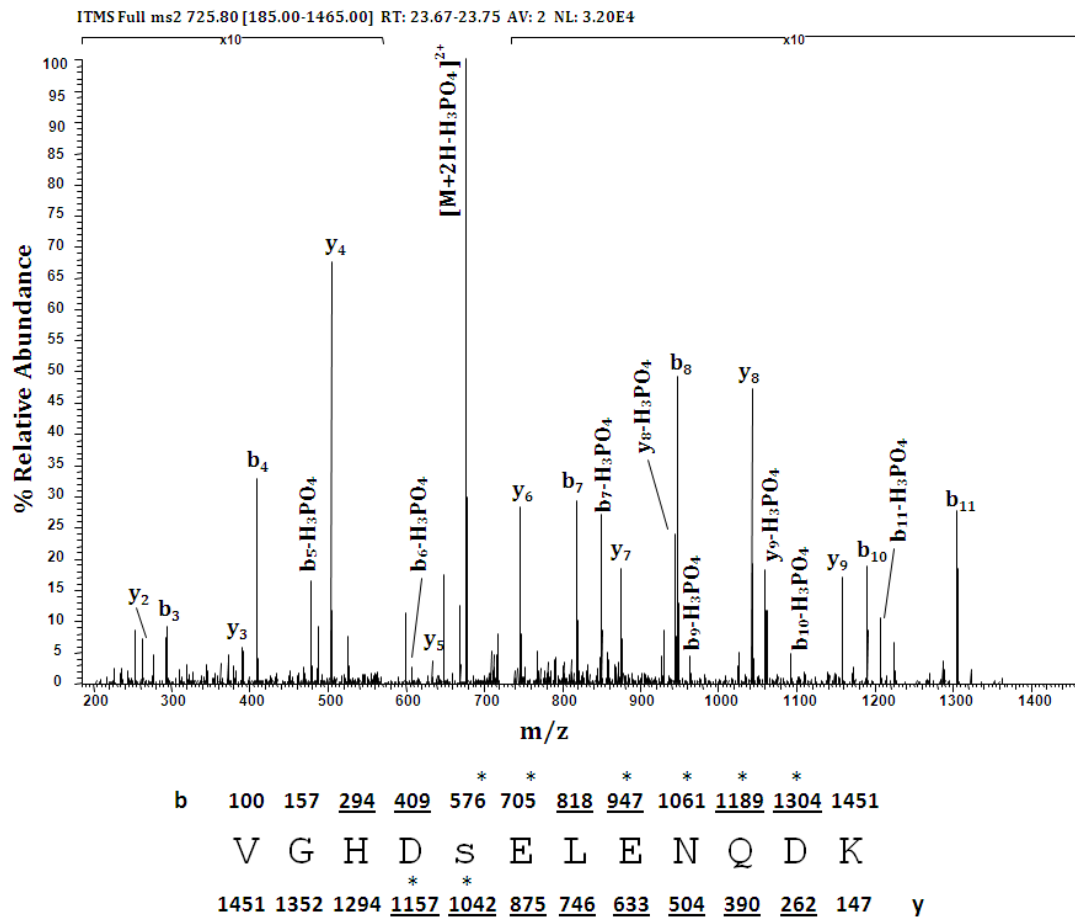
Next, all MS/MS spectra for unmodified and post-translationally modified peptides were manually validated to ensure confident identification. A convincing validation of a predicted peptide should have all the abundant ions assigned in the corresponding MS2 spectrum. The assignments include the predicted fragment ions b- and y from CAD, c- and z• from ETD, and those result from neutral losses. **Figure 2.8** displays the CAD spectrum of a singly phosphorylated dynein peptide VGHDsELENQDK. As introduced in Chapter 1, in CAD the loss of PTMs from the peptide backbone is preferable to peptide backbone fragmentation. Thus the CAD spectrum will be dominated by an abundant peak that results from the loss of phosphoric acid. As shown in **Figure 2.8**, the most abundant fragment ion resulted from the loss of phosphoric acid from the precursor ion, which indicates a low efficiency of CAD fragmentation of the peptide backbone. Also, the loss of phosphoric acid molecules from many fragment ions lowers the intensity of those fragment ions and complicates the CAD spectrum. In the given example, the position of the phosphorylation can be assigned to the only Ser in the sequence. However, not all the expected fragment ions were found. In other examples this effect can be more serious where only a few fragment ions were generated or the generated fragment ions were not strong enough for confident assignment. Especially when multiple Ser, Thr or Tyr residues are present in the same peptide,

---

complete coverage of fragment ions is required to confidently localize the phosphorylation site.

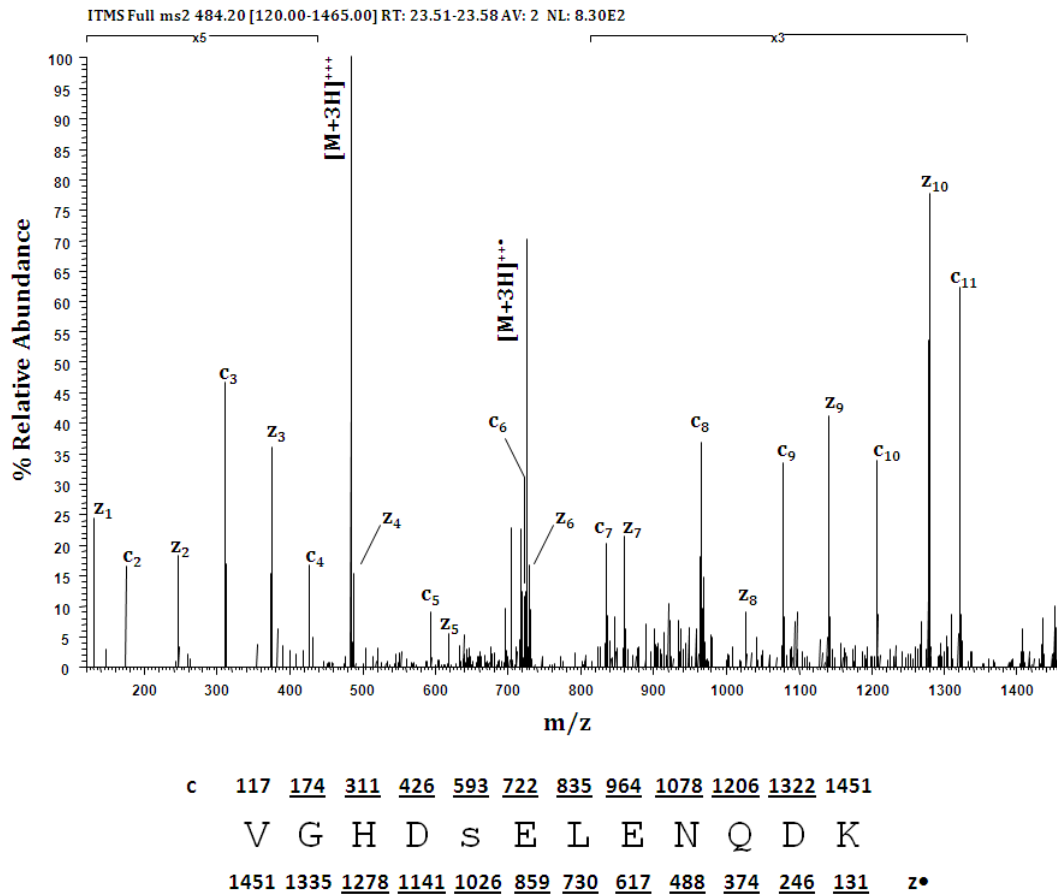
ETD fragmentation of the same phosphopeptide VGHDsELENQDK resulted in nearly complete fragment ion coverage and no neutral loss species were observed (**Figure 2.9**). As introduced in Chapter 1, during ETD fragmentation, the peptide backbone is fragmented through a nonergodic process in which intramolecular vibrational energy redistribution is not involved (36-38). Therefore labile PTMs can be preserved. The nearly complete fragment ion coverage allows for confident identification of Ser178 (IC-1B) as the phosphorylation site. In conclusion, CAD and ETD provide complementary fragmentation information to facilitate the identification of PTMs.

---



**Figure 2. 8 CAD spectrum of singly phosphorylated dynein IC-1 peptide VGHDsELENQDK.**

A precursor with the mass of 725.80 was chosen for CAD fragmentation. The experimental  $^{12}\text{C}$  isotope mass for +2 species was measured as  $m/z$  725.7974, differing by +2 ppm from the theoretical mass of 725.7959. Loss of phosphoric acid (-98 Da) from the precursor ion is labeled as  $[\text{M}+2\text{H}-\text{H}_3\text{PO}_4]^{2+}$ . Fragment ions detected are labeled within the spectrum and underlined in the peptide sequence. A lower case "s" indicates the phosphorylated Ser. Ions labeled with \* above represents the loss of phosphoric acid from those ions was observed. The phosphorylated Ser178 was confirmed by the unmodified form of  $y_7$  and the modified form of  $y_8$ . RT: retention time, AV: number of spectra averaged, NL: normalized ion counts.



**Figure 2. 9 ETD spectrum of singly phosphorylated dynein IC-1 peptide VGHDsELENQDK.**

A precursor with the mass of 484.20 was chosen for ETD fragmentation. The experimental  $^{12}\text{C}$  isotope mass for +3 species was measured as  $m/z$  484.2000, differing by less than +1 ppm from the theoretical mass of 484.1997. Fragment ions detected are labeled within the spectrum and underlined in the peptide sequence. A lower case "s" indicates the phosphorylated Ser. No loss of phosphoric acid was observed. Nearly all fragment ions were detected and demonstrate the phosphorylation at Ser178. Bracketed ions correspond to the unreacted precursor ion and the charge-reduced ion. RT: retention time, AV: number of spectra averaged, NL: normalized ion counts.

One important criterion of digestion efficiency is sequence coverage. The coverage of all the Ser, Thr, and Tyr residues is a prerequisite for the comprehensive mapping of phosphorylation sites. For example, dynein IC-1B from rat brain had 84% sequence coverage, as shown in **Figure 2.10**.

<u>MSDKSDLKAELEKQRLAQIREKKRKEEERKKKEADMQQKKEPVPDDS</u>	1-50
<u>DLDRKRRETEALLQSIGISPEPPLVPTPMSPSSKSVSTPSEAGSQDDLGP</u>	51-100
<u>LTRTLQWDTDPSVLQLQSDSELGRRLNKLGVSKVTQVDFLPREVVSYSKE</u>	101-150
<u>TQTPLATHQSEDEDEDEEMVEPKVGHDSELENQDKKQETKEAPPRELTEE</u>	151-200
<u>EKQQILHSEEFLLIFFDRTIRVIERALAEQSDIFFDYSGRELEEKDGDVQA</u>	201-250
<u>GANLSFNRFQFYDEHWSKHRVVTCDWLSLQYPELMVASYSNNEDAPHEPDG</u>	251-300
<u>VALVWNMKFKKTTPEYVFHCQSSVMSVCFARFHPNLVVGTYSGQIVLWD</u>	301-350
<u>NRSHRTPVQRTPLSAAAHTHPVYCVNVVGTQNAHNLITVSTDGKMCSWS</u>	351-400
<u>LDMLSTPQESMELVYNKSKPVAVTGMAFPTGDVNNFVVGSEEGTVYTACR</u>	401-450
<u>HGSKAGIGEVFEGHQGPVTGINCHMAVGPIDFSHLFVTSSFDWTVKLWTT</u>	451-500
<u>KHNKPLYSFEDNADYVYDVMWSPVHPALFACVDGMGRDLWNLNSDTEVP</u>	501-550
<u>TASVAIEGASALNRVRWAQGGKEVAVGDSEGRIWIYDVGELAVPHNDEWT</u>	551-600
<u>RFARTLVEIRANRADSEEEGAVELAA</u>	601-626

**Figure 2. 10** Sequence coverage map of dynein IC-1B from rat brain.

Underlined residues correspond to those included in detected peptides from both trypsin and AspN digestions.

### 2.5.3.3 Relative Peptide Quantitation

As shown in **Figure 2.6**, peak widths and heights can vary within a single LC-MS/MS analysis, which can be due to different digestion, separation, and ionization efficiencies. Relative quantitation of peptides was completed manually to estimate the abundance of each peptide detected. Relative quantitation was carried out by

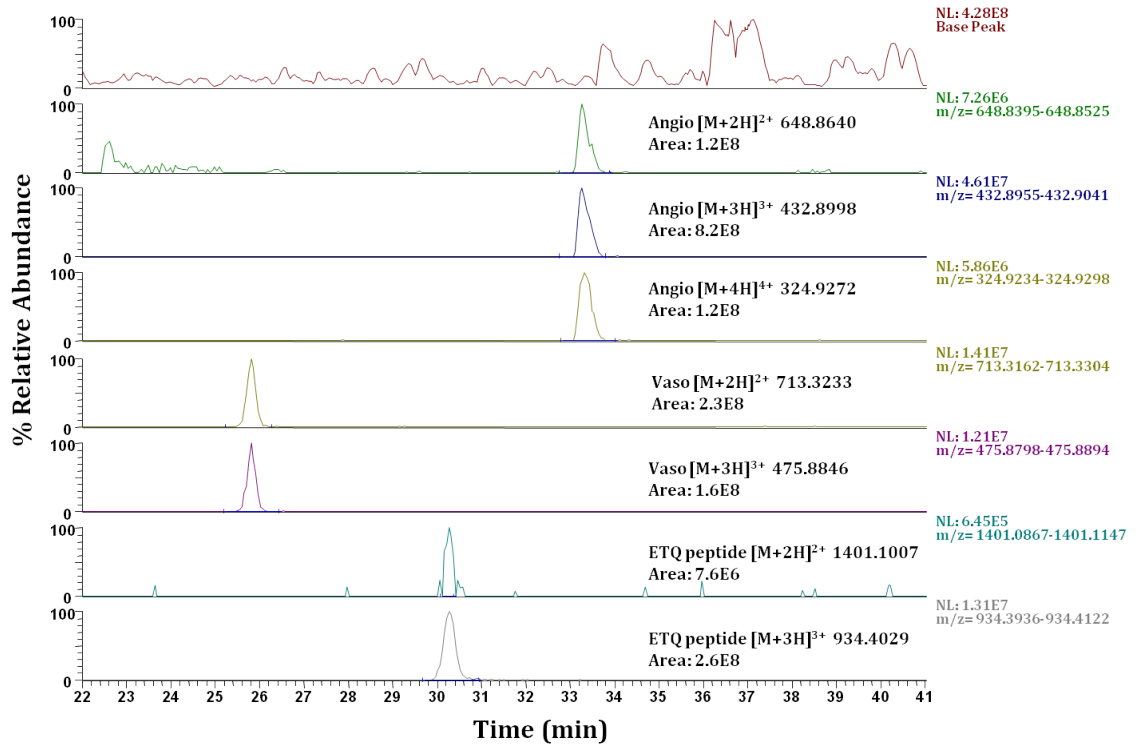
---

comparing the peak areas of two standard peptides, angiotensin and vasoactive intestinal peptide (known amount), to the peak areas of other peptides within the same MS analysis experiment.

The eluting profile for each peptide with a certain charge state was determined using an extracted ion chromatogram (XIC) with the calculated  $^{12}\text{C}$  monoisotopic mass and  $\pm 5\text{ppm}$  mass error. The integrated peak area of each ion peak was used to quantify the amount of peptide detected. Since one peptide usually has several charge states, the total amount of the peptide was quantified with the sum of the peak areas from ions of all the charge states. **Figure 2.11** displays XICs for angiotensin, vasoactive intestinal peptide, and the dynein IC peptide, ETQTPLATHQSEEDEEDEMVEPK. **Equation 2.3** is used to calculate the average peak area generated by a certain amount of angiotensin and vasoactive intestinal peptide. In this case 250 fmols of angiotensin and vasoactive intestinal peptide were loaded, so the average peak area of 250 fmols of angiotensin and vasoactive intestinal peptide was calculated by averaging the total areas of angiotensin and vasoactive intestinal peptide. Relative quantitation of each dynein peptide was determined using the averaged “250 fmols” peak area and the total peak area of that peptide, including all the charge states (**Equation 2.4**).

---





**Figure 2. 11** XICs of standard peptides angiotensin, vasoactive intestinal peptide, and ETQTPLATHQSEEDDEEMVEPK.

The charge state, experimental monoisotopic m/z and peak area for each ion are labeled. NL: normalized ion counts.

(2.3) *average peak area of standard peptides =*

$$\frac{\sum \text{area of angio} + \sum \text{area of vaso}}{2}$$

$$\sum \text{peak area of 250 fmols angio} = 1.2E8 + 8.2E8 + 1.2E8 = 10.6E8$$

$$\sum \text{peak area of 250 fmols vaso} = 2.3E8 + 1.6E8 = 3.9E8$$

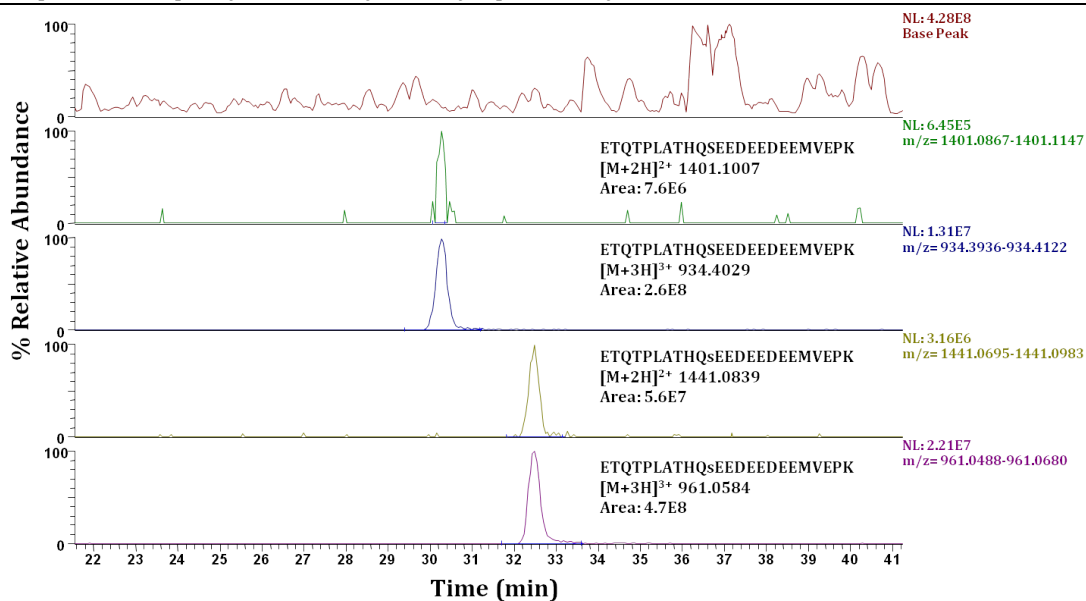
$$\text{average peak area of 250 fmols angio and vaso} = \frac{10.6E8 + 3.9E8}{2} = 7.3E8$$

$$(2.4) x = \frac{250 \text{ fmols}}{\text{average peak area of 250 fmol standard peptides}} \times \sum \text{peak area of peptide}$$

$$x(\text{ETQTPLATHQSEDEEDEEMVEPK}) = \frac{250 \text{ fmols}}{7.3E8} \times (7.6E6 + 2.6E8) = 92 \text{ fmols}$$

The percentage of modified peptide in the total amount of modified and unmodified peptide is important information in the investigation of the significance of this modification. Especially when the samples are treated differently, the changes in modification levels are the indication of a change in the biological process. Here the phosphopeptide, ETQTPLATHQsEEDDEEDEEMVEPK, is used as an example of the relative quantitation of phosphorylation. In **Figure 2.12**, XICs are shown for the unmodified and singly phosphorylated species. The percentage of the phosphorylated peptide is calculated as shown in **Equation 2.5**.

$$(2.5) \% PTM = \frac{\sum \text{area of modified peptide}}{\sum \text{area of unmodified peptide} + \sum \text{area of modified peptide}} \times 100\%$$



**Figure 2.12** XICs for dynein IC-1 peptide ETQTPLATHQSEEDDEEMVEPK and its singly phosphorylated form.

A lower case “s” indicates the phosphorylated Ser. The charge state, experimental monoisotopic m/z, and peak area for each ion are labeled. NL: normalized ion counts.

Using **Equation 2.5** it was determined that

ETQTPLATHQSEEDDEEMVEPK is phosphorylated approximately at 66%.

$$\% PTM = \frac{5.6E7 + 4.7E8}{7.6E6 + 2.6E8 + 5.6E7 + 4.7E8} \times 100\% = 66\%$$

**Tables 2.3, 2.4** and **2.5** show the phosphopeptides identified from rat brain, PC12 and cultured embryonic cortical neuron samples, respectively.

Phosphopeptides in each table were identified from several samples prepared at different times. There were no peptides from either IC-1A or IC-2A detected from the rat brain or cultured embryonic cortical neuron samples, which is consistent

with previous observations that they are not expressed in cultured embryonic

cortical neurons, and their expression in rat brain is at low levels (28, 29).

**Table 2. 3 Phosphopeptides identified from dynein purified from rat brain samples.**

The peptides were identified either from a tryptic digestion or an AspN digestion. The subunit source of each phosphopeptide and the percentage of phosphorylation are also listed. A lower case letter indicates the phosphorylated site.

Phosphopeptide (Trypsin)	Subunit	Percentage	Comments
ETEALLQSIGISPEPPLVPTPMsPSSK	IC-1	14.1%	Confirmed by IMAC
ETQTPLATHQsEEDEEDEEmVEPK	IC-1	44.6%	
VGHDsELENQDKKQETK	IC-1	94.2%	
ANRADsEEEGAVELAA	IC-1	53.6%	
SVSTPSEAGsQDDLGPLTR	IC-1	0.7%	
EAEALLQSMGLTTDSPIVPPPMsPSSK	IC-2	14.4%	
KPASVSPTTPPsPTEGEAS	LIC-1	33.4%	
LIRDFQEYVEPGEDFPAsPQRR	LIC-1	100.0%	
Phosphopeptide (AspN)		Percentage	Comments
DFLPREVVSYSKETQTPLATHQsEEDEEDEEMVEPKVGH	IC-1	65.5%	
DFLPREVVSYSKETQTPLATHQsEE	IC-1	62.1%	
DRITRKPASVSPTTPPsPTEGEAS	LIC-1	39.1%	
DRITRKPASVsPTtPPsPTEGEAS	LIC-1	16.6%	Tri-phosphopeptide, a mixture of different combinations
DRITRKPASVSPtPPsPTEGEAS			
DFQEYVEPGEDFPAsPQRRATAAQE	LIC-1	14.0%	

**Table 2. 4 Phosphopeptides identified from dynein purified from PC12 samples.**

The peptides were identified either from a tryptic digestion or an AspN digestion. The subunit source of each phosphopeptide and the percentage of phosphorylation are also listed. A lower case letter indicates the phosphorylated site.

Phosphopeptide (Trypsin)	Subunit	Percentage (%)
EAEALLQSMGLTTDSPIVPPPMsPSSK	IC-2	63.7
KPASVSPTTPPsPTEGEAS	LIC-1	43.8
Phosphopeptide (AspN)	Subunit	Percentage (%)
DSPIVPPPMsPSSKSVSTPSEAGSQ	IC-2	71.0

**Table 2. 5 Phosphopeptides identified from dynein purified from cultured embryonic cortical neuron samples.**

The peptides were identified from a tryptic digestion. The subunit source of each phosphopeptide and the percentage of phosphorylation are also listed. A lower case letter indicates the phosphorylated site.

Phosphopeptide (Trypsin)	Subunit	Percentage (%)
ETEALLQSIGISPEPPLVPTPMsPSSK	IC-1	69.2
EAEALLQSMGLTTDSPIVPPPMsPSSK	IC-2	47.5
LIRDFQEYVEPGEDFPAsPQRR	LIC-1	100%

### 2.5.4 Targeted Phosphopeptide Analyses

While data-dependent MS2 precursor selection can provide the MS2 spectra of as many possible peptides from an unknown peptide mixture, targeted MS2 precursor selection allows for as many possible MS2 spectra for a known peptide. In a data-dependent experiment, the five most abundant species are chosen for MS2. After a species is chosen three times within 30 s for an MS2, it will be put on an exclusion list and not be chosen anymore for 20 s, in order to obtain MS2 information for lower abundant peptides. However, in a targeted experiment, after each MS1 is taken, a user-selected species is always chosen as the precursor for the following MS2. Therefore, there will be more MS2 spectra generated for the selected peptide during the entire elution time of this peptide. Mono-phosphopeptide EAEALLQSMGLTTDSPIVPPPMSPSSK was identified with a phosphorylation site on one of the three serines near the C-terminus. However, it could not be localized due to weak fragmentation and low signal-to-noise ratio (S/N). In the data-dependent experiment, only 1-2 CAD spectra were generated for

the +3 species of this peptide. However, in the targeted experiment, 6 informative

CAD spectra were generated. **Figure 2.13** shows the averaged CAD spectra from the

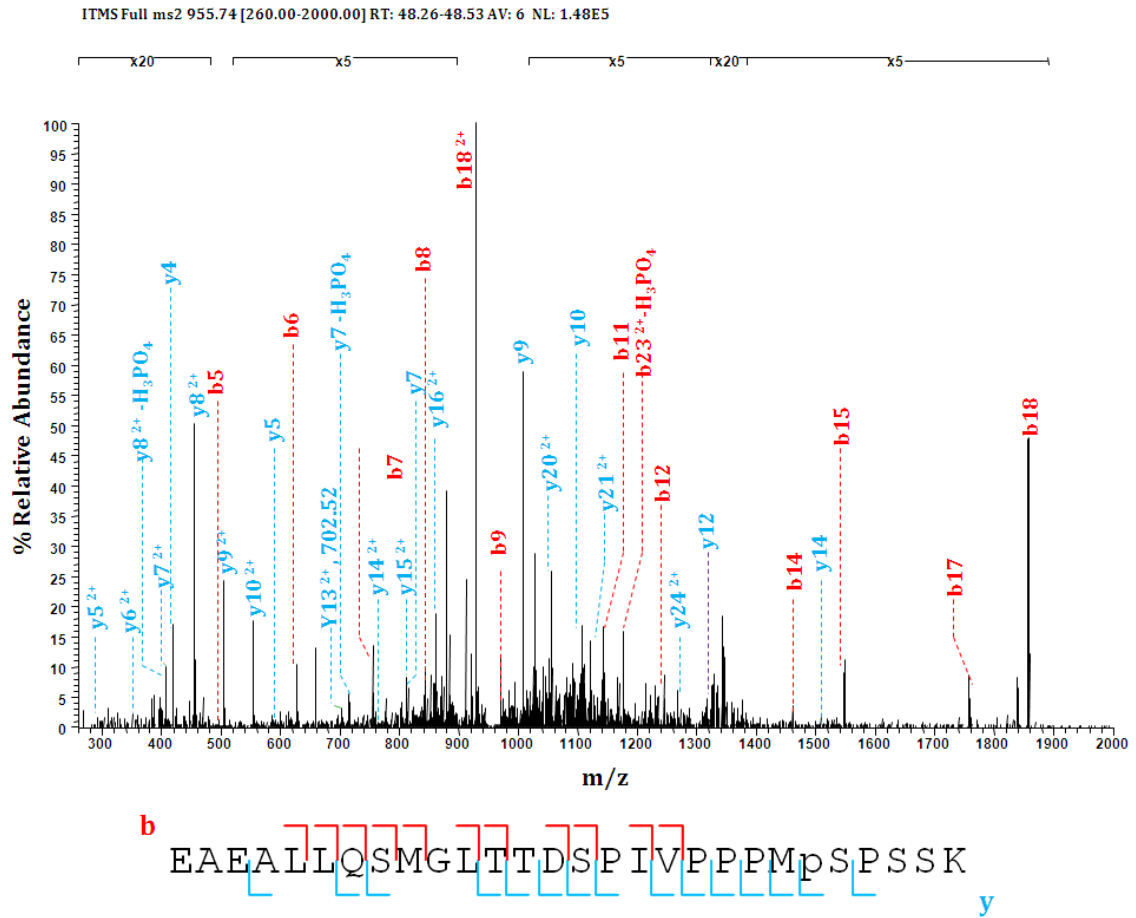
targeted experiment. It is clear that a non-modified  $y_4$  ion and a modified  $y_5$  ion

allow for the localization of the phosphorylation site at Ser81 (IC-2B)

EAEALLQSMGLTTDSPIVPPPMsPSSK. The observed enhancement in S/N increases

linearly with the square root of the number of scans recorded. Therefore,

generating and averaging more spectra is one way to enhance S/N.



**Figure 2. 13 Averaged CAD spectra of the phosphopeptide EAEALLQSMGLTTDSPIVPPPMsPSSK from a targeted precursor selection experiment.**

The targeted species was a +3 ion with mass range  $955.74 \pm 1.5$ Da. Fragment ions detected are labeled within the spectrum and in the peptide sequence. A lower case 'p' indicates that the following Ser is phosphorylated. The phosphorylated Ser was confirmed by the non-modified form of  $y_4$  and the modified form of  $y_5$ . RT: retention time, AV: number of spectra averaged, NL: normalized ion counts.

---

### 2.5.5 Phosphorylation Enrichment Analyses

Despite being one of the most common PTMs, phosphorylation is often substoichiometric. Also, the more abundant non-phosphopeptides are preferentially chosen for data-dependent MS2 analysis. Therefore, it is necessary to remove non-phosphopeptides to minimize the sample complexity and allow for the detection of low abundant phosphopeptides.

**Tables 2.6, 2.7, and 2.8** show the dynein phosphopeptides detected after IMAC enrichment from rat brain, PC12, and cultured embryonic cortical neuron samples, respectively. The phosphopeptides listed were not detected from previous MS analysis runs without IMAC enrichment. It is obvious that enriching for phosphopeptides greatly increases the chance for more phosphopeptides to be detected by the mass spectrometer. Phosphopeptides from LIC-2 and HC-1 were not detected before IMAC but were after the enrichment.

---



**Table 2. 6 Phosphopeptides detected from an IMAC enrichment experiment on dynein from rat brain.**

The peptides were identified either from a tryptic digestion or an AspN digestion. A lower case letter indicates the phosphorylated site. The subunit source of each phosphopeptide is also listed.

Phosphopeptide (Trypsin)	Subunit	Comments
ETEALLQsIGISPEPPLVPTPMsPSSK	IC-1	Di-phosphopeptide
ADsEEEGAVELAA	IC-1	
KEPVPDDsDLDR	IC-1	
KEPVPDDsDLDRK	IC-1	
KFMKDFQDYIEPEGcQGSPQRR	LIC-2	Not oxidized and oxidized methionine
KFMKDFQDYIEPEGcQGSPQR	LIC-2	Not oxidized and oxidized methionine
FMKDFQDYIEPEGcQGSPQRR	LIC-2	Not oxidized and oxidized methionine
FMKDFQDYIEPEGcQGSPQR	LIC-2	Not oxidized and oxidized methionine
DFQDYIEPEGcQGSPQRR	LIC-2	
DFQDYIEPEGcQGSPQR	LIC-2	
IAFIMDEsNVLDsGFLER	HC-1	Not oxidized and oxidized methionine
TDsTSDGRPAWMR	HC-1	
TRTDsTSDGRPAWmR	HC-1	
Phosphopeptide (AspN)	Subunit	Comments
DFLPREVVSYSKETQTPLATHQsEEDEEDEEMV EPKVGHDsELENQ	IC-1	Di-phosphopeptide, not very sure about the phosphorylation sites localization
DsELGRRRLNKLGVSKVTQV	IC-1	
DRITRKPASVsPTTPPsPTEGEAS	LIC-1	Di-phosphopeptide, a mixture of different combinations. There could be other possible combinations.
DRITRKPASVSPTtPPsPTEGEAS	LIC-1	
DYIEPEGcQGSPQRRGPLTSGSDE	LIC-2	
KPDSMVTNSsTENEAs	LIC-2	

**Table 2. 7 Phosphopeptides detected from IMAC enrichment experiment on dynein from PC12 sample.**

The peptides were identified either from a trypsin digestion or an AspN digestion. A lower case letter indicates the phosphorylated site. The subunit source of each phosphopeptide is also listed.

Phosphopeptide (Trypsin)	Subunit	Comments
SVStPSEAGSQDSGDGAVGSR	IC-2	
SVSTPSEAGsQDsGDGAVGSR	IC-2	Di-phosphopeptide
TLHWDTDPSALQLHsDSDLGR	IC-2	
D[SGDGAVGS]RRGPIKLGMAKITQV	IC-2	Mono-phosphopeptide, not sure of the phosphorylation site location (possible sites are in [])

QPPTAAGRPVDA <sub>s</sub> PR	LIC-1	
sGQKPVLSDVHAELDR	LIC-1	
SVSSNVASV <sub>s</sub> PIPAGSK	LIC-1	
VPGG <sub>s</sub> PRtPNR	LIC-1	Di-phosphopeptide
GGPASVPSA <sub>s</sub> PGTSVK	LIC-2	
GPLT[SGS]DEDNVALPLGDNVLTHNLGIPVLVVcTK	LIC-2	Mono-phosphopeptide, not sure of the phosphorylation site location (possible sites are in [])
GPLTsG <sub>s</sub> DEDNVALPLGDNVLTHNLGIPVLVVcTK	LIC-2	Di-phosphopeptide, not very sure
RGPLT[SGS]DEDNVALPLGDNVLTHNLGIPVLVVcTK	LIC-2	Mono-phosphopeptide, not sure of the phosphorylation site location (possible sites are in [])
KTGsPG <sub>s</sub> PSAGGVQSTAK	LIC-2	Di-phosphopeptide
RKDsAIQQQVANLQMK	HC-1	
KYtGEDFDEDLR	HC-1	
Phosphopeptide (AspN)	Subunit	Comments
DSPIVPPPM <sub>s</sub> PSSK[SVST]PSEAGSQ	IC-2	Di-phosphopeptide, not sure of the location of the second phosphorylation site, best guess is the last T (possible sites are in [])
DFPAsPQRRATAAQE	LIC-1	
DFQEYVEPGEDFPAsPQRRATAAQE	LIC-1	
DRITRKPASVSPTTPPsPTEGEAS	LIC-1	
DRDDsVVLPLGA	LIC-1	
DRITRKPASVSPTTPPsPTEGEAS	LIC-1	
DRITRKPASVsPTTPPsPTEGEAS	LIC-1	Di-phosphopeptide, a mixture of different combinations. There could be other possible combinations.
DRITRKPASVSPTtPPsPTEGEAS	LIC-1	
DYIEPEEGcQG <sub>s</sub> PQRRGPLTSGSDE	LIC-2	
DYIEPEEGcQG <sub>s</sub> PQRRGPL[TS GS]DE	LIC-2	Di-phosphopeptide, not sure of the location of the second phosphorylation site
DsAIQQQVANLQMKIVQE	HC-1	

**Table 2. 8 Phosphopeptides detected from IMAC enrichment experiment on dynein from cultured embryonic cortical neuron samples.**

The peptides were identified from a trypsin digestion. A lower case letter indicates the phosphorylated site. The subunit source of each phosphopeptide is also listed.

Phosphopeptide (Trypsin)	Subunit	Comments
ETEALLQsIGISPEPPLVPTPMsPSSK	IC-1	Di-phosphopeptide
LIRDFQEYVEPGEDFPAsPQR	LIC-1	
DFQEYVEPGEDFPAsPQRR	LIC-1	
DFQEYVEPGEDFPAsPQR	LIC-1	
KFMKDFQDYIEPEGcQGSPQRR	LIC-2	
KFMKDFQDYIEPEGcQGSPQR	LIC-2	
FMKDFQDYIEPEGcQGSPQRR	LIC-2	
FMKDFQDYIEPEGcQGSPQR	LIC-2	
DFQDYIEPEGcQGSPQRR	LIC-2	
DFQDYIEPEGcQGSPQR	LIC-2	

### 2.5.6 Identification of Neurotrophin-sensitive Dynein Phosphorylation

A previous analysis of PC12 samples performed by a former Hunt lab member, Erin Jeffery, showed that a brief exposure of PC12 cells to NGF increased the MS signal intensity of the phosphopeptide EAEALLQSMGLTTDSPIVPPPMsPSSK threefold to fourfold relative to unstimulated cells (31). This peptide is found in the IC-2B (S81) and IC-2C (S81) isoforms.

Neurons express both IC-1 and IC-2 isoforms. When dynein purified from rat brain was analyzed with MS, beside the IC-2 phosphopeptide EAEALLQSMGLTTDSPIVPPPMsPSSK, an IC-1 phosphopeptide, ETEALLQSMGLTTDSPIVPPPMsPSSK, was also identified (**Table 2.3**). This IC-1 phosphopeptide corresponds to the IC-2 phosphopeptide with one amino acid

difference. It is found in both the IC-1B (S80) and IC-1C (S80) isoforms.

Phosphorylated IC-1 S80 and IC-2 S81 were also found in cultured embryonic cortical neurons (**Table 2.5**).

It was estimated that in the rat brain sample, 14% of both the IC-1 S80 and the IC-2 S81 was phosphorylated. In the cultured embryonic cortical neurons with the depletion of BDNF, 33% of the IC-1 S80 and 27% of the IC-2 S81 was phosphorylated. In cultured embryonic cortical neurons with the addition of BDNF, the percentages increased to 69% and 48%, respectively, which was evidence that the newly identified phosphorylation sites are neurotrophin-sensitive.

To investigate IC phosphorylation and its role in the recruitment of dynein to Trk-containing signaling endosomes, a phospho-specific antibody ( $\alpha$ -pS-IC) to the phosphorylation site on the two ICs was prepared (the following discussions are based on the experiments performed in the Pfister Laboratory described in (31)). As shown in **Figure 2.14A**, the antibody reacted with the dynein ICs (74kD) in PC12 cell lysates and in dynein immunopurified from rat brain. Pretreatment of the immunopurified dynein with lambda phosphatase removed the immunoreactivity. The changes of IC phosphorylation levels in response to neurotrophin application in PC12 cells and cultured embryonic cortical neurons were then characterized using this antibody. In PC12 cells, the IC phosphorylation level had increased 2.3-fold to its maximal level after 20 minutes of treatment with NGF and it remained relatively constant at least for 40 minutes (**Figure 2.14B**). Since BDNF can be secreted by

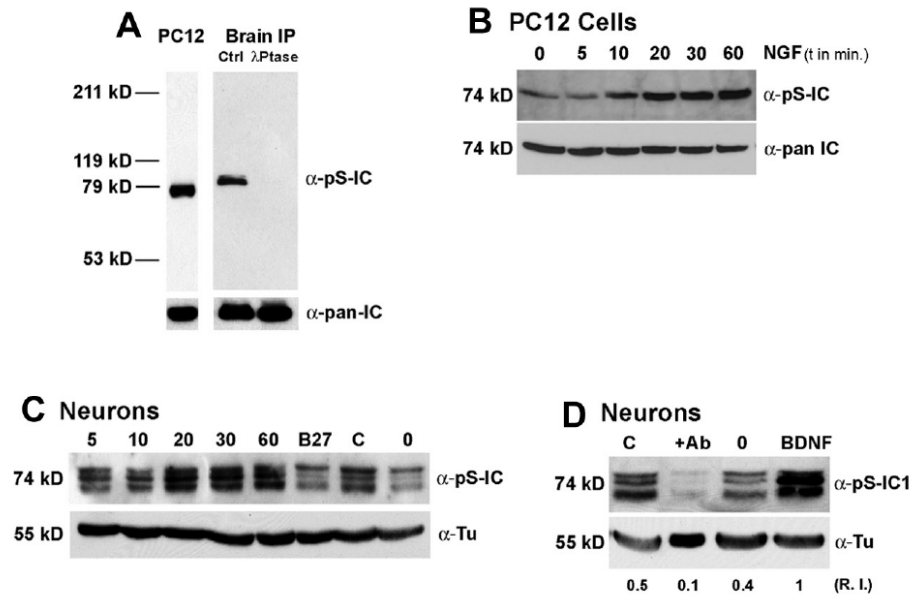
---

neurons, to determine whether cortical neuron IC phosphorylation was mediated by

BDNF, the BDNF secreted by neurons was depleted by replacing the conditioned medium (39). The phospho-specific IC antibody showed a reduced level of phosphorylation in cultured embryonic cortical neurons under the BDNF-depleted culture conditions compared to the normal culture conditions (**Figure 2.14C**).

Within 5 minutes after the addition of BDNF to the depleted medium, there was an increase in the phosphorylation levels of the ICs, and by 10 minutes it had increased 2.9-fold to its maximal level. To determine whether the residual phosphorylation IC observed in cells with the depletion of neurotrophin might be due to BDNF secreted by the cells, an antibody that functionally blocks the binding of BDNF to TrkB receptors was added to the fresh medium. The presence of the antibody in the neurotrophin depleted medium further lowered the amount of phospho-antibody staining, further demonstrating that the level of IC phosphorylation is sensitive to the concentration of BDNF in the medium (**Figure 2.14D**).

---



**Figure 2. 14 Neurotrophin stimulation of cytoplasmic dynein IC phosphorylation.**

A, Characterization of  $\alpha$ -pS-IC, an antibody to the PPMS\*PSSK phosphorylation site. Top left lane, blot of a PC12 cell lysate probed with  $\alpha$ -pS-IC. Right lanes, brain immunoprecipitate (IP) from a rat brain lysate. The IP was divided into two equal fractions, and half was treated with lambda phosphatase ( $\lambda$ Ptase) and the other with buffer as a control (Ctrl), both were probed with  $\alpha$ -pS-IC. The phospho-specific antibody detects at least two intermediate chain bands in the control sample, but none in the phosphatase-treated sample. Molecular weight markers are indicated on the left of the lane. Bottom, portions of identical gels probed with  $\alpha$ -pan-IC (a pan IC antibody that reacts with all intermediate chain isoforms), showing that IC is present in all three lanes. B, Time-dependent stimulation of cytoplasmic dynein IC phosphorylation by neurotrophin, NGF, in PC12 cells. PC12 cells were incubated with NGF for the indicated times (in minutes). Blots of cell lysates were probed with different antibodies:  $\alpha$ -pS-IC and  $\alpha$ -pan-IC (loading control). C, Time-dependent stimulation of neuronal dynein IC phosphorylation by the neurotrophin BDNF. Blots of lysates of cultured embryonic cortical neurons probed with  $\alpha$ -pS-IC and  $\alpha$ -Tu (tubulin antibody as a loading control). The numbers above the other lanes are the times in minutes after the addition of BDNF to cells in neurotrophin-depleted medium. B27, cells to which the B27 supplement was added to neurotrophin depleted medium (for 60 min); C, Control cells grown under standard culture conditions; 0, cells grown in neurotrophin depleted medium. D, An antibody that functionally blocks BDNF significantly reduces phosphorylation of neuronal ICs. Blots of neuron cultures grown under different conditions are shown. C, control cells grown under standard culture conditions; 0, cells depleted of neurotrophin by medium replacement; +Ab, cells depleted of neurotrophin by medium replacement and the addition of a function-blocking antibody to BDNF; and BDNF, cells grown in medium depleted of neurotrophin then incubated with BDNF for 20 minutes. The blot was probed with  $\alpha$ -pS-IC and  $\alpha$ -Tu (loading control). R.I., relative intensity (relative to the level of BDNF-treated cells) These figures are modified from (31) with permission.

### 2.5.7 ERK1/2 and Phosphorylation of the Cytoplasmic Dynein Intermediate

#### Chain

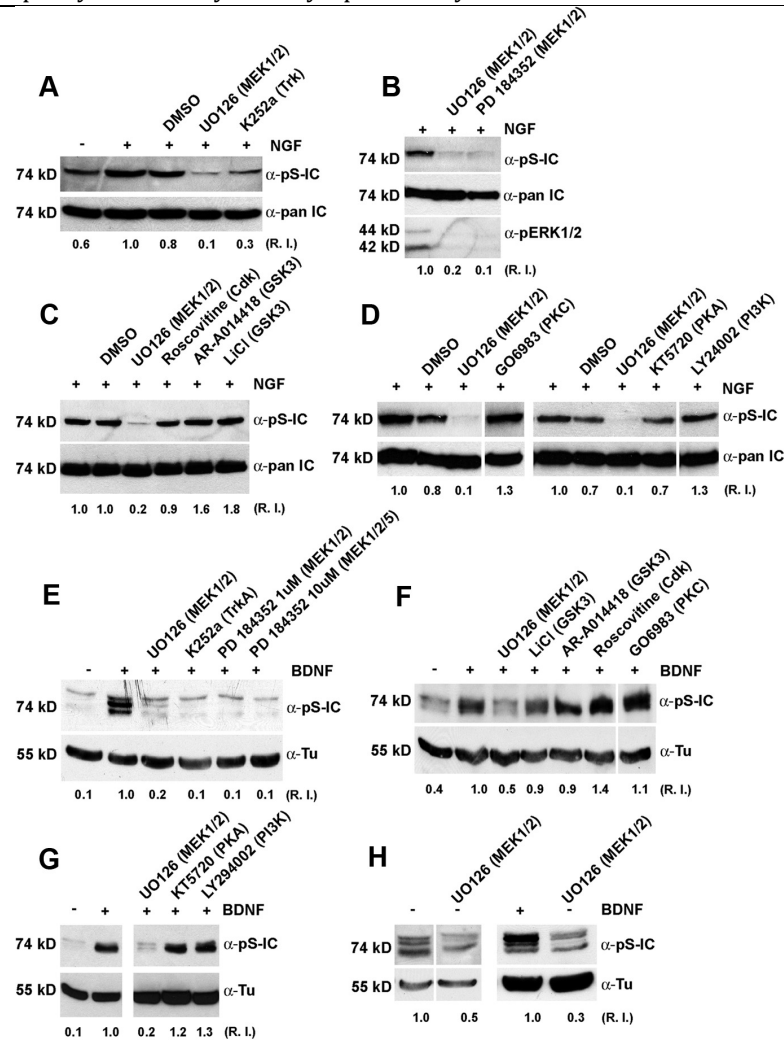
The effect of kinase inhibitors was investigated on the neurotrophin-sensitive IC phosphorylation to identify the kinases that are responsible for IC phosphorylation. **Figure 2.15A, E** shows that inhibition of the Trk kinases with K252a blocked the neurotrophin induced IC phosphorylation in PC12 cells and cultured embryonic cortical neurons, which provided evidence that the kinase was either Trk or a downstream effector of the Trk kinases. ERK, cell cycle kinases, and glycogen synthase kinase 3 (GSK3) were proposed by the ScanSite algorithm as kinases with the potential to phosphorylate the ICs on this site based on the motif (40). In particular, this site fell within a consensus sequence (PXSP) for phosphorylation by ERK (41). ERK1/2 was inhibited by the addition of U0126 which can block MEK, the activator of ERK1/2. The result showed that this inhibitor reduced the IC-2 phosphorylation by NGF in PC12 cells and the IC-1 and IC-2 phosphorylation by BDNF in neurons (**Figure 2.15A, E, H**). In addition, PD 184352 was found to have a similar effect although it is structurally unrelated to U0126 and inhibits MEK1/2 activation in a different way (**Figure 2.15B**). ERK5 is the downstream effector of TrkB that promotes neuronal survival (42). Two different concentrations of PD 184352 were used to distinguish between ERK1/2 and ERK5. While high concentrations of PD 184352 inhibit ERK1/2 and ERK5, lower concentrations of PD 184352 inhibit ERK1/2 (43). It was found that the lower

---

concentration of PD 184352 inhibited the neurotrophin-stimulated phosphorylation of both IC-1 and IC-2 in neurons and IC-2 in PC12 cells, providing evidence that ERK1/2 is the more likely IC kinase (**Figure 2.15B, E**). Inhibitors of GSK3 and the cell cycle kinases did not block the neurotrophin-stimulated IC phosphorylation in PC12 cells or neurons (**Figure 2.15C, F**). Inhibitors of PI3K and PKC, components of two other major Trk-stimulated signaling pathways, also did not inhibit neurotrophin-stimulated IC phosphorylation in the two model cell systems (**Figure 2.15D, F, G**). The Trk inhibitor K252a is known to inhibit protein kinase A (PKA). Therefore, KT5720, a PKA inhibitor that does not inhibit Trk kinase, was used to confirm that PKA did not contribute to IC phosphorylation (**Figure 2.15D, G**). To further demonstrate the role of ERK1/2 in the phosphorylation of the IC, constitutively active MEK1 was transfected into PC12 cells. It was observed that constitutively active MEK was sufficient to induce ERK activation and IC phosphorylation in the absence of NGF (data not shown).

---





**Figure 2. 15 IC phosphorylation requires active Trk and ERK1/2.**

The indicated kinase inhibitors and the target kinase or vehicle control (dimethyl sulfoxide, DMSO) were added to PC12 cells and neurons, and blots of cell lysates were probed with different antibodies as indicated ( $\alpha$ -pan-IC or  $\alpha$ -Tu as loading controls). The intensities of the IC bands were quantified relative to neurotrophin-treated cells. R.I., relative intensity. A–D, PC12 cells. The inhibitors were added to the PC12 cell culture medium for 30 minutes before addition of NGF. The cells were lysed 30 min after addition of NGF. A, Addition of NGF increased the level of IC phosphorylation 1.7 fold over control cells without NGF. Addition of the Trk kinase inhibitor, K252a, blocked the NGF induced stimulation of IC phosphorylation and reduced the IC phosphorylation level to below that of control cells. UO126, which blocks the activation of MEK and thus the activation of ERK1/2, also inhibited the neurotrophin stimulation of IC phosphorylation. B, Two structurally unrelated MEK inhibitors UO126 and PD184352 blocked ERK1/2 activation ( $\alpha$ -pERK1/2), and the NGF induced stimulation of IC phosphorylation ( $\alpha$ -pS-IC). C, Inhibitors of cell cycle kinases (cyclin-dependent kinase, Cdk) and GSK3 did not block the NGF induced stimulation of phosphorylation. D, Inhibitors of PKC, PI3K, and PKA did not block the NGF stimulation of phosphorylation. E–H, Neurons. Cultured embryonic cortical neurons were grown on dishes and depleted of growth factors. The inhibitors were added to the culture medium 30 min before addition of BDNF. The cells were lysed 15 min after addition of BDNF. E, Inhibitors of the Trk kinase and the MEK1/2 blocked the BDNF induced stimulation of IC phosphorylation. F, Inhibitors of Cdk, GSK3, and PKC did not block the BDNF-induced stimulation of phosphorylation. G, Inhibition of PKA or PI3K did not block IC phosphorylation. H, Cultured embryonic cortical neurons were grown under regular culture conditions. Left,

Blot showing a comparison of IC phosphorylation levels in control (untreated) cells to cells treated with U0126 for 30 min before lysis. Right, Blot comparing IC phosphorylation levels in cells treated with either BDNF or U0126 for 30 min before lysis. These figures are modified from (31) with permission.

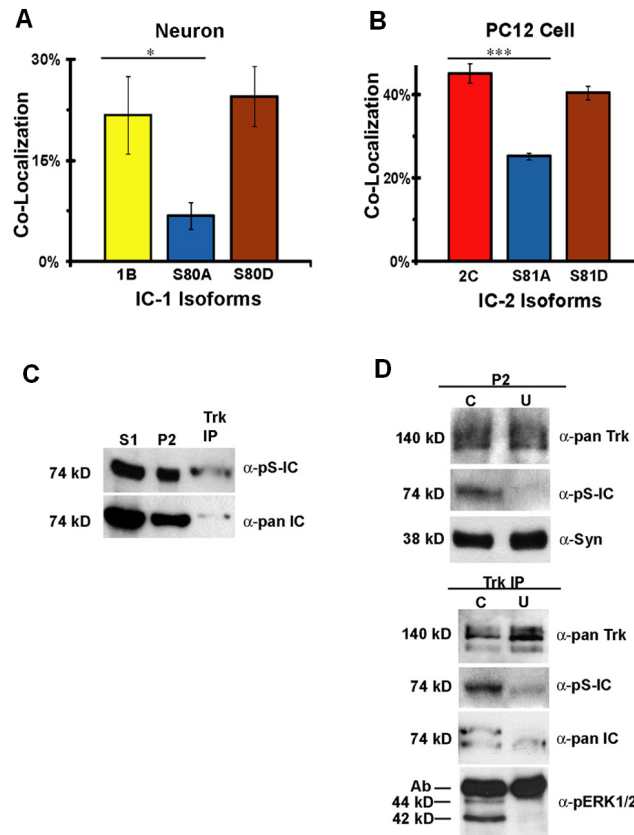
### 2.5.8 Intermediate Chain Phosphorylation Modulates Dynein Association with Trk-containing Organelles

To examine the functional significance of the observed neurotrophin-stimulated IC phosphorylation, the distributions of the ICs and Trks in living cells were analyzed by expressing fluorescently tagged Trks and ICs in hippocampal neurons and PC12 cells. After depleting the cells of the relevant endogenous IC isoform using siRNA, the levels of colocalization of Trk with the S/D phosphomimic and S/A dephosphomimic mutants relative to WT IC were compared. **Figure 2.16A** shows that in cultured hippocampal neurons, there was no significant difference in the extent of colocalization of TrkB puncta with WT IC-1B and IC-1B S80D phosphomimic mutation puncta. However, there was a significant decrease in the level of colocalization of the S80A dephosphomimic mutant with TrkB. In PC12 cell neurites, more than 40% of the WT IC-2C and IC-2C S81D phosphomimic colocalized with TrkA (**Figure 2.16B**) while the colocalization of the IC-2C S81A dephosphomimic mutant with TrkA was decreased significantly. These results are consistent with the hypothesis that IC phosphorylation enhances dynein binding to signaling endosomes.

---

Next, the association of dynein with immunoaffinity-purified Trk-containing organelles was investigated (**Figure 2.16C**). Firstly, the ratio of intensities of the phospho-IC to the pan-IC was quantified. No difference was observed when the cytosol (S1) and crude total membrane fractions (P2) were compared. However, there was a 2.4-fold enrichment of the phospho-dynein associated with immunopurified Trk-containing membrane-bound organelles relative to the cytosol and crude total membrane fraction (**Figure 2.16C**). Next, the amount of phospho-IC in P2 and in organelles purified from NGF-treated cells and cells that were pretreated with the MEK1/2 inhibitor U0126 before the addition of NGF were compared. As shown in **Figure 2.16D**, pretreatment with U0126 substantially reduced the amount of phospho-IC dynein found in crude total membrane fractions and the Trk-containing organelles. Most of the decrease in total IC associated with Trk-containing organelles isolated from U0126-treated cells was due to the loss of 92% of the upper, phospho-IC band. Importantly, there was no compensatory increase in the level of the lower, dephospho-IC band in the U0126-treated Trk-containing organelles. It was also observed that after pretreatment of cells with the MEK inhibitor U0126, there was almost no activated ERK1/2 associated with the Trk-containing organelles.

---



**Figure 2. 16 IC dephosphorylation reduces dynein binding to Trk-containing endosomes.**

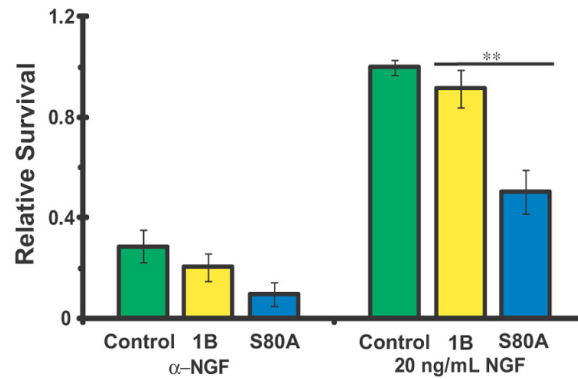
A. The dephosphomimetic IC-1B S80A mutation reduces the colocalization of dynein with TrkB in cultured neurons. Axons of living cells were simultaneously imaged in the GFP and RFP channels with DualView, and the percentage of overlapping puncta was quantified. B. The dephosphomimetic mutation IC-2C S81A has reduced colocalization with TrkA in living PC12 cells. C. Phosphorylated IC is enriched on immunoaffinity-purified Trk-containing organelles. PC12 cells were briefly stimulated with NGF to induce formation of signaling endosomes. The cells were homogenized, and the initial cytosol (S1) and crude total membrane fractions (P2) were obtained by centrifugation. Trk-containing membranes were isolated from the P2 by immunoaffinity (Trk IP), and the blots were probed with  $\alpha$ -pS-IC and  $\alpha$ -pan-IC. D. The MEK kinase inhibitor U0126 reduces the amount of dynein associated with Trk-containing organelles. Top, Blots of equal volumes of crude total membrane fractions (P2) prepared from NGF-stimulated control (C) and U0126 (U) pretreated PC12 cells were probed with pS-IC, and antibodies to the membrane proteins synaptophysin (Syn) and Trk (pan Trk) were used as loading controls. Treatment with U0126 reduced the amount of phospho-IC in the crude total membrane pellets by 90%. Bottom, Blots of immunoaffinity-purified Trk-containing organelles prepared from the P2s from control (C) or U0126 (U) pretreated cells were probed with pan-Trk, pS-IC, pan-IC, and pERK1/2 antibodies. Both immunoprecipitates contain Trk. Activated ERK1/2 was present in the immunoprecipitate from control cell membranes, but not the immunoprecipitate from U0126-treated cells. U0126 treatment of PC12 cells before the addition of NGF reduced the amount of the dynein ICs that copurified with Trk-containing organelles. This figure is modified from (31) with permission.

---

### 2.5.9 Phosphorylation and NGF-dependent Neuronal Survival

Dynein is known to mediate long-distance neurotrophin survival signaling (44). To determine the role of S80 phosphorylation on NGF-dependent survival, experiments were performed on cultured sympathetic neurons whose survival requires transport of NGF-TrkA endosomes. When cells were transfected with WT IC-1B-mRFP, there was no effect on cell survival compared to control transfected cells (**Figure 2.17**). However, when cells were transfected with the dephosphomimic mutant IC-1B-S80A-mRFP, a significant decrease in NGF-dependent neuronal survival was observed. This observation supports the role of this phosphorylation site in recruiting dynein to Trk-containing signaling endosomes. These data, in combination with the reduced colocalization of the dephosphomimic IC-1B mutant with the Trk-containing organelles and the reduced amount of dynein on immunoaffinity-purified Trk-containing organelles from U0126-treated cells, all support a model in which phosphorylation enhances dynein binding to organelles for retrograde transport.

---



**Figure 2. 17 Expression of the dephospho-IC-1B S80A mutant in sympathetic neurons reduces NGF-dependent cell survival.**

Sympathetic neurons were cultured and transfected with either the IC-1B WT or S80A dephospho-mutant by nucleofection, and the survival of the transfected cells in the presence of antibody to NGF ( $\alpha$ -NGF) or presence of 20 ng/mL NGF was determined. When the S80A mutant was expressed, there was a 45% decrease in cell survival relative to WT transfected cells in the presence of NGF. Green, Control transfected cells; yellow, WT IC-1B-mRFP transfected cells; blue, IC-1B-S80A transfected cells. This figure is modified from (31) with permission.

## 2.6 Conclusions

Mass spectrometric analysis of cytoplasmic dynein from rat revealed 35 phosphorylation sites on different subunits. Among all the phosphorylation sites, 33 sites were able to be unambiguously localized. This represents the largest number of phosphorylation sites reported on cytoplasmic dynein in any single study. The success of this PTM analysis was contributed by several factors. Antibody-based immunoprecipitation allows isolation of dynein from complex samples containing many other proteins. Phosphopeptide enrichment further reduced the sample complexity by the removal of non-phosphopeptides. On-line HPLC allowed for efficient separations of the entire dynein complex digests prior to MS/MS analysis. High resolution MS1 and the following MS2 using both CAD and ETD enabled

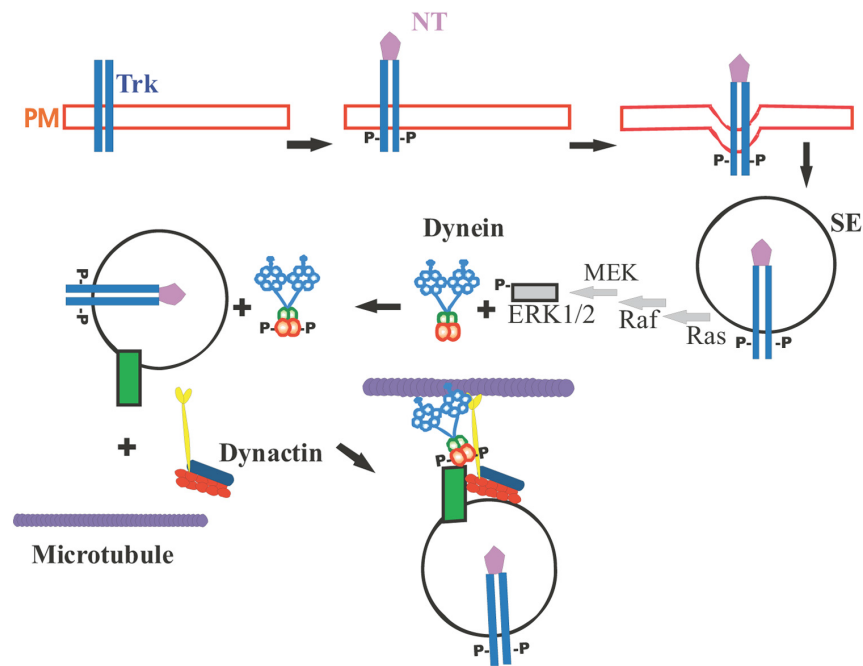
phosphopeptide identification and unambiguous site mapping of all but two sites.

Targeted MS2 experiments allowed for as many MS2 spectra generated for the interested peptide in order to enhance the S/N ratio to facilitate the confident assignment of fragment ions. Manual inspection of the complementary ETD and CAD data guaranteed the confidence in PTM characterization.

Two phosphorylated serines on dynein IC subunit were identified as neurotrophin-sensitive by mass spectrometry and antibody-based neurotrophin stimulation experiments. The following studies identified a mechanism that is used to recruit the cytoplasmic dynein to the signaling endosome upon neurotrophin stimulation for retrograde transport. Firstly, kinase prediction algorithms, specific pharmacological inhibition and *in vivo* expression of constitutively active MEK all support the identification of ERK1/2 as the most likely kinase to phosphorylate the ICs at IC-1 81S and IC-2 80S. The signaling pathway is shown in **Figure 2.18**. Upon binding to neurotrophin, the Trk kinase is auto-phosphorylated on its cytoplasmic tail and becomes internalized within a signaling endosome, with the neurotrophin remaining bound to Trk. Downstream of Trk, Ras is activated by GTP (guanosine triphosphate) binding, and it activates Raf, which activates MEK, which phosphorylates and activates the ERK1/2 MAP kinase. ERK1/2 phosphorylates the cytoplasmic dynein IC-1 and IC-2. Next, it was demonstrated that phosphorylation on IC-1 81S and IC-2 80S enhances dynein binding to the Trk signaling endosomes. Live cell imaging studies showed that the dephosphomimic mutations of the two

---

sites reduced the colocalization of dynein with TrkB. In immunoaffinity-purified Trk-containing organelles, enriched phosphorylated IC was detected. Also, MEK kinase inhibitor reduced the amount of phospho-IC dynein associated with Trk-containing organelles. Finally, a sympathetic neuron survival assay showed that the expression of the dephospho-IC-1B S80A mutant reduced NGF-dependent cell survival. All of these data support a model in which IC phosphorylation enhances dynein binding to Trk signaling endosomes for retrograde transport.



**Figure 2. 18 Model for ERK1/2 phosphorylation of ICs regulating cytoplasmic dynein binding to endosomes in response to neurotrophin.**

NT: neurotrophin. PM: plasma membrane. SE: signaling endosome. This figure is modified from (31) with permission.



---

## 2.7 References

1. Gibbons, I.R. (1995) *Dynein family of motor proteins - present status and future questions*. Cell Motil. Cytoskeleton 32, 136-144
  2. Pfister, K. K., Shah, P. R., Hummerich, H., Russ, A., Cotton, J., Annuar, A. A., King, S. M., and Fisher, E. M. C. (2006) *Genetic analysis of the cytoplasmic dynein subunit families*. PLoS Genet. 2, 11-26
  3. Eschbach, J. and Dupuis, L. (2011) *Cytoplasmic dynein in neurodegeneration*. Pharmacol. Ther. 130, 348-363
  4. Miller, F. D. and Kaplan, D. R. (2001) *On trk for retrograde signaling*. Neuron 32, 767-770
  5. Heerssen, H. M., Pazyra, M. F., and Segal, R. A. (2004) *Dynein motors transport activated trks to promote survival of target-dependent neurons*. Nat. Neurosci. 7, 596-604
  6. Bradshaw, R. A. and Dennis, E. A. (2010) *Chapter 1 - cell signaling: Yesterday, today, and tomorrow in Handbook of Cell Signaling, 2nd Ed.*, Academic Press, Salt Lake City UT
  7. Sideman, S. (2008) *Preface - cardiac control pathways: Signaling and transport phenomena*. Ann. NY Acad. Sci. 1123, XX-XLI
  8. Offenh usser, A. and Rinaldi, R. (2009) *Nanobioelectronics - for electronics, biology, and medicine*. Springer, New York, NY
  9. Choudhary, C. and Mann, M. (2010) *Decoding signalling networks by mass spectrometry-based proteomics*. Nat. Rev. Mol. Cell Biol. 11, 427-439
  10. Walsh, C. T. (2006) *Posttranslational Modification of Proteins: Expanding Nature's Inventory*, Roberts and Company Publishers, Greenwood Village, CO
  11. Cohen, P. and Goedert, M. (1998) *Engineering protein kinases with distinct nucleotide specificities and inhibitor sensitivities by mutation of a single amino acid*. Chem. Biol. 5, R161-R164
  12. Cohen, P. (2000) *The regulation of protein function by multisite phosphorylation – a 25 year update*. Trends Biochem. Sci. 25, 596-601
-

13. Delom, F. and Chevet, E. (2006) *Phosphoprotein analysis: From proteins to proteomes*. Proteome Sci. 4, 15
  14. Rylatt, D. B. and Cohen, P. (1979) *Amino-acid sequence at the site on rabbit skeletal-muscle glycogen synthase phosphorylated by the endogenous glycogen synthase kinase-2 activity*. FEBS Lett. 98, 71-75
  15. Kaufmann, H., Bailey, J. E., and Fussenegger, M. (2001) *Use of antibodies for detection of phosphorylated proteins separated by two-dimensional gel electrophoresis*. Proteomics 1, 194-199
  16. Posewitz, M. C. and Tempst, P. (1999) *Immobilized gallium(III) affinity chromatography of phosphopeptides*. Anal. Chem. 71, 2883-2892
  17. Anderson, L. and Porath, J. (1986) *Isolation of phosphoproteins by immobilized metal (Fe-3+) affinity-chromatography*. Anal. Biochem. 154, 250-254
  18. Kinoshita, E., Yamada, A., Takeda, H., Kinoshita-Kikuta, E., and Koike, T. (2005) *Novel immobilized zinc(II) affinity chromatography for phosphopeptides and phosphorylated proteins*. J. Sep. Sci. 28, 155-162
  19. Vale, R. D. (2003) *The molecular motor toolbox for intracellular transport*. Cell 112, 467-480
  20. Pfister, K. K., Fisher, E. M. C., Gibbons, I. R., Hays, T. S., Holzbaur, E. L. F., McIntosh, J. R., Porter, M. E., Schroer, T. A., Vaughan, K. T., Witman, G. B., King, S. M., and Vallee, R. B. (2005) *Cytoplasmic dynein nomenclature*. J. Cell Biol. 171, 411-413
  21. Burgess, S. A., Walker, M. L., Sakakibara, H., Knight, P. J., and Oiwa, K. (2003) *Dynein structure and power stroke*. Nature 421, 715-718
  22. Vallee, R. B., Williams, J. C., Varma, D., and Barnhart, L. E. (2004) *Dynein: An ancient motor protein involved in multiple modes of transport*. J. Neurobiol. 58, 189-200
  23. Dodding, M. P. and Way, M. (2011) *Coupling viruses to dynein and kinesin-1*. EMBO J. 30, 3527-3539
  24. Ginty, D. D. and Segal, R. A. (2002) *Retrograde neurotrophin signaling: Trk-ing along the axon*. Curr. Opin. Neurobiol. 12, 268-274
-

25. Salata, M. W., Dillman, J. F., Lye, R. J., and Pfister, K. K. (2001) *Growth factor regulation of cytoplasmic dynein intermediate chain subunit expression preceding neurite extension*. J. Neurosci. Res. 65, 408-416
26. Dillman, J. F. and Pfister, K. K. (1994) *Differential phosphorylation in-vivo of cytoplasmic dynein associated with anterogradely moving organelles*. J. Cell Biol. 127, 1671-1681
27. Huang, E. J. and Reichardt, L. F. (2003) *Trk receptors: Roles in neuronal signal transduction*. Annu. Rev. Biochem. 72, 609-642
28. Pfister, K. K., Salata, M. W., Dillman, J. F., Torre, E., and Lye, R. J. (1996) *Identification and developmental regulation of a neuron-specific subunit of cytoplasmic dynein*. Mol. Biol. Cell 7, 331-343
29. Pfister, K. K., Salata, M. W., Dillman, J. F., Vaughan, K. T., Vallee, R. B., Torre, E., and Lye, R. J. (1996) *Differential expression and phosphorylation of the 74-kDa intermediate chains of cytoplasmic dynein in cultured neurons and glia*. J. Biol. Chem. 271, 1687-1694
30. Greene, L. A. and Tischler, A. S. (1976) *Establishment of a noradrenergic clonal line of rat adrenal pheochromocytoma cells which respond to nerve growth-factor*. Proc. Natl. Acad. Sci. U. S. A. 73, 2424-2428
31. Mitchell, D. J., Blasier, K. R., Jeffery, E. D., Ross, M. W., Pullikuth, A. K., Suo, D., Park, J., Smiley, W. R., Lo, K. W. -, Shabanowitz, J., Deppmann, C. D., Trinidad, J. C., Hunt, D. F., Catling, A. D., and Pfister, K. K. (2012) *Trk activation of the ERK1/2 kinase pathway stimulates intermediate chain phosphorylation and recruits cytoplasmic dynein to signaling endosomes for retrograde axonal transport*. J. Neurosci. 32, 15495-15510
32. Geer, L. Y., Markey, S. P., Kowalak, J. A., Wagner, L., Xu, M., Maynard, D. M., Yang, X. Y., Shi, W. Y., and Bryant, S. H. (2004) *Open mass spectrometry search algorithm*. J. Proteome Res. 3, 958-964
33. Perkins, D. N., Pappin, D. J. C., Creasy, D. M., and Cottrell, J. S. (1999) *Probability-based protein identification by searching sequence databases using mass spectrometry data*. Electrophoresis 20, 3551-3567
34. Makarov, A., Denisov, E., Kholomeev, A., Baischun, W., Lange, O., Strupat, K., and Horning, S. (2006) *Performance evaluation of a hybrid linear ion trap/orbitrap mass spectrometer*. Anal. Chem. 78, 2113-2120
-

35. Marshall, A. G. (1985) *Fourier-transform ion-cyclotron resonance mass-spectrometry*. Acc. Chem. Res. 18, 316-322
36. Kruger, N. A., Zubarev, R. A., Carpenter, B. K., Kelleher, N. L., Horn, D. M., and McLafferty, F. W. (1999) *Electron capture versus energetic dissociation of protein ions*. Int. J. Mass Spectrom. 182, 1-5
37. Kruger, N. A., Zubarev, R. A., Horn, D. M., and McLafferty, F. W. (1999) *Electron capture dissociation of multiply charged peptide cations*. Int. J. Mass Spectrom. 185, 787-793
38. Cerda, B. A., Horn, D. M., Breuker, K., Carpenter, B. K., and McLafferty, F. W. (1999) *Electron capture dissociation of multiply-charged oxygenated cations. A nonergodic process*. Eur. Mass Spectrom. 5, 335-338
39. Chang, M. Y., Son, H., Lee, Y. S., and Lee, S. H. (2003) *Neurons and astrocytes secrete factors that cause stem cells to differentiate into neurons and astrocytes, respectively*. Mol. Cell. Neurosci. 23, 414-426
40. Obenauer, J. C., Cantley, L. C., and Yaffe, M. B. (2003) *Scansite 2.0: Proteome-wide prediction of cell signaling interactions using short sequence motifs*. Nucleic Acids Res. 31, 3635-3641
41. Songyang, Z., Lu, K. P., Kwon, Y. T., Tsai, L. H., Filhol, O., Cochet, C., Brickey, D. A., Soderling, T. R., Bartleson, C., Graves, D. J., DeMaggio, A. J., Hoekstra, M. F., Blenis, J., Hunter, T., and Cantley, L. C. (1996) *A structural basis for substrate specificities of protein ser/thr kinases: Primary sequence preference of casein kinases I and II, NIMA, phosphorylase kinase, calmodulin-dependent kinase II, CDK5, and Erk1*. Mol. Cell. Biol. 16, 6486-6493
42. Watson, F. L., Heerssen, H. M., Bhattacharyya, A., Klesse, L., Lin, M. Z., and Segal, R. A. (2001) *Neurotrophins use the Erk5 pathway to mediate a retrograde survival response*. Nat. Neurosci. 4, 981-988
43. Bain, J., Plater, L., Elliott, M., Shpiro, N., Hastie, C. J., Mclauchlan, H., Klevernic, I., Arthur, J. S. C., Alessi, D. R., and Cohen, P. (2007) *The selectivity of protein kinase inhibitors: A further update*. Biochem. J. 408, 297-315
44. Heerssen, H., Pazyra, M., and Segal, R. (2004) *Dynein motors transport activated trks to promote survival of target-dependent neurons*. Nat. Neurosci. 7, 596-604
-

---

### **3 Identification of Post-translational Modifications of Human**

#### **Desmoplakins in Regulating Interactions with Intermediate Filaments**

##### **3.1 Abstract**

Desmoplakin (DP) is a cytoskeletal linker protein that tethers the intermediate filament (IF) network to specialized intercellular junctions in the heart and skin, and is required for maintaining mechanical integrity of these tissues. It was previously demonstrated that phosphorylation of S2849 within the DP C-tail switches DP from being tightly bound to IF to a more dynamic state, allowing its translocation into junctions during desmosome assembly. The high density of Ser/Thr/Tyr residues within the C-tail raises the possibility that complex patterns of phosphorylation tune DP function. Using LC-MS/MS implemented with front-end ETD analysis of DP, six novel phosphorylation sites were identified, all of which fall within consensus motifs for glycogen synthase kinase (GSK) 3 phosphorylation cascades (S/T XXX S/T). Further, GSK3 inhibition led to DP behavior resembling that of the phospho-deficient DP mutant S2849G: it enhanced DP's interaction with the IF network and delayed DP trafficking and incorporation into desmosomes. The identification of multiple phosphorylation sites revealed the presence of putative arginine-claw conformations within the C-terminal tail, a compact protein fold characterized by the formation of salt-bridges between phosphorylated serine and arginine side chains. It is hypothesized that phosphorylation-induced arginine claw

---

formation within the DP regulatory region alters the dynamics of DP binding to IF.

Additionally, LC-MS/MS analysis revealed four novel DP arginine methylation sites including a putative arginine-claw residue R2834. A mutation in this site, R2834H, causes Arrhythmogenic Right Ventricular Cardiomyopathy (ARVC) and can lead to sudden cardiac death. Both global inhibition of protein methylation using inhibitors of protein arginine methyltransferases and targeted mutation of R2834H DP enhanced DP association with IFs and delayed desmosome assembly. LC-MS/MS and coimmunoprecipitation analyses indicated that R2834H ablates phosphorylation of other serines beside S2849 and weakens DP interactions with GSK3 in cells. Collectively, all the data are consistent with the idea that phosphorylation and methylation cooperate to control DP-IF interactions, and the interference with this regulatory machinery contributes to the pathogenesis of skin and heart diseases.

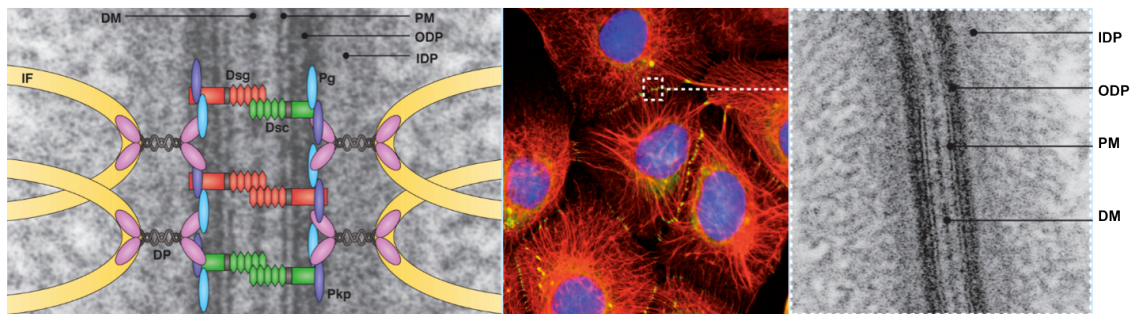
## **3.2 Introduction**

### **3.2.1 Desmosome and Desmoplakin**

Cell adhesion is essential for embryogenesis and for maintenance of tissue integrity. It also regulates critical processes such as cytoskeletal organization and cell polarity (1). Intercellular junctions are comprised of protein complexes that work together to form the cell adhesion by providing contact between neighboring cells (2). The desmosome is a specialized intercellular protein complex that

---

provides structural integrity to cardiac and epithelial tissues during mechanical stress (3-5). It connects the IF cytoskeleton to the plasma membrane at sites of cell-cell contact. Desmosomes are dynamic, adaptable organelles that are highly regulated and very ordered as seen by ultrastructural analysis using electron microscopy (**Figure 3.1**) (6, 7).



**Figure 3. 1 The desmosome and the desmoplakin.**

Left, cartoon representation of desmosomal components. The desmosome is composed of the desmosomal cadherins (Dsg and Dsc), the armadillo proteins (Pkp and Pg), and the plakins (DP). Middle, confocal immunofluorescence shows the keratin IF-desmosome attachments where keratin and DP are red and green, respectively, and DP-keratin co-localization is yellow. Right, the zoomed in ultra structural analysis of desmosome by electron microscopy. IF: intermediate filament; DM: dense midline; PM: plasma membrane; ODP: outer dense plaque. IDP: inner dense plaque. This figure is modified from (8) with permission.

Desmosomes are heterogeneous complexes composed of proteins from three different families: cadherins, armadillo proteins, and plakins (**Figure 3.1**).

Desmosomal cadherins form the adhesive core of cell-cell contacts by making direct interactions using their extracellular domains (3). As seen with electron microscopy, **Figure 3.1 shows** that the cadherins cluster together to form the dense midline (DM). The armadillo proteins bind to the tails of the cadherins and a member of the plakin family, desmoplakin (DP). DP is a key component of the

desmosome complex as its C-terminus anchors the desmosome to the IF networks

(4).

DP is composed of three regions, an N-terminal plakin domain, a central coiled-coil rod domain, and a C-terminal domain (**Figure 3. 2**). It is believed that the N-terminal plakin domain forms direct and indirect interactions with the armadillo proteins and the cadherins. The central coiled-coil rod domain mediates oligomerization. The C-terminal domain contains direct binding domains to IF and plays a key role in governing the spatial segregation of desmosomal components (9). It contains three plakin repeat domain A, B, and C, and a 68 residue flexible C-terminal tail (**Figure 3.2**). The B and C subdomains contain a positively charged groove for the direct interaction with the negatively charged IF network which provides mechanical strength to maintain the cellular architecture in eukaryotic cells (10). The DP C-terminal tail does not contain the direct contact sites with IF, however, it regulates the DP-IF interactions during desmosome assembly and disassembly. Regulation of the DP-IF interaction is crucial to attain adhesive strength and to maintain tissue integrity.

---



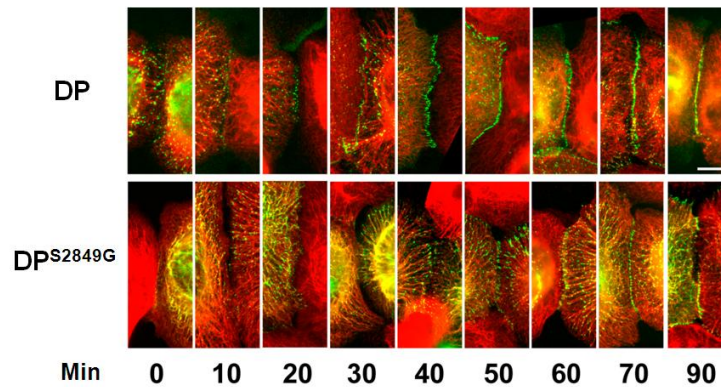


phosphorylation of DP and signals for desmosome formation, thus promoting

effective intercellular adhesion (18). While the literature has previously focused on the role of PKC $\alpha$  phosphorylation, it remains unknown if another kinase is directly responsible for the phosphorylation of DP (19, 20).

Previous data from the Green laboratory support the hypothesis that phosphorylation of the C-terminal tail is a molecular switch that drives DP translocation during desmosome assembly and promotes cellular adhesion. *In vitro* binding assays and immunofluorescence co-localization studies show that inhibiting phosphorylation at residue S2849 in the DP C-tail, through site-directed mutagenesis (DP<sup>S2849G</sup>), leads to a 10-fold increase in DP-IF associations (14, 21-23). Additionally, live cell imaging of the phospho-deficient DP mutant depicts a delayed trafficking to the desmosome at new cell-cell contacts (**Figure 4**) (12). It is critical that the intercellular junctions maintain the modularity to be quickly turned over during tissue remodeling, therefore, S2849G could have devastating effects in an *in vivo* model.

---



**Figure 3. 3 Phospho-deficient DP<sup>S2849G</sup> enhances DP-IF colocalization and delays DSM assembly in a Ca<sup>2+</sup> switch assay.**

Squamous carcinoma cell line (SCC9) transiently expressing DP-GFP or DP<sup>S2849G</sup>-GFP (green) were fixed at 10-minute intervals after a calcium switch (0–90 minutes) followed by keratin staining (red). DP-GFP appeared at cell borders within 10 minutes and became prominent by 40 minutes. DP<sup>S2849G</sup>-GFP appeared weakly at cell borders at 20 minutes, with a dampened increase in intensity at 40 and 60 minutes. This figure is modified from (12) with permission.

The Green laboratory has generated a polyclonal DP-specific antibody to recognize phosphorylation at residue S2849. Phosphorylation detection of DP, immunoprecipitated from <sup>32</sup>P orthophosphate labeled cells, indicates the presence of multiple phospho-isoforms on a two-dimensional gel. In addition, the high density of Ser/ Thr/ Tyr residues in the C-terminal tail raises the possibility that phosphorylation events at separate residues are cooperative and sequential (24). Moreover, phosphorylation cascades are emerging as a molecular mechanism to regulate protein-protein interactions as multisite phosphorylation adds a layer of complexity for intracellular signaling networks to control cellular processes.

GSK3 is the primary kinase that processively phosphorylates substrates at multiple sites in sequential cascades. This kinase recognizes substrates that are

primed by the phosphorylation from a separate kinase at a location four amino acids

C-terminal to the site (25). Following the initial priming phosphorylation event, GSK3 processively phosphorylates S/T residues with the consensus motif "S/T XXX S/T" in a C- to N-terminal direction sequentially for up to 14 residues in a single polypeptide chain (25). This mechanism mediates a diverse set of cellular functions including protein translation, gene transcription, microtubule organization, and glycogen metabolism (26-28). Found in both cytoplasmic and nuclear compartments, GSK3 specifically localizes to the plasma membrane to target many of its substrates including beta-catenin and desmosome armadillo protein, plakoglobin (Pg) (29). Actin crosslinking factor-7 (ACF-7) is another member of the plakin gene family and contains sequence homology to DP in the GSR repeat domain (30). GSK3 sequentially phosphorylates ACF-7 in its GSR repeat domain to coordinate microtubule dynamics and cell polarization. Similarly, the GSR repeat domain in the DP C-terminal tail contains 15 GSK3 consensus sites (S/T XXX S/T). Therefore, DP C-terminal tail could be a favorable substrate for GSK3 phosphorylation.

### 3.2.3 Desmoplakin Mutations in Diseases

Over 50 genetic mutations have been identified implicating DP as a key target for inherited diseases. These mutations in DP have been shown to promote the pathogenesis of skin-blistering disorders and cardiac myopathies, and even cause death with severe loss of function (31, 32). A critical mutation of an Arg to a His in

---

### 3.2.4 Central Hypothesis

[illegible]

S2849 (green) is the phosphorylation site that is reported to be critical to DP function (12). R2834 (marked with a star) is the mutated site found in ARVC patients (29). Amino acids highlighted in red are previously reported phosphorylation sites (11). Orange arrows denote GSK3 phospho-targets that proceed sequentially. All the serine residues are in bold.

---

### 3.3 Materials

#### 3.3.1 Reagents

##### **Mallinckrodt, Inc. (Paris, KY)**

Acetonitrile, HPLC grade,  $\geq 99.8\%$  purity

##### **New England Biolabs, Inc. (Ipswich, MA)**

Histone H2A, Human, Recombinant

##### **PQ Corporation (Valley Forge, PA)**

Kasil® 1624

##### **Roche Diagnostics (Indianapolis, IN)**

Endoproteinase AspN

##### **Sigma Aldrich (St. Louis, MO)**

Ammonium bicarbonate

Angiotensin I acetate salt hydrate,  $\geq 90\%$  purity (human)

Azulene

Cyanogen bromide, CNBr

Fluoranthene, 99% purity

---

---

Formamide

Glacial acetic acid,  $\geq 99.9\%$  purity

Potassium ferricyanide,  $\text{K}_3\text{Fe}(\text{CN})_6$

Sodium thiosulfate pentahydrate,  $\text{Na}_2\text{S}_2\text{O}_3 \cdot 5\text{H}_2\text{O}$

Trifluoroacetic acid, TFA

Vasoactive intestinal peptide fragment 1-12,  $\geq 97\%$  purity

### **Thermo Scientific (Rockford, IL)**

Pierce® Water, LC-MS Grade

### **3.3.2 Equipment & Instrumentation**

#### **Agilent Technologies (Palo Alto, CA)**

HP1100 high performance liquid chromatograph

#### **Branson (Danbury, CT)**

Branson 1200 Ultrasonic Bath

#### **Dr. Maisch (Ammerbuch-Entringen, Germany)**

Reprosil-Pur 120 C18 (3  $\mu\text{m}$  diameter, 100 Å pore size)

---

---

Reprosil-Pur 120 C18 (5  $\mu\text{m}$  diameter, 100 Å pore size)

**Eppendorf (Hauppauge, NY)**

5415R Benchtop Centrifuge

**Labconco Corporation (Kansas City, MO)**

Centrivap Concentrator

**New Objective, Inc. (Woburn, MA)**

PicoClear<sup>®</sup> microcapillary union

**Polymicro Technologies, LLC (Pheonix, AZ)**

360  $\mu\text{m}$  o.d. x 50  $\mu\text{m}$  i.d. polyimide coated fused silica capillary

360  $\mu\text{m}$  o.d. x 75  $\mu\text{m}$  i.d. polyimide coated fused silica capillary

**Sutter Instrument Co. (Navato, CA)**

P-2000 microcapillary laser puller

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

LTQ-Orbitrap hybrid mass spectrometer

LTQ-Orbitrap Velos hybrid mass spectrometer

---



### **3.4 Methods**

The sample preparation and biology/pathology-based experiments were completed by Lauren Albrecht, graduate student in Dr. Kathleen J. Green's laboratory in the Department of Pathology at the Northwestern University, Feinberg School of Medicine.

#### **3.4.1 Cell Culture and Affinity Purification**

Human Embryonic Kidney (HEK) 293 cells were plated onto 8 10-cm plates, grown to 85 % confluency, and transfected with 144 µg of DP-Stag DNA and TurboFect amine. The transfected cells were then grown for 48 hours and lysed in 400 mL MS lysis buffer (50 mM Tris-HCl pH 8, 1% IGEPAL CA-630, 150 mM NaCl, 1 mM EDTA, 1 mM EGTA, 1X protease inhibitor mix, 3 mM PMSF, 1 mM dithiothreitol, 1X phosphatase inhibitor cocktail IV). Cells were spun down and the supernatant was combined with 80 µL (in total) S protein-agarose beads for 2 hours in four different tubes (each tube contains cells from two 10 cm dishes). The beads were spun down at 1000 rpm for 2 minutes and then the supernatant was removed. S protein-agarose beads were washed with wash buffer I (50 mM Tris-HCl pH 8, 1 M NaCl, 2 mM PMSF, 1% IGEPAL CA-630) five times and wash buffer II (50 mM Tris pH 8, 500 mM NaCl, 2 mM PMSF) three times. The final purified DP-S-tag beads were stored in 20 µL of wash buffer II.

---

To generate DP-S-tag mutants R2834H and S2849G, HEK 293 cells were transfected with 15  $\mu$ L R2834H or S2849G DP-S-tag DNA per dish. To generate the co-expression of HA-tagged constitutively active GSK3 and wild type (WT) DP-S-tag, HEK 293 cells were transfected with 7.5  $\mu$ L HA-CA-GSK3 DNA and 7.5  $\mu$ L of WT DP-S-tag DNA. The following affinity purification of the DP-S-tag was as previously described.

### **3.4.2 Immunoblot Analyses**

After affinity purification, DP-S-tag was resolved on a 10% Tris glycine gel after which proteins were transferred to a 0.45  $\mu$ m nitrocellulose membrane and immunoblotted against DP-S-tag or phosphorylated DP-S-tag. Immunoblots were then scanned and analyzed using Image J (33).

### **3.4.3 In-solution Protein Cleavage and Sample Preparation for LC-MS/MS Analyses**

Affinity purified DP-S-tag was retained on agarose beads and washed three times with 100  $\mu$ L of water. DP-S-tag was first cleaved C-terminal to Met using cyanogen bromide (CNBr). The DP-S-tag beads were suspended in 50  $\mu$ L of 70% TFA in water and one small crystal (<1mg) of CNBr was added to the solution. The reaction proceeded 20 hours at 4°C. Following cleavage, the supernatant containing peptides was collected and dried down in a centrivap concentrator. For AspN digestion, the sample was resuspended in 30  $\mu$ L of 100 mM ammonium bicarbonate

---

and AspN was added at an enzyme to substrate ratio of 1:20. The digestion was carried out for 6 hours at room temperature and then quenched by adding acetic acid to adjust the pH to 3. The acidified sample was stored at -35 °C.

#### **3.4.4 In-gel Proteolytic Digestions and Sample Preparation for LC-MS/MS**

##### **Analyses**

In-gel digestion was performed on the silver-stained gel band containing DP-S-tag. The gel band was destained by incubating in 50 µL of destaining solution (30 mM  $K_3Fe(CN)_6$  and 100 mM  $Na_2S_2O_3 \cdot 5H_2O$  at 1:1 volume ratio) for 3 minutes, in 100 µL of  $H_2O$  for 20 minutes for three times and finally in 100 mM ammonia bicarbonate for 20 minutes. The gel band was dehydrated in acetonitrile for 10 minutes on shaker and dried down in a centrivap concentrator. The proteins were reduced in 50 µL of 10 mM dithiothreitol for 1 hour at 37°C and alkylated in 50 µL of 55mM iodoacetamide in the dark for 45 minutes on a shaker. Next, the gel band was twice dehydrated in 50 µL of acetonitrile for 10 minutes and dried down in a centrivap concentrator, followed by incubation with 50 µL of 12.5 ng/µL trypsin to be rehydrated on ice. After 45 minutes, the supernatant was replaced with 100 mM ammonia bicarbonate and the digestion was carried out at 37°C for 6 hours. Three different solutions were made to extract peptides from the gel: extraction solution X contained 5% formic acid and 25% acetonitrile; extraction solution Y contained 5% formic acid and 50% acetonitrile; and extraction solution Z contained 5% formic acid and 75% acetonitrile. Peptide extraction included the following procedures.

---

Firstly, 25  $\mu\text{L}$  of 100 mM ammonia bicarbonate was added to the sample tube and then all of the supernatant was transferred to a new Eppendorf tube. The supernatants from each of the following steps were all transferred to the same tube. Next, the gel band was incubated with 50  $\mu\text{L}$  of extraction solution X for 15 minutes and the supernatant was collected. The gel band was then rehydrated with  $\text{H}_2\text{O}$  for 10 minutes and the supernatant was collected as well. This step was repeated for extraction solutions Y and Z except that no rehydration was needed for the last extraction. The collected supernatant was dried down, resuspended in 0.1% acetic acid and stored in  $-35\text{ }^\circ\text{C}$ .

#### 3.4.5 HPLC Capillary Column Assembly

An analytical column was constructed using 360  $\mu\text{m}$  o.d. x 50  $\mu\text{m}$  i.d. fused silica capillary with a 2 mm Kasil® 1624 frit. The analytical column was packed with 7 cm of C18 resin (3  $\mu\text{m}$  diameter, 100 Å pore size) and equipped with a laser-pulled electrospray emitter tip. A precolumn was constructed using 360  $\mu\text{m}$  o.d. x 75  $\mu\text{m}$  i.d. fused silica capillary with a 2 mm Kasil® 1624 frit and packed with 5 cm of C18 resin (5  $\mu\text{m}$  diameter, 100 Å pore size). Both columns were conditioned with several picomoles of angiotensin and histone H2A tryptic digests three times to optimize chromatographic performance. The two columns were connected with a microcapillary union. The chromatographic performance was examined by analyzing 1 pmol of histone H2A tryptic digests prior to use to assure separation efficiency.

---

### 3.4.6 HPLC-MS/MS Analyses

A 5% fraction of the DP-S-tag digest (800 fmol) was pressure-loaded onto the precolumn on a pressure bomb at a flow rate of 1  $\mu$ L/min. Internal standard peptides (250 fmol, angiotensin and vasoactive intestinal peptide) were added to the sample prior to loading for relative quantification. The sample was desalted by rinsing the precolumn with 0.1 M acetic acid on an HPLC for 10 minutes at a flow rate of 1  $\mu$ L/min. The precolumn was then connected to the analytical column using a microcapillary union. Peptides were gradient eluted at a flow rate of 60 nL/min from the HPLC column and electrosprayed into the mass spectrometer. The LC gradient was as follows: 0-60% of solvent B in 60 min, 60-100% solvent B in 10 min, 100% solvent B hold for 5 min, 100% - 0% B in 5 min, 100% Solvent A for 10 min (solvent A: 0.1 M acetic acid in water; solvent B: 70% acetonitrile and 0.1 M acetic acid in water).

Data-dependent MS/MS analyses included a high resolution (60,000 at  $m/z$  400) MS1 scan in the Orbitrap and 10 low resolution MS2 scans in the LTQ. The top five most abundant ions were selected for MS2 analysis. For each precursor, both CAD and ETD were performed to obtain complementary spectra for optimal peptide identification. Parameters for data dependence MS/MS data acquisition included monoisotopic precursor selection enabled, charge state rejection enabled for  $z = +1$  and unassigned species, dynamic exclusion enabled: repeat count of 3, repeat duration of 30 s, exclusion list duration of 20 s, exclusion list size of 500, and

---

precursor isolation window of 3 m/z. ETD reaction time was 30 ms. FTMS AGC

target was 5e5; ITMSn AGC target was 6e4; and ETD reagent target was 3e5.

Targeted ETD MS analysis included high resolution MS1 (60,000 at m/z 400) and MS2 (30,000 at m/z 400) scan in the Orbitrap-velos. Precursor isolation window for MS2 was 3 m/z. ETD reaction time was 30 ms. Fragment ions resulting from 10 rounds of ETD were accumulated in the C-trap prior to detection in the Orbitrap. FTMS AGC target was 1e5, FTMSn AGC target was 5e4, and ETD reagent AGC target was 3e5.

#### **3.4.7 Mass Spectrometric Data Analyses**

MS/MS data from data-dependent experiments were searched against a database containing the sequence of DP-S-tag or its mutations using the Open Mass Spectrometry Search Algorithm (OMSSA) software. MS/MS data were searched using either b and y fragment ions from CAD spectra or c and z• fragment ions from ETD spectra. Database searches used no enzyme specificity and an E-value cutoff of 1. Variable modifications included in the database search were oxidation of Met, phosphorylation of Ser, Thr and Tyr. A precursor mass tolerance of  $\pm 0.01$  Da was used for high-resolution MS1 data and a fragment ion mass tolerance of  $\pm 0.35$  Da was used for low resolution MS2 data. The removal of reduced charge species within ETD data prior to searches was selected. All peptide sequences were manually validated by inspection of the accurate precursor mass and the

---

corresponding CAD or ETD spectra. Some peptide sequences were also determined by *de novo* interpretation of the CAD and ETD spectra, assisted with the use of Mascot sequence queries. Mascot search conditions utilized were: precursor mass tolerance of  $\pm 0.01$  Da, fragment ion mass tolerance of  $\pm 0.30$  Da, no enzyme specificity, variable phosphorylation on Ser/Thr/Tyr, and variable oxidation on Met.

MS/MS data from targeted ETD experiments were manually validated by inspection of the accurate precursor mass and the corresponding ETD spectra.

### 3.5 Results & Discussion

#### 3.5.1 DP-S-tag Purification

Since the expression of full length DP is challenging, a C-terminal DP truncation construct fused to an S-tag (DP-S-tag) was transfected and expressed in HEK 293 cells. The sequence of DP-S-tag is shown in **Figure 3.5**. Purification of DP-S-tag was based on the high affinity interaction between the N-terminal S-peptide tag on the expressed protein and S-protein bound to agarose beads.

```

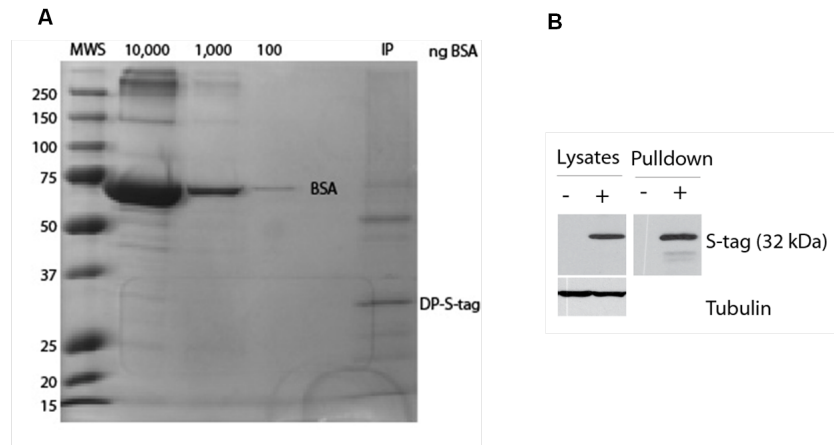
      MAHHHHHHSS GKETAAAKFE RQHMSFPPPS GLVPRGSAGS GTIDDDDKSP GFSSTISELV
2628-2687 LEKISITEGI ERGIVDSITG QRLLEAQACT GGIIHPTTGQ KLSLQDAVSQ GVIDQDMATR
2688-2747 LKPAQKAFIG FEGVKGKKKM SAAEAVKEKW LPYEAGQRFL EFQYLTGGLV DPEVHGRIST
2748-2807 EEAIRKGFID GRAAQRLQDT SSYAKILTCP KTKLKISYKD AINRSMVEDI TGLRLLEAAS
2808-2867 VSSKGLPSPY NMSSAPGSRS GSRSGSRSGS RSGSRSGSRR GSFDATGNSS YSYSYSFSSS
2868-2871 SIGH

```

**Figure 3. 5 DP-S-tag sequence.**

The sequence in grey corresponds to amino acids contributed by the S-tag (underlined sequence) and extended linker region. Sequence is numbered based on full length human DP isoform I.

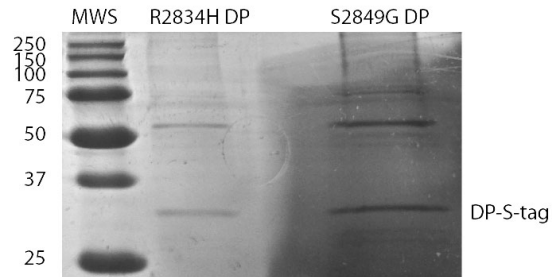
Epitope protein tags provide the opportunity to selectively purify the protein of interest. The S-protein affinity tag exploits the high affinity binding characteristics of S-protein and S-peptide, two products generated by subtilisin cleavage of Ribonuclease A (RNaseA) (24). **Figure 3.6** displays purified DP-S-tag as visualized using SDS-PAGE and Western blot analysis.



**Figure 3. 6 SDS-PAGE and Western blot of affinity purified DP-S-tag.**

A. SDS-PAGE of affinity purified DP-S-tag. The first lane was loaded with molecular weight standards with molecular weight (kDa) labeled on the left. The next three lanes were loaded with different amount of BSA for relative quantification. The right lane was loaded with affinity purified DP-S-tag and the DP-S-tag band was labeled (32 kDa). The gel was stained with coomassie. B. Western blot analysis of DP-S-tag. DP-S-tag was probed with antibody against S-tag from cell lysates and affinity purified DP-S-tag sample (pulldown). "+" indicates the sample was from cells transfected with DP-S-tag DNA and "-" indicates control sample from cells with no transfection. Tubulin was probed with tubulin antibody as loading control.





**Figure 3. 7 SDS-PAGE of affinity purified DP-S-tag mutates R2834H and S2849G.**

The first lane was loaded with molecular weight standards with molecular weight (kDa) labeled on the left. The second and third lanes were loaded with affinity purified DP-S-tag mutates R2834H and S2849G, respectively. The DP-S-tag bands were labeled (32 kDa). The gel was stained with coomassie.

### 3.5.2 Protein Cleavage and LC-MS/MS Analyses of DP-S-tag

Since the primary goal of this project was to characterize PTMs on the C-terminal tail GSR repeat domain, generating peptides of proper size in that region became a critical task. Although 14 residues in the C-terminal tail region had been previously identified as phosphorylation sites, it was possible that more remain to be identified, because those sites were identified from a trypsin digestion, and many serines were not mapped in the region rich in arginines (11). Therefore, another protein cleavage strategy is required to generate peptides with proper sizes and charge states to cover the DP C-terminal tail GSR repeat domain region. Also, generating large peptides with multiple phosphorylatable sites is ideal so that simultaneous phosphorylations on different sites can be detected. Analysis of the amino acid composition of DP-S-tag C-terminal tail GSR domain region resulted in the selection of a combination of AspN and CNBr as cleavage reagents to generate three major peptides as shown in **Figure 3.8**. AspN is an endoprotease that cleaves

C-terminal to Asp. CNBr is a small molecule that non-enzymatically cleaves proteins

C-terminal to Met residues and converts them into homoserine lactones or

homoserines (**Figure 3.9**). The number of charges that can be carried by each

peptide was predicted by counting the number of basic residues (Arg/Lys/N-

terminus). Then based on charge state, the proper fragmentation method was

determined for each peptide. As discussed in Chapter 1, CAD prefers short peptides

(less than 15 residues) with low charge state while ETD provides efficient

fragmentation of longer peptides (more than 15 residues) with a higher charge state

( $z \geq +3$ ). Peptide DITGLRLLAASVSSKGLPSPYNM contains 3 basic residues and

therefore both CAD and ETD can be utilized for PTM identification. Peptide

SSAPGSRSGSRSGSRSGSRSGSRSGSRSGSF contains 14 serines and was the most

interesting target for phosphorylation identification. The presence of seven

arginine residues allows this peptide to carry as many as eight charges and

therefore it requires ETD to generate informative MS2 spectra. The C-terminal

peptide DATGNSSYSYSYSFSSSIGH does not contain additional basic residues

besides the N-terminus; therefore its CAD spectra could provide structural

information.

---



Also it was found that increasing the water concentration had negative effect on the cleavage yield. The cleavage yield decreased by 50% in 30% TFA compared to 70% TFA, mainly due to the decreased acidity.

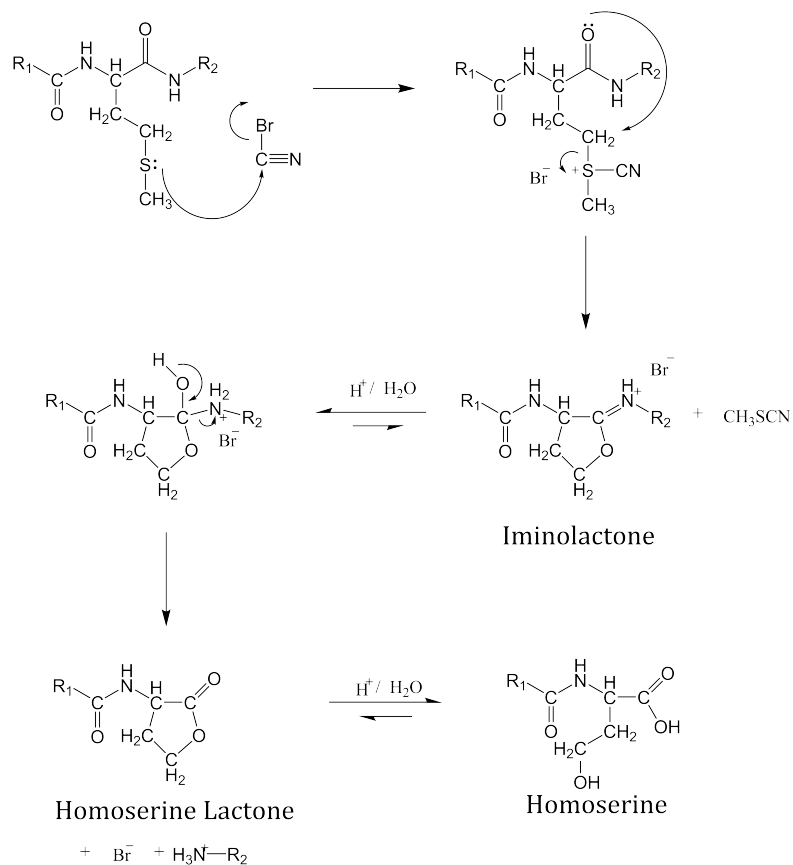
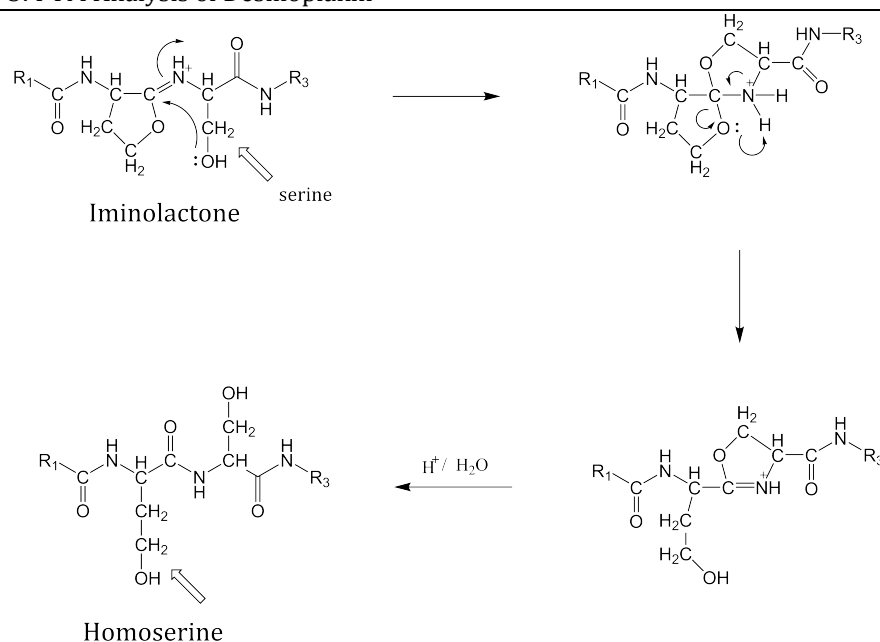


Figure 3. 9 Scheme I for CNBr cleavage of Met-X bonds where X is any amino acid.



**Figure 3. 10** Scheme II for CNBr cleavage of Met-X bonds where X is a serine or threonine.

An AspN digestion of the DP-S-tag sample was performed to evaluate sample quality. The digestion was followed by a 90-minute HPLC gradient separation and MS/MS analysis. Most identified peptides were DP-S-tag peptides; however, some peptides were not consistent with the specificity of AspN. Further database searches revealed the presence of non-specifically cleaved peptides. **Table 3.1** shows a list of peptides with cleavages in a non-specific manner detected by MS/MS at the DP-S-tag C-terminus. The theoretical peptide generated by AspN, DITGLRLLEAASVSSKGLPSPYNMSSAPGSRSGSRSGSRSGSRSGSRSGSRRGSF, was not detected. Besides the DP-S-tag C-terminus, non-specific peptides were also found in other regions of DP-S-tag.

**Table 3. 1 Non-specific peptides detected by MS/MS at DP-S-tag C-terminus after AspN digestion.**

Lower case “m” indicates the Met is oxidized.

DITGLRL
DITGLRLL
DITGLRLLE
DITGLRLLEA
DITGLRLLEAA
DITGLRLLEAASVSSK
DITGLRLLEAASVSSKGLPSPY
DITGLRLLEAASVSSKGLPSPYN
DITGLRLLEAASVSSKGLPSPYNm

Several experiments were performed to investigate the reason for the observed non-specific cleavages. Firstly, the specificity of AspN was examined by doing the same digestion and LC-MS/MS experiment on an immunopurified dynein sample (Chapter 2) and DP-S-tag sample. Both datasets were subjected to a no-enzyme OMSSA search against the dynein database and the DP-S-tag database, respectively. Significant numbers of non-specific cleavages were observed in DP-S-tag sample while no such cleavages were found in the dynein sample. Additionally, the LysC digestion on DP-S-tag showed significant non-specific cleavages as well. These results showed that the non-specific cleavages were not caused by AspN itself. It was suspected that some proteases from the HEK 293 cells were co-purified with DP-S-tag and caused the non-specific cleavages. To prove this hypothesis, purified DP-S-tag was washed off the agarose beads and SDS-PAGE was performed to separate all the proteins based on molecular weight. Then the gel band containing DP-S-tag (32kD) was cut out and subject to in-gel trypsin digestion. Not

surprisingly, LC-MS/MS analysis of the in-gel digested sample revealed highly specific tryptic peptides, proving that the removal of other proteins can prevent non-specific cleavages of the DP-S-tag. Also, overnight incubation of the DP-S-tag sample at 37°C without the addition of any protease resulted in DP-S-tag peptides, providing more evidence of the presence of proteases that were co-purified with DP-S-tag.

To minimize the contaminant proteins in the sample, the ratio of cell lysate and S-protein agarose beads was adjusted to reduce the binding space on the agarose beads for non-DP-S-tag proteins. In addition, the salt concentration in the wash buffers was increased and more wash steps were performed in the purification process to remove weakly bound non-specific proteins. Also, to prevent the proteolysis in the purification process, protease inhibitors were added to all of the wash buffers. An evaluation of these efforts using ApsN digestion and LC-MS/MS analysis showed that the amount of DP-S-tag peptides with specific cleavages improved by threefold. However, the non-specific cleavage issue still existed.

To further prevent proteolysis by the proteases that were already present in the sample, the CNBr reaction was performed before AspN digestion. Since the CNBr reaction was carried out in 70% TFA, which can denature most proteins, the proteases might lose their activity in this highly acidic condition. In addition, the CNBr reaction was performed at 4°C to limit any enzymatic activity. LC-MS/MS

---

analysis of peptides resulting from CNBr reaction still showed non-specific

cleavages but in lower amounts (**Table 3.2**). These cleavages could have occurred during sample storage or even during the CNBr reaction.

**Table 3. 2 Selected non-specific peptides detected by LC-MS/MS after CNBr reaction.**

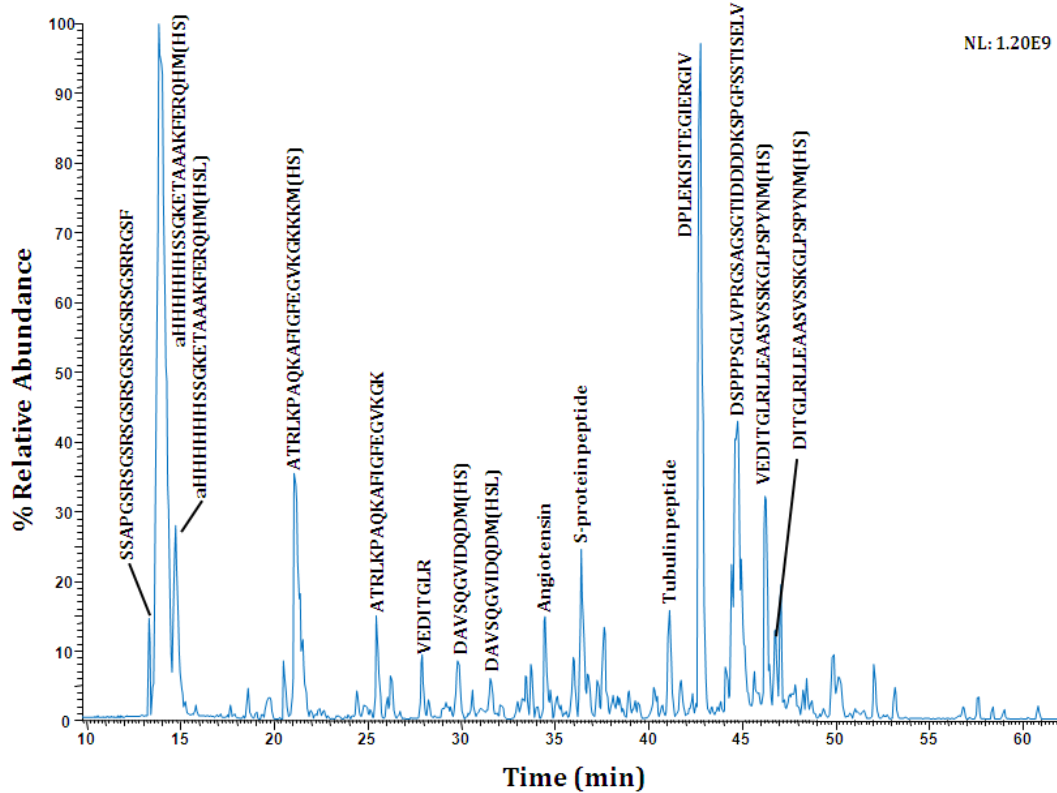
The theoretical peptides generated by CNBr were VEDITGLRLLEAASVSSKGLPSPYNM and ATRLKPAQKAFIGFEGVKGKKM, where the Met residues have been converted to homoserine lactones or homoserines, were detected by LC-MS/MS.

VEDITGLR
VEDITGLRL
VEDITGLRLL
VEDITGLRLLEAASVSSK
VEDITGLRLLEAASVSSKGLPSPYN
ATRLKPAQKAFIG
ATRLKPAQKAFIGFEGVK
ATRLKPAQKAFIGFEGVKG
ATRLKPAQKAFIGFEGVKGK
ATRLKPAQKAFIGFEGVKGKK
ATRLKPAQKAFIGFEGVKGKKK

After AspN digestion of the CNBr-treated sample, the resulting peptides were subjected to LC-MS/MS analysis. **Figure 3.10** shows the base peak chromatogram of the DP-S-tag sample. Most peaks were identified as DP-S-tag peptides. Some contaminants were identified as peptides from the S-protein that was used in DP-S-tag purification and others were from proteins such as tubulin. **Table 3.3** provides the relative quantification of the peptides generated with CNBr and AspN cleavage. While the N-terminal peptide presented at 3 pmol level, most other peptides were below 1 pmol. This difference could be due to non-specific cleavages as well as differences in cleavage, separation and ionization efficiency. As expected, peptide



SSAPGSRSGSRSGSRSGSRSGSRSGSRRGSF was seen in reduced yield due to the lower efficiency of the CNBr cleavage of Met-Ser. Additionally, this peptide is extremely hydrophilic due to the repeating serines and arginines; therefore it elutes very early in the HPLC gradient (**Figure 3.11**). Its poor binding affinity to the HPLC column resulting from high hydrophilicity could have resulted in its loss during HPLC column loading. Peptide DATGNSSYSYSYSFSSSSIGH was the least abundant peptide in this list, possibly due to non-specific cleavage and its low ionization efficiency.



**Figure 3. 11 Base peak chromatogram resulting from LC-MS/MS analysis of CNBr and AspN cleavage of affinity purified DP-S-tag.**

The sequences of the major peaks are labeled within the chromatogram. NL: normalized ion count.

**Table 3. 3 Relative quantification of DP-S-tag peptides with CNBr and AspN cleavage.**

In this experiment 1 pmol (based on in-gel estimation) of DP-S-tag peptides were loaded on to HPLC column. Peptide 1 is the N-terminal peptide where the original N-terminus Met was cleaved in cells and the new N-terminus Ala was found to be acetylated (indicated with lower case “a”). Peptides 1-4 contain methionines in the C-terminus which are from CNBr cleavage and converted to homoserine lactones or homoserines. Peptide 3 is a mis-cleaved product, containing an AspN inside the sequence. The amount (fmol) of each peptide is determined using internal standard peptides as described in Chapter 2.

No.	Peptide	Amount (fmol)
1	aHHHHHHSSGKETAAAKFERQHM	3469
2	ATRLKPAQKAFIGFEGVKGKKKM	1643
3	VEDITGLRLLEAASVSSKGLPSPYNM	745
4	DITGLRLLEAASVSSKGLPSPYNM	353
5	SSAPGSRSGSRSGSRSGSRSGSRSGSRRGSF	542
6	DATGNSSYSYSYSFSSSSIGH	104

### 3.5.3 PTM Analyses of DP-S-tag

#### 3.5.3.1 Phosphorylation

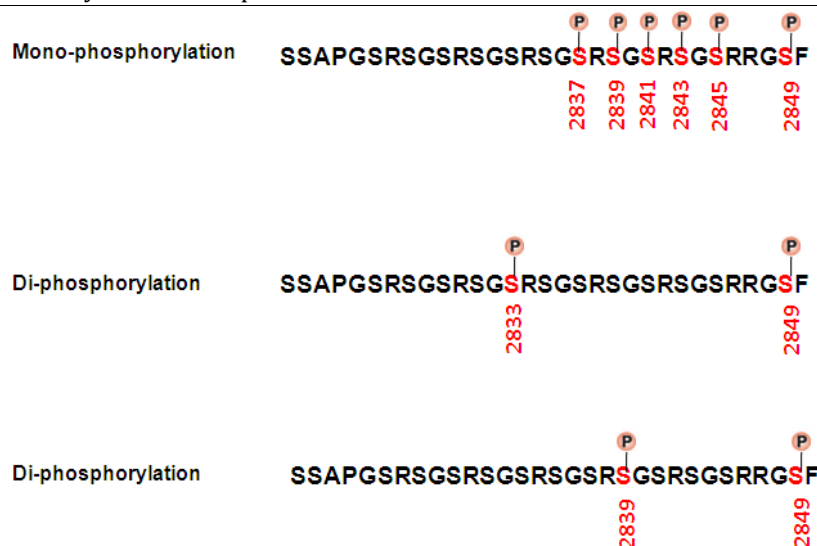
The mono-phosphorylated form of peptide

SSAPGSRSGSRSGSRSGSRSGSRSGSRRGSF was identified as 10% of the total amount of this peptide. Due to the low amount of phosphorylated peptide, the S/N of the ETD spectra was low and therefore the localization of the phosphorylation site was difficult. Also, there was evidence that the ETD spectra resulted from a mixture of precursors with phosphorylation on several different sites. In addition, since the charge state of the chosen precursor was 6, the resulting fragment ions possessed charge states  $z=1-4$ . Low resolution ETD spectra do not provide charge state information for fragment ions with  $z \geq 2$ ; therefore the confident assignment of fragment ions became difficult.

The configuration of front-end ETD enables the ions resulting from multiple rounds of ETD to be accumulated in the C-trap prior to detection in the Orbitrap mass analyzer (35). In this case, the S/N increases linearly with the number of C-trap fills. A targeted experiment was performed on the ion with charge state of +6 and average  $m/z$  of 521.86 using ten C-trap fills and high resolution mass analysis in the Orbitrap. The ETD fragment signal was significantly enhanced after 10 rounds of fragment ion accumulation. High resolution MS2 spectra provided charge state information for the fragment ions, facilitating the assignment of fragment ions. These improvements allowed for the identification of six mono-phosphorylation sites on this peptide, Ser2837, Ser2839, Ser2841, Ser2843, Ser2845 and Ser2849 as shown in **Figure 3.12**. The identified phosphorylation sites fall within the consensus sites for two separate putative GSK3 phosphorylation cascades (S/T XXX S/T).

Additionally, di-phosphorylation was also found on this peptide at 1% level. It had two phosphorylated forms, one form with phosphorylations on Ser2833 and Ser2849, and the other form with phosphorylation on Ser2839 and Ser2849 (**Figure 3.12**). Among the seven identified phosphorylation sites, only Ser2849 was reported before.

---



**Figure 3. 12 Mono-phosphorylation and di-phosphorylation sites identified on DP-S-tag peptide SSAPGSRSGSRSGSRSGSRSGSRSGSRSGSRRGSF.**

### 3.5.3.2 Methylation

Peptide SSAPGSRSGSRSGSRSGSRSGSRSGSRSGSRRGSF was also found to contain arginine methylation. Mono-, di-, tri-, and tetra-, penta-, hexa-methylation were identified on arginines within this peptide. **Figure 3.13** shows the XIC of all the methylated forms; only the most abundant charge state is shown for each form. Besides quantitation using peak area as described in chapter 2, the peak height can also be utilized for relative semi-quantitative comparison between different species. For example, in **Figure 3.13**, the normalized ion count (NL) of mono-methylated form is 3.55E7 and the NL of tri-methylated form is 1.24E7; therefore, the amount of mono-methylated form is roughly threefold of the tri-methylated form. Front-end ETD combined with multiple C-trap fills enabled the localization of the methylation sites shown in **Table 3.5**. **Figure 3.14** shows the ETD spectrum of di-methylated

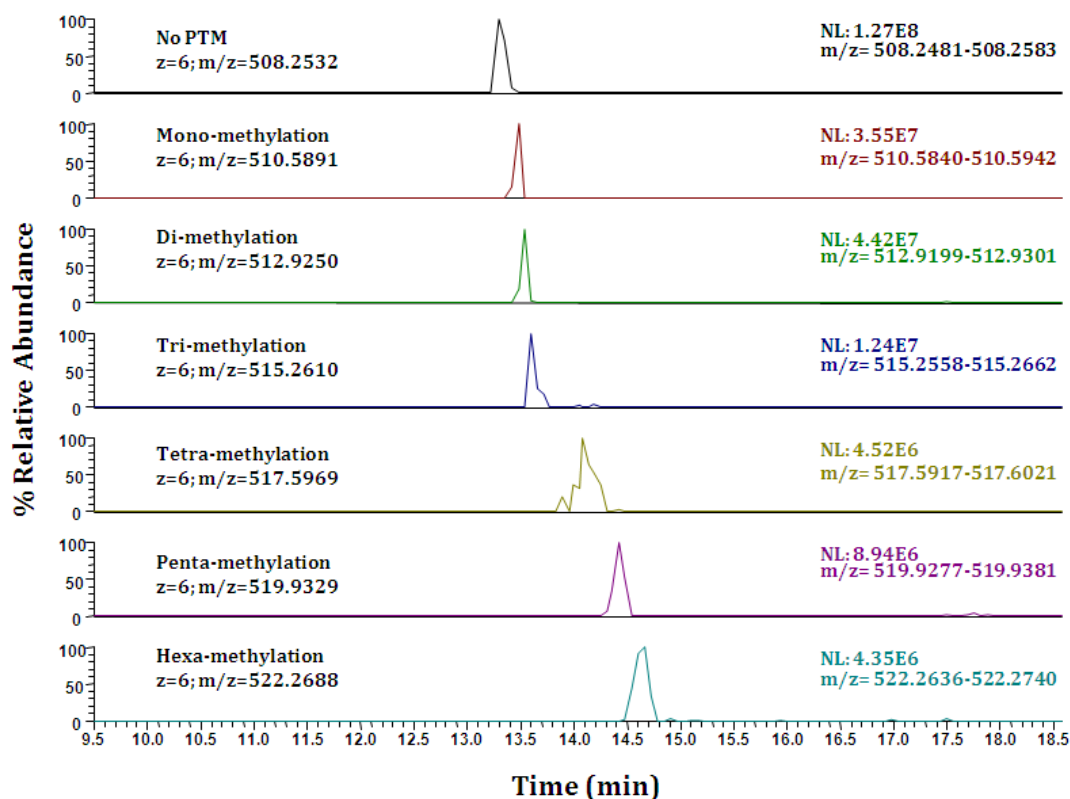
SSAPGSRSGSRSGSRSGSRSGSRSGSRSGSF as an example of the identification of methylation using ETD.

Protein methylation, another common PTM in biological systems, is the covalent attachment of a methyl group from S-adenosyl methionine (SAM) to the nitrogen in either the amino group on the N-terminus, or the Lys/Arg side chain (36). Protein methylation is catalyzed by methyltransferases and it has been estimated that 1% of all genes within the mammalian genome encode for methyltransferases (36). Methylation occurs in three different ways on the guanidino group: monomethylated, symmetrically di-methylated and asymmetrically di-methylated, each of which has potentially different functional consequences (37). Methylation can dramatically change protein character by increasing hydrophobicity and steric bulk (36). This can consequently inhibit or enhance protein-protein interactions through specific recognition of the methylated residue.

The structures of symmetrical di-methylation and asymmetrical di-methylation are shown in **Figure 3.15**. In ETD, electrons can be transferred to the guanidino groups, resulting in the loss of small neutral molecules from the guanidine groups. Differently methylated guanidino groups produce different small neutral molecules; therefore, small neutral molecule loss analysis of the ETD spectra could provide structural information regarding the methylated arginines (**Figure 3.15**). Analysis of high-resolution ETD spectra of methylated peptide

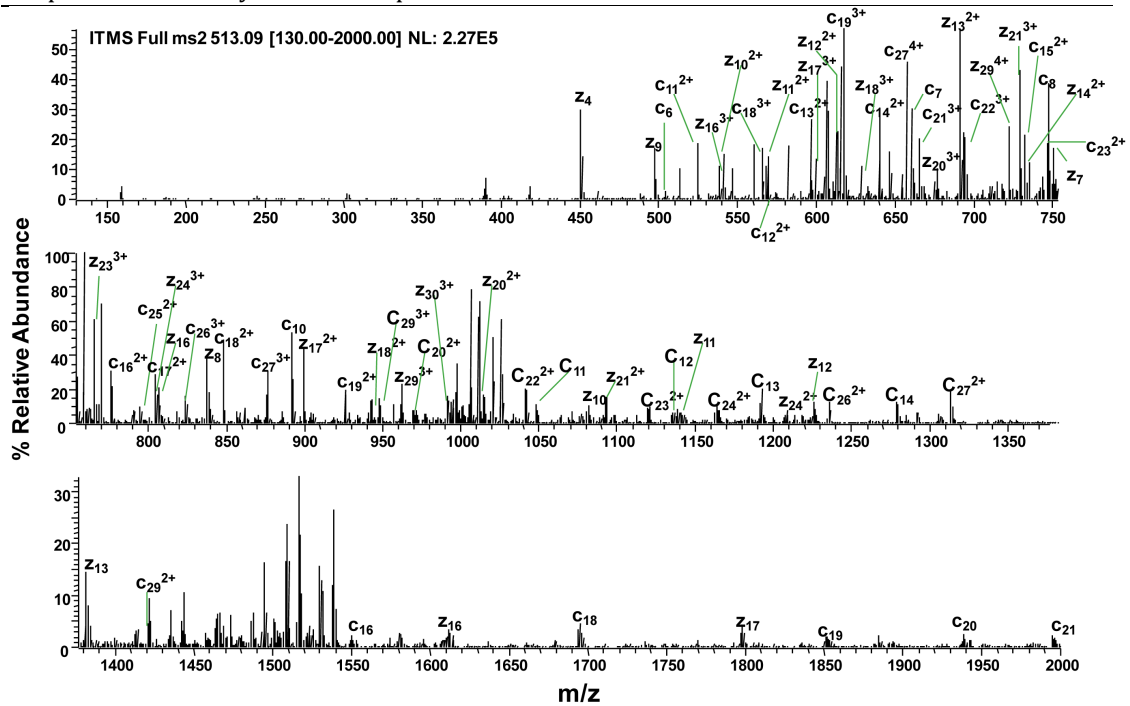
---

SSAPGSRSGSRSGSRSGSRSGSRSGSRRGSF suggested that the methylation is more possibly asymmetric (**Figure 3.16**). Pharmacological studies will be performed in the future to further confirm this conclusion.



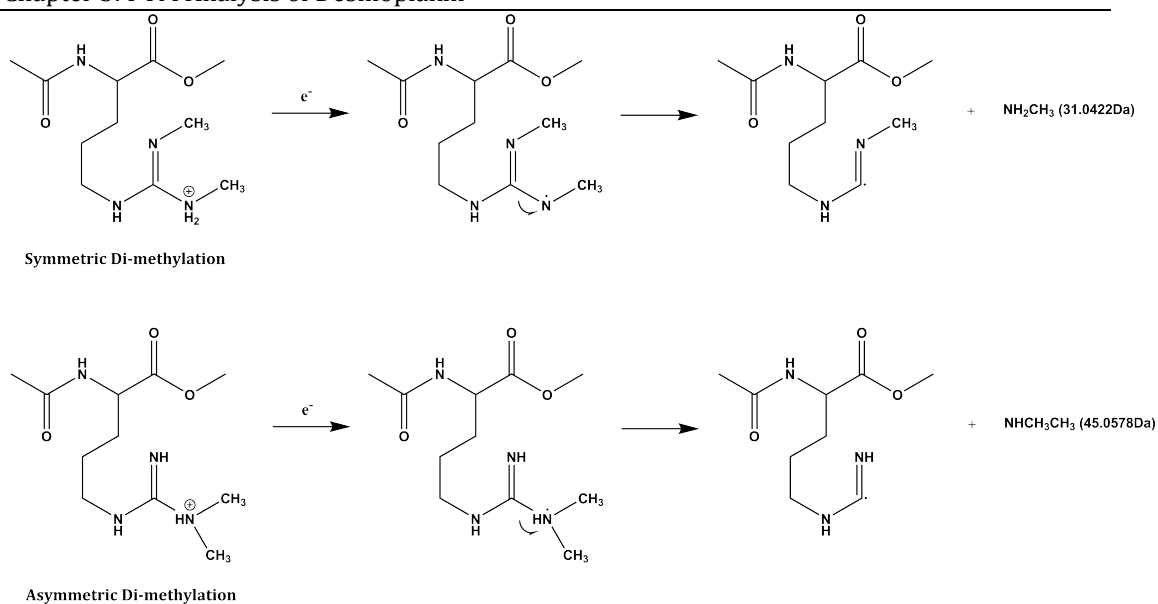
**Figure 3. 13** XICs of 1-6 methylation detected on DP-S-tag peptide SSAPGSRSGSRSGSRSGSRSGSRRGSF.

Only the most abundant charge state is shown for each peptide. The charge state and experimental monoisotopic m/z are labeled. NL: normalized ion counts.



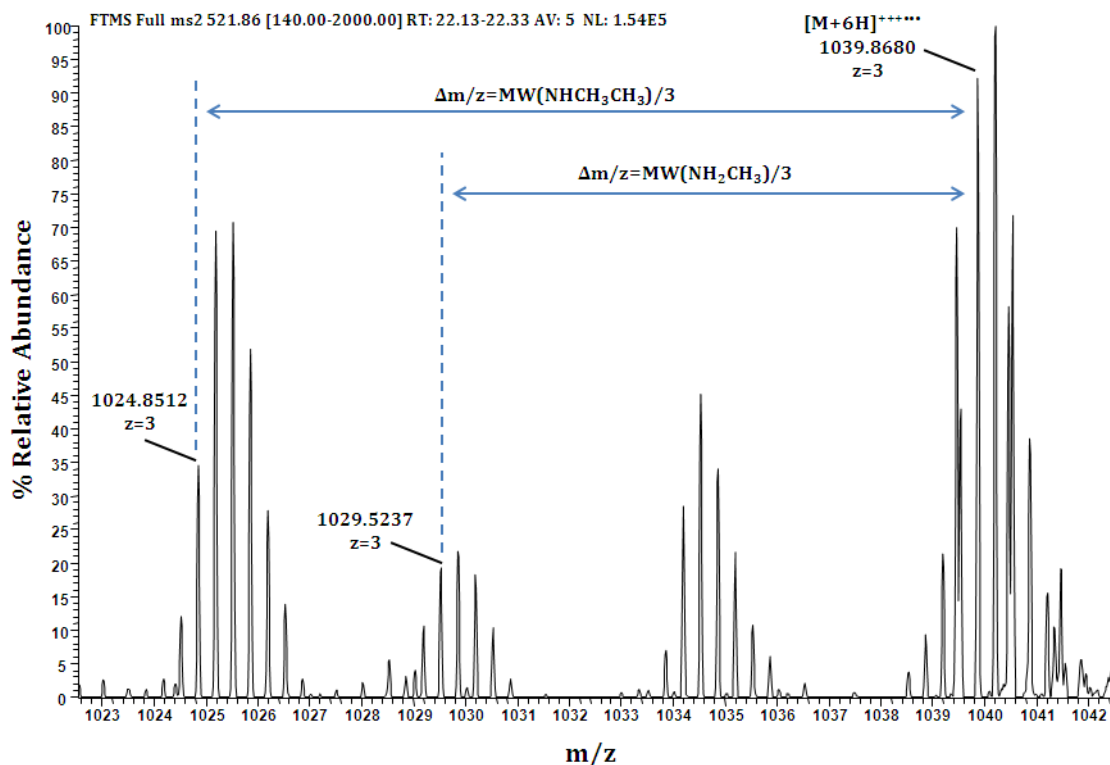
**Figure 3. 14 ETD spectrum of di-methylated PD-S-tag peptide SSAPGSRSGSRSGSRSGSRSGSRSGSRSGF.**

A precursor with charge state of +6 and m/z of 513.09 was chosen for ETD fragmentation. Fragment ions detected are labeled within the spectrum and in the peptide sequence. The di-methylated arginine is indicated by "me" in the front. RT: retention time, AV: number of spectra averaged, NL: normalized ion counts.



**Figure 3. 15 The mechanism of loss of small neutral molecules from symmetrically and asymmetrically methylated peptides in ETD.**





**Figure 3. 16 Neutral loss analysis of ETD spectrum of penta-methylated PD-S-tag peptide SSAPGSRSGSRSGSRSGSRSGSRSGSRSGSF.**

A precursor with  $z = +6$  and  $m/z$  of 521.86 is chosen for ETD fragmentation. The  $m/z$  analysis is performed in Orbitrap to generate high-resolution ETD spectrum that can resolve isotopic peaks for each fragment ions. The region of  $m/z$  1022-1043 is zoomed in to show the loss of small neutral molecules. Peak with  $m/z$  1039.8680 corresponds to the charge reduced species of the precursor and its charge state is +3. The mass different between peak with  $m/z$  of 1039.8680 and peak with  $m/z$  of 1029.5237 corresponds to the mass of  $NH_2CH_3$  (31.0422 Da). The mass different between peak with  $m/z$  of 1039.8680 and peak with  $m/z$  of 1024.8512 corresponds to the mass of  $NHCH_3CH_3$  (45.0578 Da). If the methylation is symmetric, the peak with  $m/z$  of 1029.5237 would be the only peak appeared. If the methylation is asymmetric, the peaks with  $m/z$  1029.5237 and 1024.8512 would appear with the ratio approximately 1:2, which is close to the ratio showed in this spectrum. It is also possible that symmetric and asymmetric methylation occur simultaneously. However, similar analysis of hexa-methylated SSAPGSRSGSRSGSRSGSRSGSRSGSRSGSF also supports the conclusion that the methylation is more possibly asymmetric.

In addition, the combination of phosphorylation and methylation was also identified on this peptide as shown in **Figure 3.17**. Relative quantitation showed that peptides with both modifications are present at the 0.1%-1% level, corresponding to hundreds of attomoles to a few fmols in this experiment. Only the two relatively abundant forms “di-methylation + mono-phosphorylation” and “tri-methylation + mono-phosphorylation” had ETD data to localize the modified sites as shown in **Table 3.5**. Collectively, seven phosphorylation sites and four methylation sites were identified on this peptide (**Figure 3.18**). Except Ser2849, all the PTMs identified are novel sites.

---

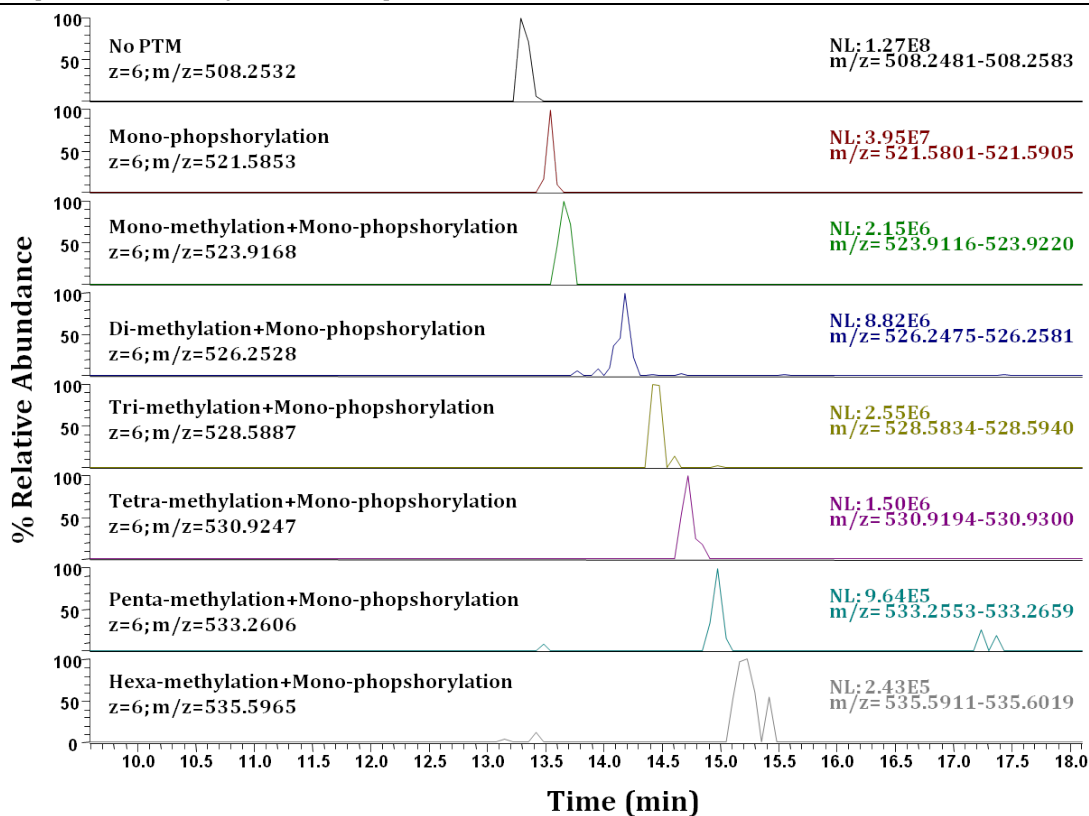


Figure 3. 17 XICs of 1-6 methylation and phosphorylation detected on DP-S-tag peptide SSAPGSRSGSRSGSRSGSRSGSRSGSRRGSF.

Only the most abundant charge state is shown for each peptide. The charge state and experimental monoisotopic m/z are labeled. NL: normalized ion counts.



Figure 3. 18 All the phosphorylation and methylation sites identified on DP-S-tag peptide SSAPGSRSGSRSGSRSGSRSGSRSGSRRGSF.

Amino acids labeled with "P" are the phosphorylation sites and amino acids labeled with "Me" are the methylation sites. The sequence number of each site in DP is also labeled.

**Table 3. 4 A summary of methylated DP-S-tag peptide SSAPGSRSGSRSGSRSGSRSGSRRGSF.**

The location of methylation sites are indicated with “me” (modified with single methyl group) or “me,me” (modified with two methyl groups). “Phospho” indicates a phosphorylation site.

Mono-methylation	PTM Sites
SSAPGSR(me)SGSRSGSRSGSRSGSRRGSF	R2826
SSAPGSRSGSRSGSR(me)SGSRSGSRSGSRRGSF	R2834
SSAPGSRSGSRSGSRSGSR(me)SGSRSGSRRGSF	R2838
SSAPGSRSGSRSGSRSGSRSGSRSGSR(me)RGSF	R2846
Di-methylation	
SSAPGSR(me,me)SGSRSGSRSGSRSGSRRGSF	R2826
SSAPGSRSGSRSGSR(me,me)SGSRSGSRSGSRRGSF	R2834
SSAPGSRSGSRSGSRSGSR(me,me)SGSRSGSRRGSF	R2838
SSAPGSRSGSRSGSRSGSRSGSRSGSR(me,me)RGSF	R2846
Tri-methylation	A different combinations of R2826, R2834, R2838 and R2846
Tetra-methylation	
Penta-methylation	
SSAPGSR(me,me)SGSRSGSR(me)SGSRSGSRSGSR(me,me)RGSF	R2826, R2834, and R2846
SSAPGSR(me,me)SGSRSGSRSGSR(me)SGSRSGSR(me,me)RGSF	R2826, R2838, and R2846
Hexa-methylation	
SSAPGSR(me,me)SGSRSGSR(me,me)SGSRSGSRSGSR(me,me)RGSF	R2826, R2834, and R2846
SSAPGSR(me,me)SGSRSGSRSGSR(me,me)SGSRSGSR(me,me)RGSF	R2826, R2838, and R2846
Mono-phosphorylation+Di-methylation	
SSAPGSRSGSRSGSR(me,me)SGSRSGS(phospho)RSGSRRGSF	R2834 and S2841
SSAPGSRSGSRSGSR(me,me)SGSRSGSRS(phospho)GSRRGSF	R2834 and S2843
Mono-phosphorylation+Tri-methylation	
SSAPGSR(me)SGSRSGSR(me,me)SGSRSGS(phospho)RSGSRRGSF	R2826, R2834 and S2841
SSAPGSRSGSRSGSR(me,me)SGSR(me)SGSRS(phospho)GSRRGSF	R2834, R2838 and S2843

### 3.5.4 PTM Analyses of DP-S-tag S2849G

As discussed previously, *in vitro* binding assays and immunofluorescence co-localization studies show that inhibiting phosphorylation at residue S2849 in the DP C-terminus through site-directed mutagenesis (DP<sup>S2849G</sup>), leads to a 10-fold increase in DP-IF associations (12). Additionally, live cell imaging of the phospho-deficient DP mutant depicts delayed trafficking to the desmosome at new cell-cell contacts. To investigate the influence of S2849 phosphorylation on phosphorylation events of other serines, LC-MS/MS analysis was performed on DP-S-tag S2849G. Surprisingly, no phosphorylated form was identified for peptide

SSAPGSRSGSRSGSRSGSRSGSRSGSRSGSF while methylations were still found in mono-, di-, tri-, tetra-, penta-, and hexa-forms. Therefore, S2849 phosphorylation could be a primer for the following phosphorylation cascade and inhibition of S2849 phosphorylation blocks the sequential phosphorylation events. However, arginine methylation could be independent of S2849 phosphorylation.

### 3.5.5 PTM Analyses of DP-S-tag R2834H

The PTM pattern of DP-S-tag R2834H was also investigated since R2834H has been identified in patients with ARVC and heart-specific expression of DPR2834H *in vivo* leads to cardiomyocyte apoptosis, cardiac fibrosis, and cardiac dysfunction in adult mice (29). The SSAPGSRSGSRSGSRSGSRSGSRSGSRSGSF peptide was detected at an extremely low level in this sample, possibly due to the

---

non-specific digestion issue. While about 1 pmol sample was subject to LC-MS/MS analysis, only 40 fmol of unmodified peptide was detected. Mono-phosphorylated peptide was detected but at 1 fmol level. Nonetheless, the multiple C-trap fills enabled by front-end ETD successfully enhanced ETD fragment signal and allowed for the localization of phosphorylation site only on Ser2849. This result provided evidence that R2834 methylation inhibits the phosphorylation events beside S2849 phosphorylation. In addition, mono-, di-, and tri-methylation were found on this peptide.

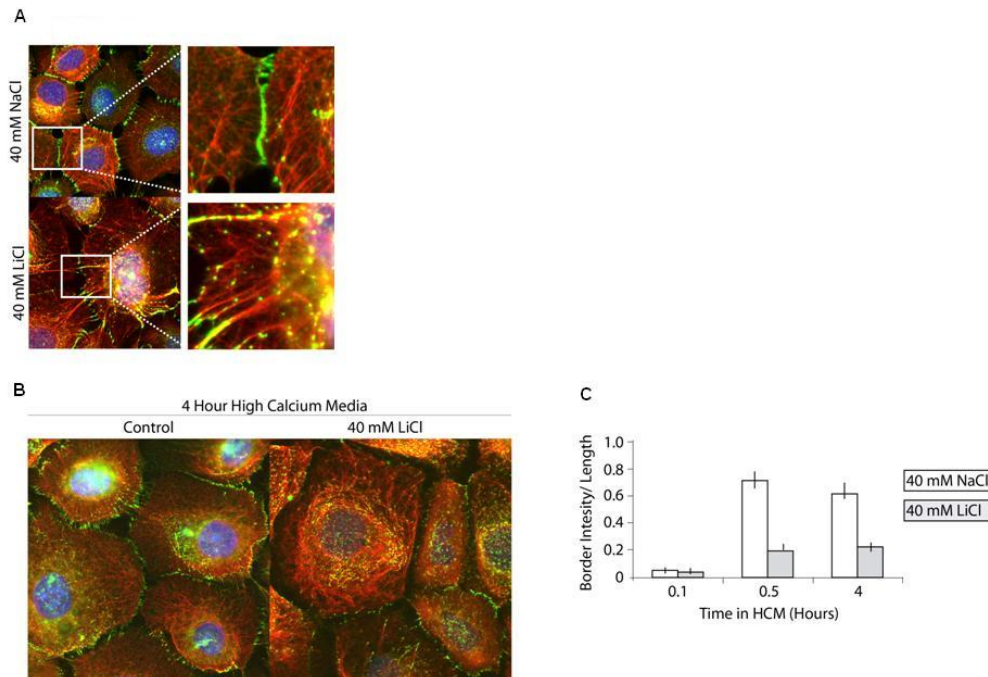
---

### 3.5.6 GSK3 Signaling Promotes Desmosome Assembly\*

Within the C-terminal tail, DP contains 15 consensus sites for GSK3 phosphorylation cascades (S/T XXX S/T) (**Figure 3.4**). GSK3 is the primary kinase responsible for promoting processive phosphorylation cascades, a mechanism that increases the rate of multi-site phosphorylation and has precedence where downstream signaling depends on fully phosphorylated forms of substrate (25). Given the importance of GSK3 as a diverse signaling mediator, pharmacological approaches were applied to inhibit GSK3 phosphorylation to gain initial insights into the role of GSK3 signaling on DP. Co-localization immunofluorescence microscopy indicated that GSK3 inhibition promoted enhanced localization of DP along IF, suggesting that GSK3 signaling affects DP interactions with IF (**Figure 3.19A**). To examine the role of GSK3 signaling on desmosome assembly, calcium switch analysis in combination with GSK3 inhibition was performed. GSK3 inhibition delayed DP trafficking to new sites of cell-cell contact and delayed desmosome formation compared to control cells upon incubation in high calcium media (**Figure 3.19B, C**). These data establish a role for GSK3 signaling in mediating DP-IF interactions to promote desmosome assembly.

---

\*The following discussions are based on the experiments performed in the Green Laboratory.



**Figure 3.19 GSK3 signaling promotes desmosome assembly.**

A. Co-localization immunofluorescence microscopy of the association of DP and IF under normal condition and GSK3 inhibition condition. In SCC9 cells, GSK3 activity was inhibited with 40 mM LiCl while the experimental control was treated with 40 mM NaCl. Green: DP; red: IF. B. Co-localization immunofluorescence microscopy combining calcium switch analysis and GSK3 inhibition. Cells were first incubated in a low calcium media that disrupts intercellular junctions. Upon incubation in high calcium media, DP was trafficking to new sites of cell-cell contact. C. Quantitation of cell border intensity/cell length in different time points of the calcium switch analysis under normal condition and GSK3 inhibition condition. HCM: high calcium media.

### 3.5.7 GSK3 Is Recruited to DP to Modulate DP-IF Complexes

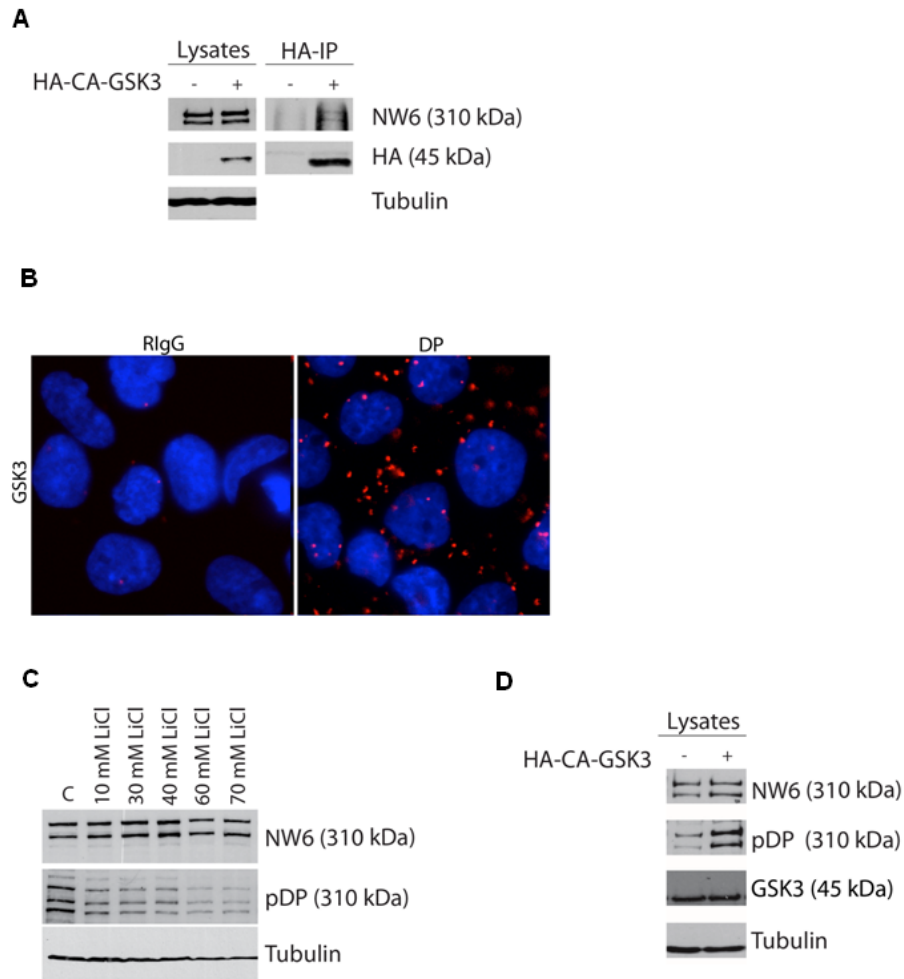
The binding interaction between DP and GSK3 was biochemically characterized using co-immunoprecipitation assay and proximity ligation assay analysis. As shown in **Figure 3.20A**, when HA-tagged constitutively active GSK3 (HA-CA-GSK3) was expressed in cells, endogenous DP was detected from the immunopurified HA-CA-GSK3, which demonstrated the direct interactions between GSK3 and endogenous DP. To substantiate the direct interaction between DP and



GSK3 in cells, proximity ligation assay analysis was used to provide visualization

and localization of individual protein-protein interactions based on the detection of coupling between secondary antibodies that are bound to target proteins. In **Figure 3.20B**, the detection of fluorescent dots (red) indicates the presence and localization of DP interactions with GSK3 that are absent in rabbit IgG control images. In addition, the pharmacological inhibition of GSK3 was sufficient to decrease phosphorylation of endogenous DP in a dose-dependent manner (**Figure 3.20C**) while expression of CA-GSK3 enhanced phosphorylation (**Figure 3.20D**). These data introduce a novel role for GSK3 in the signaling machinery recruited to DP-IF complexes to modulate cytoskeletal-membrane attachments during desmosome assembly.

---



**Figure 3. 20 GSK3 is recruited to DP to modulate DP-IF complexes**

A. Blots of lysates and immunopurified HA-CA-GSK3 from HEK 293 cells expressing HA-CA-GSK3 probed with antibodies to DP (NW6), HA-tag (HA), and tubulin. Tubulin was used as the loading control. B. Proximity ligation analysis (PLA) of in situ association of GSK3 and endogenous DP. PLA was performed using primary anti-GSK3 antibodies along with control antibody (rabbit IgG, RlgG) or DP antibody, in conjunction with secondary antibodies (PLA probes) attached with unique short DNA strands. When the PLA probes were in close proximity, the DNA strands can interact through a subsequent addition of two other circle-forming DNA oligonucleotides. After joining of the two added oligonucleotides by enzymatic ligation, they are amplified and viewed with a fluorescence microscope. Red: fluorescent label to show the interactions of GSK3 and DP. Blue: DAPI fluorescent label that binds to nucleus for normalization. C. Blots of HEK 293 cell lysates with pharmacological inhibition of GSK3 using different amounts of LiCl probed with antibodies to DP (NW6), phosphor-DP (pDP), and tubulin. c: control. D. Blots lysates of HEK 293 cells expressing HA-CA-GSK3 probed with antibodies to DP (NW6), phosphor-DP (pDP), GSK3 and tubulin.

### 3.5.8 GSK3 Promotes Phosphorylation Cascades of the DP C-terminal Tail

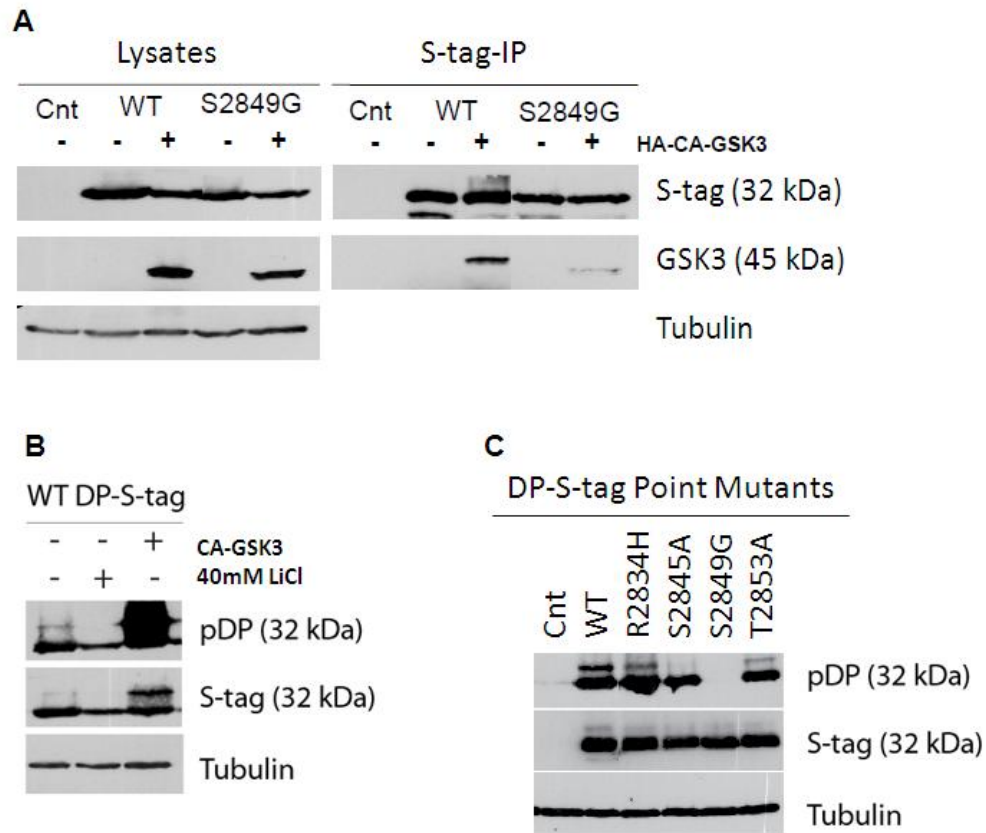
To test if GSK3 binding affinity for DP was conserved with the DP C-terminal truncation construct, co-immunoprecipitation analysis was performed on DP-S-tag. Consistent with endogenous DP, WT DP-S-tag specifically bound to HA-CA-GSK3 (**Figure 3.21A**). To ensure that enzymatic activity of GSK3 was also retained, pharmacological and genetic manipulation of GSK3 were performed in the transfected HEK 293 cells. Using the DP-specific phospho-antibody, inhibition of GSK3 was found to ablate the presence of an upper band compared to control cells where WT DP-S-tag is visualized as a doublet (**Figure 3.21B**). Additionally, co-expression of WT DP-S-tag and CA-GSK3 enhanced the presence of this upper phospho-band (**Figure 3.21B**). Using site-directed mutagenesis to further probe the presence of a GSK3 cascade, it was found that blocking phosphorylation with S2845A also resulted in a loss of the upper phospho-band recognized by the DP-specific phospho-antibody while T2853A had no effect indicating S2845 lies within the cascade but T2853 may belong to another cascade (**Figure 3.21C**). These data suggest that phosphorylation of S2849 is partially mediated by GSK3 where the upper phospho-band represents a GSK3-sensitive phospho-state of DP. Intriguingly, inhibiting phosphorylation of S2849 with site-directed mutagenesis was sufficient to weaken DP binding to GSK3 by co-immunoprecipitation analysis (**Figure 3.21A**). As discussed previously, LC-MS/MS analyses of DP-S-tag S2849G demonstrated a complete loss of phosphorylation at GSK3-mediated phospho-sites identified in this

---

study. Together, these data indicate that GSK3 promotes phosphorylation cascades of the DP C-terminal tail and phosphorylation of S2849 controls downstream phosphorylation cascades.

Collectively, GSK3 represents a novel binding partner for DP and phosphorylates the C-tail with a processive mechanism in two phosphorylation cascades. Additionally, these analyses provide further evidence that complex patterns of phosphorylation cascades fine-tune the regulatory mechanism of PTMs for the C-terminal tail. While it is tempting to assume S2849 acts as a priming site for GSK3 phosphorylation cascades, the sensitivity of this site itself to GSK3 enzymatic activities suggest a less traditional role for S2849 to control phosphorylation cascades.

---



**Figure 3. 21 GSK3 promotes phosphorylation cascades of the DP C-terminal tail.**

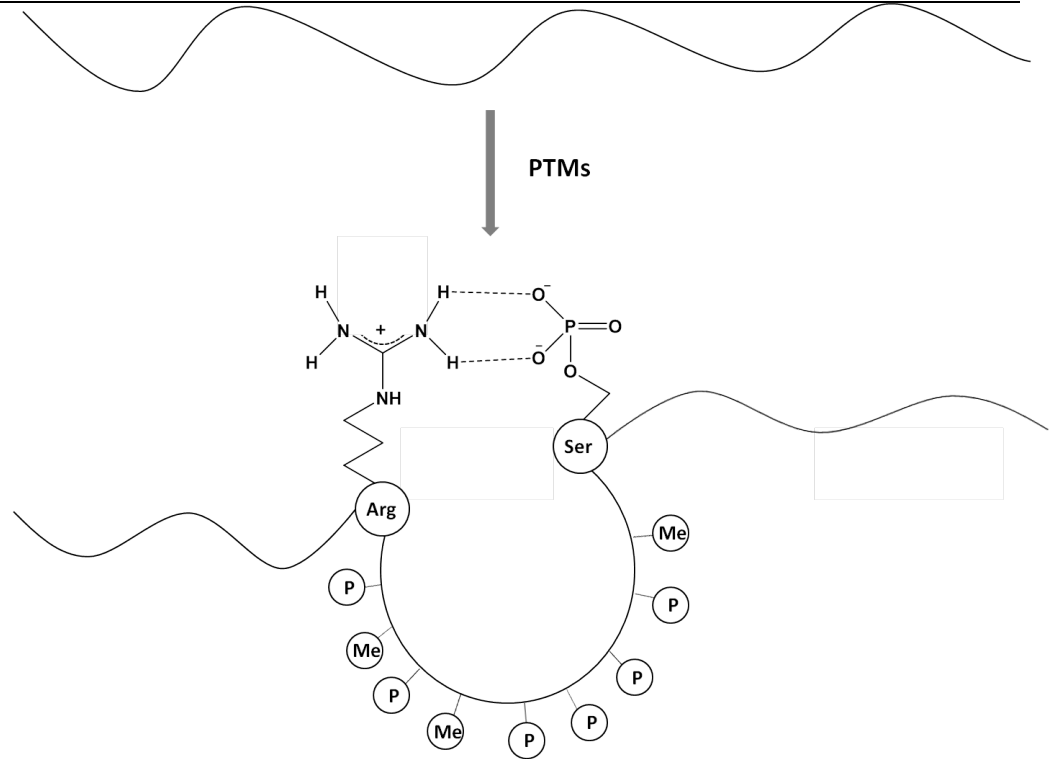
A. Blots of lysates and immunopurified DP-S-tag from HEK 293 cells expressing DP-S-tag in different conditions probed with antibodies to S-tag, GSK3, and tubulin. Tubulin was used as the loading control. Control cells did not express DP-S-tag, WT cells express WT DP-S-tag, and S2849G cells express DP-S-tag S2849G. The expression/no expression of HA-CA-GSK3 in WT cells and S2849G cells were also compared. B. Blots of immunopurified WT DP-S-tag from HEK 293 cells in different conditions probed with antibodies to phospho-DP (p-DP), S-tag, and tubulin. The cells were transfected with/without CA-GSK3 and treated with/without 40 mM LiCl. C. Blots of immunopurified DP-S-tag from HEK 293 cells expressing WT and point mutants of DP-S-tag probed with antibodies to phospho-DP (p-DP), S-tag, and tubulin.

### 3.5.9 Sequential Phosphorylation Induces Arginine Claw Conformation of the

#### DP C-terminal Tail

Phosphorylation has been shown to control protein function by inducing long-range alterations of protein structure such as the formation of arginine claws, a compact protein fold characterized by the hydrogen bond formation between phospho-moieties and guanidinium groups of arginine side-chains (**Figure 3.22**) (38). The identification of a phosphorylation cascade within the arginine-rich GSR-repeat domain led to the investigation of the presence of an arginine claw conformation in the C-terminal tail. Prior to phosphorylation, the C-tail is predicted to be in a flexible conformation that lacks tertiary structure in solution. Upon phosphorylation, the C-tail adopts a compact claw-like structure that possesses an enhanced thermostability (**Figure 3.22**). It is hypothesized that the arginine claw conformation of the C-tail serves as an electronegative docking groove that positions multisite phosphorylated serine residues outward to mediate interactions with other proteins. It is possible that the role of phosphorylation cascades of the C-tail is to produce a bulk electrostatic response to the negatively charged domain of IF proteins, loosening the association of DP and IF during desmosomal assembly.

---



**Figure 3. 22 Hypothesized phosphorylated-induced arginine claw conformation of the DP C-terminal tail confers the structure to this naturally unstructured region.**

### 3.5.10 Crosstalk Between Arginine Methylation and Phosphorylation

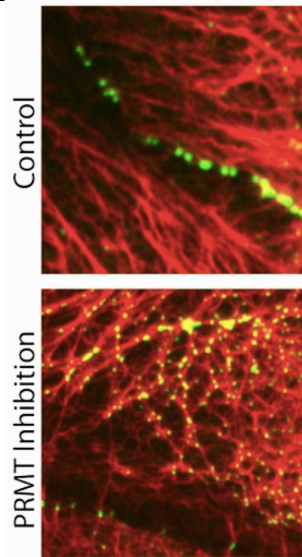
### Contributes to the Regulatory Roles of the C-terminal Tail on DP behavior

PTM crosstalk is that the presence of one modification influences the appearance of others (39). Positive crosstalk is a situation in which one PTM promotes the addition or removal of a second PTM, or enables recognition by a binding protein that carries out a second modification (40). Although PTMs are often thought of as acting independently, there are examples in which two or more PTM sites in a protein have a combinatorial effect on activity (11, 41). However, both the basis and prevalence of PTM cross-talk remain largely unknown. The DP C-

terminal tail represents a 'hot spot' with short clusters of modifiable residues. In addition to the large fraction of residues available for phosphorylation, the C-tail also contains consensus sites for arginine methylation (RG), a modification recently implicated as a prominent regulatory feature in a variety of cellular processes which acts by altering protein-protein interactions (42). LC-MS/MS with ETD revealed for the first time four DP arginine residues that are modified by mono- or di-methylation. The identification of peptides that contain methylated arginine residues and GSK3 phosphorylated residues simultaneously suggests that these modifications have synergistic regulatory roles. To address the possible importance of methylation on DP behavior, a pharmacological approach was used to inhibit PRMTs to block protein methylation. Global inhibition of protein methylation enhanced DP association with IFs and delayed desmosome assembly in a manner to reminiscent of kinase-inhibited cells (**Figure 13.9A, Figure 3.23**). This result suggests that arginine methylation contributes to the regulatory roles of the C-tail on DP behavior. As discussed previously, LC-MS/MS with ETD revealed the loss of phosphorylation sites besides S2849 in the methylation-deficient point mutant DP-S-tag R2834H, showing that the methylated R2834 is critical for the GSK3 phosphorylation cascade. Collectively, these data provide evidence for a model where phosphorylation and methylation PTMs cooperate to control DP-IF interactions and that interference with this regulatory machinery contributes to the pathogenesis of skin and heart diseases.

---





**Figure 3.23 Co-localization immunofluorescence microscopy of the association of DP and IF under normal condition and PRMT inhibition condition.**

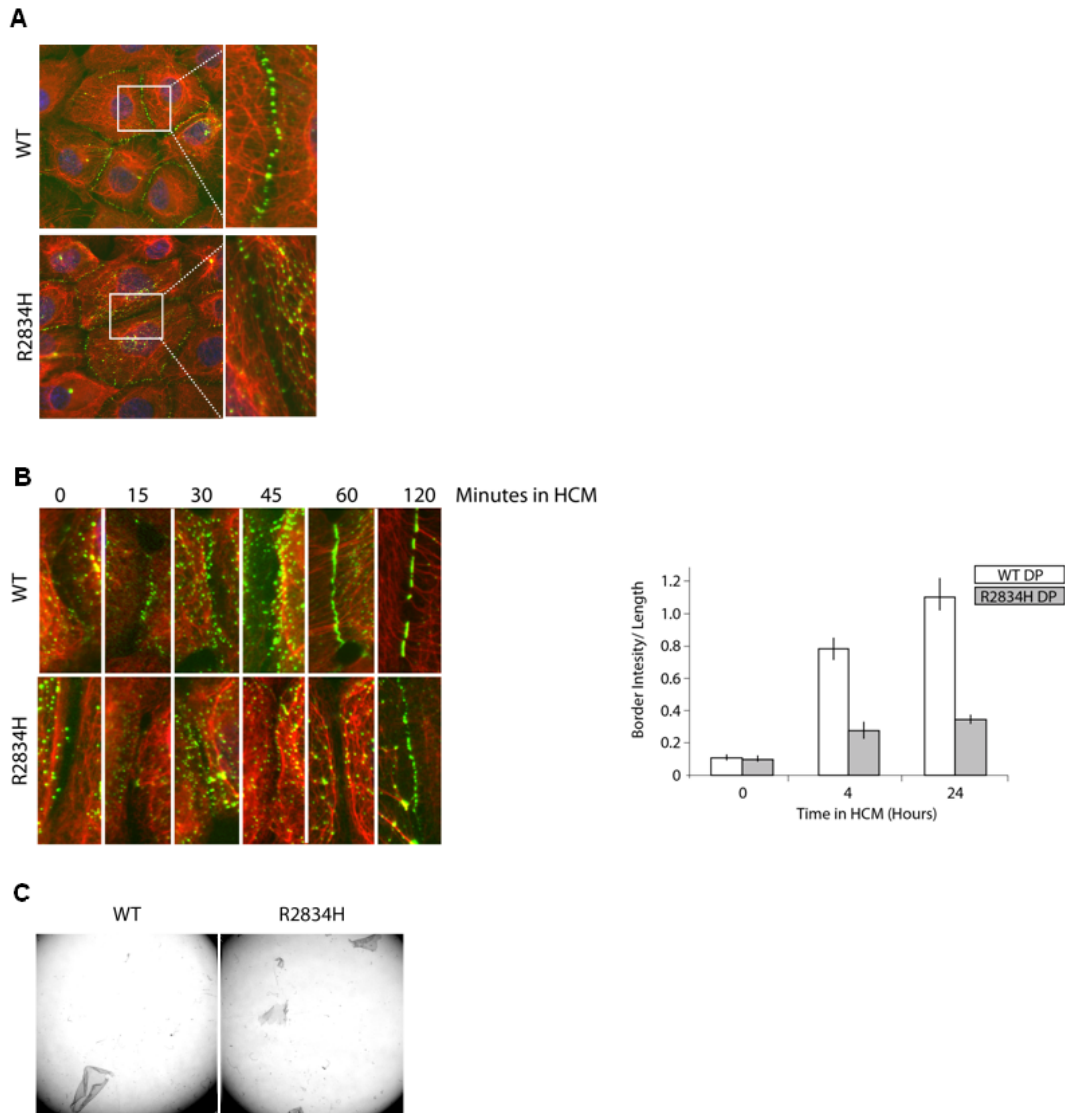
The co-localization immunofluorescence microscopy was performed on SCC9 cells treated with/without PRMT inhibitors. Green: DP; red: IF.

### 3.5.11 ARVC Point Mutation Alters Regulatory Roles of the DP C-terminal Tail

The identification of the ARVC point mutation at residue R2834H in the DP C-terminal tail supports the idea that the phosphorylation-induced structural alteration of the C-tail represents a key regulatory mechanism in mediating DP function and when misregulated promotes the pathogenesis of disease. To dissect the molecular mechanisms underlying the pathogenesis of ARVC, both cell biology and biochemical approaches were utilized. First, the extent to which the R2834H point mutation alters DP localization and trafficking in cells was examined. Compared to DP WT, DP R2834H partially mimicked kinase-inhibited cells where DP localization was enhanced along IF (**Figure 24A**). Accordingly, DP R2834H delayed DP trafficking to borders during desmosome assembly in a calcium switch

assay (**Figure 24B**). To assess the affect of DP R2834H on intercellular adhesion a Dispase Assay was performed. For this analysis cells were grown to confluency and lifted as cell sheets with dispase enzyme. Cell monolayers were then subjected to mechanical stress and the numbers of fragments produced are quantified representing adhesive strength. In the absence of endogenous DP, DP R2834H-expressing cells resulted in a significant increase in fragmentation compared to WT DP cells indicating that R2834H alters the integrity of intercellular junctions (**Figure 24C**). Collectively, these data demonstrate a similarity of R2834H DP behavior to that of the kinase inhibited cells and phospho-deficient S2849G DP mutant, suggesting that this point mutant alters phosphorylation of DP as a mechanism contributing to disease pathogenesis.

---



**Figure 3. 24 ARVC point mutation alters regulatory roles of the DP C-terminal tail.**

A. Co-localization immunofluorescence microscopy of the association of DP (WT and DP R2834H) and IF in SCC9 cells expressing WT DP-S-tag and methylation-deficient point mutant DP-S-tag R2834H. Green: DP; red: IF. B. Methylation-deficient DP R2834H enhances DP-IF colocalization and delays DSM assembly in a calcium switch analysis. SCC9 cells expressing DP or DP R2834H were fixed at different time points after a calcium switch (0–120 min) followed by keratin staining. Left: Quantitation of cell border intensity/cell length in different time points of the calcium switch analysis for WT DP and DP R2834H. C. Dispase Assay to assess the affect of R2834H on intercellular adhesion. Confluent monolayers of SCC9 cells were released using dispase and mechanical stress was applied. The cell fragments were viewed with a dissecting scope.

### 3.6 Conclusions

Mass spectrometric analyses of truncated, tagged desmoplakin C-terminal tail (DP-S-tag) expressed in HEK 293 revealed six novel serine phosphorylation sites and four novel arginine methylation sites on the GSK repeat domain region. Several factors contributed to the success of this PTM analysis. Firstly, epitope protein tags (S-tag) provided the opportunity to selectively purify the protein of interest. Next, the combination of CNBr and AspN as cleavage methods allowed for the generation of peptides containing multiple serines and arginines with appropriate size and charge state for the flowing on-line HPLC separations and ETD MS/MS analysis. In addition, front-end ETD enabled multiple rounds of ETD and accumulation of fragment ions in the C-trap prior to mass analysis in the Orbitrap, linearly enhancing the S/N to facilitate confident ETD spectral interpretation.

In subsequent studies, it was demonstrated that crosstalk between arginine methylation and GSK3-mediated phosphorylation cascades tune DP-IF interactions. The inhibition of arginine methylation of DP was shown to contribute to the underlying molecular mechanisms promoting ARVC. Furthermore, a structural model was provided for the PTM regulation of DP function through the formation of an arginine claw structure within the DP C-terminal tail region.

Over 50 genetic mutations have been identified implicating DP as a key target for inherited disease. These mutations in DP have been shown to promote the

---

pathogenesis of skin-blistering disorders and cardiac myopathies and even cause

death with severe loss of function. However, the importance of this arginine methylation in human pathologies remains largely unknown. Future studies on the role of arginine methylation in desmosome assembly may reveal key aspects of the control of the signaling machinery recruited to form complexes to promote intercellular adhesion. Additionally, identification of arginine methylation of DP may unveil the underlying molecular mechanisms that promote the pathogenesis of ARVC.

Phosphorylation cascades could either represent highly specific modules that mediate protein-protein interactions or alternatively represent the requirement of a robust negative charge signal. The DP C-terminal tail represents an ideal model to further the understanding of the basic principles of protein code directing downstream signaling events. These data raise the possibility that, similar to protein binding motifs, dense clustering of modifiable residues function as discrete information units mediating different signals depending on modification state in various cellular contexts and when dysregulated lead to disease pathogenesis.

---

### 3.7 References

1. Lien, W., Stepniak, E., and Vasioukhin, V. (2008) *Dissecting the role of cadherin-catenin proteins in mammalian epidermis*. Proc. Natl. Acad. Sci. U. S. A. 105, 15225-15226
  2. Furukawa, F., Takigawa, M., Matsuyoshi, N., Shirahama, S., Wakita, H., Fujita, M., Horiguchi, Y., and Imamura, S. (1994) *Cadherins in cutaneous biology*. J. Dermatol. 21, 802-813
  3. Getsios, S., Huen, A. C., and Green, K. J. (2004) *Working out the strength and flexibility of desmosomes*. Nat. Rev. Mol. Cell Biol. 5, 271-281
  4. Godsel, L. M., Getsios, S., Huen, A. C., and Green, K. J. (2004) *The molecular composition and function of desmosomes*. Handb. Exp. Pharmacol., 137-193
  5. Simpson, C. L. and Green, K. J. (2007) *Identification of desmogleins as disease targets*. J. Invest. Dermatol. 127, E15-6
  6. Skerrow, C. J., Clelland, D. G., and Skerrow, D. (1989) *Changes to desmosomal antigens and lectin-binding sites during differentiation in normal human-epidermis - a quantitative ultrastructural-study*. J. Cell. Sci. 92, 667-677
  7. Lee, P. D., Ge, B., Greenwood, C. M. T., Sinnett, D., Fortin, Y., Brunet, S., Fortin, A., Takane, M., Skamene, E., Pastinen, T., Hallett, M., Hudson, T. J., and Sladek, R. (2006) *Mapping cis-acting regulatory variation in recombinant congenic strains*. Physiol. Genomics 25, 294-302
  8. Desai, B. V., Harmon, R. M., and Green, K. J. (2009) *Desmosomes at a glance*. J. Cell. Sci. 122, 4401-4407
  9. Garrod, D. and Chidgey, M. (2008) *Desmosome structure, composition and function*. Biochim. Biophys. Acta-Biomembr. 1778, 572-587
  10. Choi, H. J., Park-Snyder, S., Pascoe, L. T., Green, K. J., and Weis, W. I. (2002) *Structures of two intermediate filament-binding fragments of desmoplakin reveal a unique repeat motif structure*. Nat. Struct. Biol. 9, 612-620
  11. Dephoure, N., Zhou, C., Villen, J., Beausoleil, S. A., Bakalarski, C. E., Elledge, S. J., and Gygi, S. P. (2008) *A quantitative atlas of mitotic phosphorylation*. Proc. Natl. Acad. Sci. U. S. A. 105, 10762-10767
-

12. Godsel, L. M., Hsieh, S. N., Amargo, E. V., Bass, A. E., Pascoe-Mcgillicuddy, L. T., Huen, A. C., Thorne, M. E., Gaudry, C. A., Park, J. K., Myung, K., Goldman, R. D., Chew, T. L., and Green, K. J. (2005) *Desmoplakin assembly dynamics in four dimensions: Multiple phases differentially regulated by intermediate filaments and actin*. J. Cell Biol. 171, 1045-1059
  13. Yin, T. F. and Green, K. J. (2004) *Regulation of desmosome assembly and adhesion*. Semin. Cell Dev. Biol. 15, 665-677
  14. Stappenbeck, T. S., Lamb, J. A., Corcoran, C. M., and Green, K. J. (1994) *Phosphorylation of the desmoplakin cooh terminus negatively regulates its interaction with keratin intermediate filament networks*. J. Biol. Chem. 269, 29351-29354
  15. Amar, L. S., Shabana, A. M., Oboeuf, M., Martin, N., and Forest, N. (1999) *Involvement of desmoplakin phosphorylation in the regulation of desmosomes by protein kinase C, in HeLa cells*. Cell Adhes. Commun. 7, 125-138
  16. Gaudry, C. A., Palka, H. L., Dusek, R. L., Huen, A. C., Khandekar, M. J., Hudson, L. G., and Green, K. J. (2001) *Tyrosine-phosphorylated plakoglobin is associated with desmogleins but not desmoplakin after epidermal growth factor receptor activation*. J. Biol. Chem. 276, 24871-24880
  17. Wallis, S., Lloyd, S., Wise, I., Ireland, G., Fleming, T. P., and Garrod, D. (2000) *The alpha isoform of protein kinase C is involved in signaling the response of desmosomes to wounding in cultured epithelial cells*. Mol. Biol. Cell 11, 1077-1092
  18. Bass-Zubek, A. E., Hobbs, R. P., Amargo, E. V., Garcia, N. J., Hsieh, S. N., Chen, X., Wahl, J. K., III, Denning, M. F., and Green, K. J. (2008) *Plakophilin 2: A critical scaffold for PKC alpha that regulates intercellular junction assembly*. J. Cell Biol. 181, 605-613
  19. Pasdar, M., Li, Z., and Chan, H. (1995) *Desmosome assembly and disassembly are regulated by reversible protein-phosphorylation in cultured epithelial-cells*. Cell Motil. Cytoskeleton 30, 108-121
  20. Sheu, H. M., Kitajima, Y., and Yaoita, H. (1989) *Involvement of protein kinase-C in translocation of desmoplakins from cytosol to plasma-membrane during desmosome formation in human squamous-cell carcinoma-cells grown in low to normal calcium-concentration*. Exp. Cell Res. 185, 176-190
  21. Lapouge, K., Fontao, L., Champiaud, M., Jaunin, F., Frias, M. A., Favre, B., Paulin, D., Green, K. J., and Borradori, L. (2006) *New insights into the molecular basis of desmoplakin- and desmin-related cardiomyopathies*. J. Cell. Sci. 119, 4974-4985
-

22. Meng, J. J., Bornslaeger, E. A., Green, K. J., Steinert, P. M., and Ip, W. (1997) *Two-hybrid analysis reveals fundamental differences in direct interactions between desmoplakin and cell type-specific intermediate filaments*. J. Biol. Chem. 272, 21495-21503
23. Fontao, L., Favre, B., Riou, S., Geerts, D., Jaunin, F., Saurat, J. H., Green, K. J., Sonnenberg, A., and Borradori, L. (2003) *Interaction of the bullous pemphigoid antigen 1 (BP230) and desmoplakin with intermediate filaments is mediated by distinct sequences within their COOH terminus*. Mol. Biol. Cell 14, 1978-1992
24. Daugherty, R. L. and Gottardi, C. J. (2007) *Phospho-regulation of beta-catenin adhesion and signaling functions*. Physiology 22, 303-309
25. Woodgett, J. R. and Cohen, P. (1984) *Multisite phosphorylation of glycogen-synthase - molecular-basis for the substrate-specificity of glycogen-synthase kinase-3 and casein kinase-ii (glycogen-synthase kinase-5)*. Biochim. Biophys. Acta 788, 339-347
26. Mukai, F., Ishiguro, K., Sano, Y., and Fujita, S. C. (2002) *Alternative splicing isoform of tau protein kinase I/glycogen synthase kinase 3 beta*. J. Neurochem. 81, 1073-1083
27. Beals, C. R., Sheridan, C. M., Turck, C. W., Gardner, P., and Crabtree, G. R. (1997) *Nuclear export of NF-ATc enhanced by glycogen synthase kinase-3*. Science 275, 1930-1933
28. Aberle, H., Bauer, A., Stappert, J., Kispert, A., and Kemler, R. (1997) *Beta-catenin is a target for the ubiquitin-proteasome pathway*. EMBO J. 16, 3797-3804
29. Yang, Z., Bowles, N. E., Scherer, S. E., Taylor, M. D., Kearney, D. L., Ge, S., Nadvoretzkiy, V. V., DeFreitas, G., Carabello, B., Brandon, L. I., Godsel, L. M., Green, K. J., Saffitz, J. E., Li, H., Danieli, G. A., Calkins, H., Marcus, F., and Towbin, J. A. (2006) *Desmosomal dysfunction due to mutations in desmoplakin causes arrhythmogenic right ventricular dysplasia/cardiomyopathy*. Circ. Res. 99, 646-655
30. Wu, X., Shen, Q., Oristian, D. S., Lu, C. P., Zheng, Q., Wang, H., and Fuchs, E. (2011) *Skin stem cells orchestrate directional migration by regulating microtubule-ACF7 connections through GSK3 beta*. Cell 144, 341-352
31. Jonkman, M., Pasmooij, A., Pasmans, S., van den Berg, M., ter Horst, H., Timmer, A., and Pas, H. (2005) *Loss of desmoplakin tail causes lethal acantholytic epidermolysis bullosa*. Am. J. Hum. Genet. 77, 653-660
-



32. Al-Owain, M., Wakil, S., Shareef, F., Al-Fatani, A., Hamadah, E., Haider, M., Al-Hindi, H., Awaji, A., Khalifa, O., Baz, B., Ramadhan, R., and Meyer, B. (2011) *Novel homozygous mutation in DSP causing skin fragility-woolly hair syndrome: Report of a large family and review of the desmoplakin-related phenotypes*. Clin. Genet. 80, 50-58
33. Collins, T. J. (2007) *ImageJ for microscopy*. BioTechniques 43, S25-S30
34. Kaiser, R. and Metzka, L. (1999) *Enhancement of cyanogen bromide cleavage yields for methionyl-serine and methionyl-threonine peptide bonds*. Anal. Biochem. 266, 1-8
35. Earley, L., Anderson, L. C., Bai, D. L., Mullen, C., Syka, J. E. P., English, A. M., Dunyach, J., Stafford, G. C. J., Shabanowitz, J., Hunt, D. F., and Compton, P. D. (2013) *Front-end electron transfer dissociation: A new ionization source*. Anal. Chem. 85, 8385-8390
36. Walsh, C. T. (2006) *Posttranslational Modification of Proteins: Expanding Nature's Inventory*, Roberts and Company Publishers, Greenwood Village, CO
37. Bedford, M. T. (2007) *Arginine methylation at a glance*. J. Cell. Sci. 120, 4243-4246
38. Hamelberg, D., Shen, T., and McCammon, J. A. (2007) *A proposed signaling motif for nuclear import in mRNA processing via the formation of arginine claw*. Proc. Natl. Acad. Sci. U. S. A. 104, 14947-14951
39. Swaney, D. L., Beltrao, P., Starita, L., Guo, A., Rush, J., Fields, S., Krogan, N. J., and Villen, J. (2013) *Global analysis of phosphorylation and ubiquitylation cross-talk in protein degradation*. Nature Methods 10, 676-682
40. Hunter, T. (2007) *The age of crosstalk: Phosphorylation, ubiquitination, and beyond*. Mol. Cell 28, 730-738
41. Gao, M. and Karin, M. (2005) *Regulating the regulators: Control of protein ubiquitination and ubiquitin-like modifications by extracellular stimuli*. Mol. Cell 19, 581-593
42. Boisvert, F. M., Cote, J., Boulanger, M. C., and Richard, S. (2003) *A proteomic analysis of arginine-methylated protein complexes*. Molecular & Cellular Proteomics 2, 1319-1330
-

#### **4. Size-controlled Digestion for Middle-down nanoHPLC-MS/MS**

### **Characterization of the Sequence and Post-translational Modifications of Monoclonal Antibody**

#### **4.1 Abstract**

Therapeutic monoclonal antibodies (mAbs) have brought a paradigm shift in biotechnology and pharmaceutical sectors. They are now used for the treatment of cancer, autoimmune, and inflammatory diseases. Mass spectrometry (MS) is the primary analytical tool for mAb characterization throughout drug discovery and product process development. Digesting large proteins (e.g. mAb) into 3-15 kDa large peptides for peptide sequencing (“middle-down” MS) has many advantages over traditional “bottom-up” and the more recent “top-down” approaches. Middle-down MS represents an attractive method for characterization of large proteins. This chapter demonstrates a novel, size-controlled digestion method that digests mAbs into mainly 3-8 kDa large peptides for middle-down MS analysis sequence and post-translational modifications (PTMs). Pushing a sample solution containing urea-denatured protein through an enzyme-column (capillary column packed with POROS AL beads immobilized with aspergillopepsin I) with a fixed flow rate allows for continuous digestion of the proteins into peptides with a consistent distribution of sequences and sizes. This strategy overcomes the challenge of mAb denaturation by taking the advantage of the high tolerance of aspergillopepsin I to high

---

concentration of urea. Moreover, the peptide sizes can be controlled by adjusting the flow rate of the protein sample passing through the enzyme-column. Digestion of apomyoglobin using this enzyme-column demonstrated increased peptide sizes with reduced protein residence time in the enzyme-column (i.e. increased sample flow rate). When coupling with nanoHPLC-MS/MS with electron-transfer dissociation (ETD), this middle-down strategy yielded sequence coverage of 97% for apomyoglobin. This strategy, combining both ETD and collision-activated dissociation (CAD) in nanoHPLC-MS/MS, successfully characterized a Waters standard mAb, yielding sequence coverage of 98% for light chain (LC) and 94% for heavy chain (HC). Moreover, mAb PTMs including pyroglutamic acid formation, oxidation, amidation, and glycosylation have also been identified using this novel method.

## 4.2 Introduction

### 4.2.1 Antibodies as Therapeutics

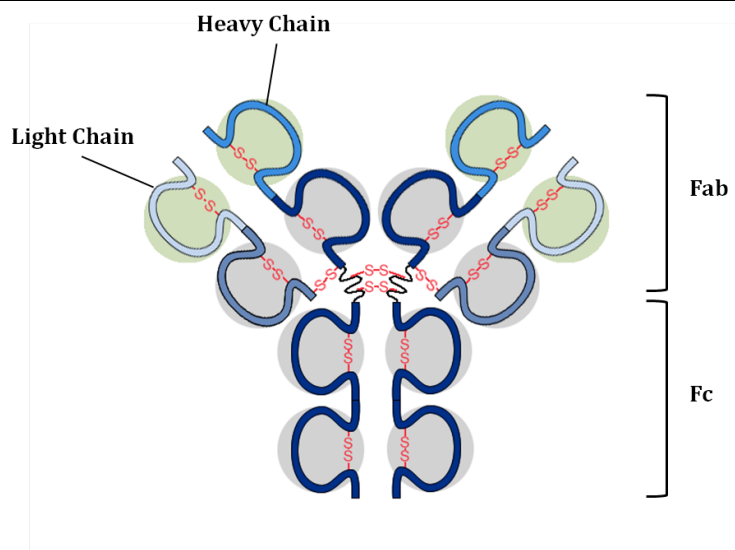
Antibodies are natural inhibitors of pathogens produced by B lymphocytes (1). The development of *in vitro* production of antibodies by cell fusion and recombinant DNA technology enabled the use of antibodies as therapeutic agents (2, 3). Recombinant mAbs have become one of the most promising drug classes for human therapeutic treatments in various diseases including inflammation, transplantation, oncology, autoimmune, cardiovascular, and infectious diseases (4,

---

5). Compare to conventional small-molecule drugs, mAbs have several advantages such as high specificity and fewer side effects (6-8).

Monoclonal human immunoglobulin gamma (IgG) antibodies are the most abundant antibodies among different classes of human antibodies (6). All the antibody drugs approved for clinical use are IgG antibodies. As shown in **Figure 4.1**, a typical human IgG (~150 kDa) is composed of two identical ~50 kDa HCs and two identical ~25 kDa LCs, which are held together by multiple disulfide bonds to form a Y-shaped structure (1). The antigen-binding sites locate within the variable domains of HC and LC (the domains with green shadows in **Figure 4.1**). These domains contain complementarity-determining regions (CDRs) that differ mostly in their sequences compared to other regions of antibody, and are responsible for the diversity and specificity of antibody binding (1). The number and location of disulfide bonds between two heavy chains and between heavy and light chains divide the human IgGs further into four subclasses: IgG1, IgG2, IgG3, and IgG4 (6). To facilitate structural characterization, IgG is often cleaved with papain or other proteases at the hinge region to form two fragments, Fab (antigen binding fragment) and Fc (crystalizable fragment) (**Figure 4.1**).

---



**Figure 4. 1** Schematic drawing of human IgG1.

Fab: antigen-binding fragment. Fc: crystalizable fragment.

#### 4.2.2 Structure Characterization of Therapeutic Antibodies

The amino acid sequence of a mAb is homogeneous, however, the final product of therapeutic antibodies have heterogeneities resulting from various modifications that occur during different stages of production, such as mutations, C-terminal lysine processing, pyroglutamic acid formation, oxidation, amidation, deamidation, glycosylation, and disulfide linkages (6, 9, 10). Identification of the primary sequence of therapeutic mAbs, as well as elucidation of the N-glycan structures, disulfide linkages and other PTMs, is critical for the evaluations of drug safety, efficacy, stability, as well as understanding the structure/function relationships (6, 11).

The demands for characterization of therapeutic mAbs are increasing with the rapid development of mAb-based pharmaceuticals. MS is the most powerful techniques for the structural characterization of therapeutic mAbs due to its high accuracy, resolution, and speed over other analytical techniques (6).

### **4.2.3 Structure Characterization Strategies of Protein Using Mass**

#### **Spectrometry**

The most widely employed MS-based strategies for protein structural characterizations are “bottom-up” approaches. This method relies on complete enzymatic digestion of the proteins into small peptides prior to MS analyses (6). Typically proteins are reduced, alkylated, and digested with proteases that cleave proteins at two or more types of amino acid residues and generate relatively short peptides (typically <3000 Da) (12, 13). These enzymes include: trypsin (most widely used, cleaves peptide bonds at the C-terminal side of Lys and Arg except when either is followed by Pro), pepsin (cleaves peptide bonds at hydrophobic and preferably aromatic amino acids such as Phe, Trp, and Tyr), chymotrypsin (cleaves peptide bonds at the C-terminal side of aromatic amino acid residues, including Tyr, Trp and Phe) (14-16). Although this approach is very sensitive and can provide high resolution structural information for proteins, the digestion process is time-consuming and labor-intensive in terms of sample preparation and data analysis (6). More importantly, these peptides are relatively small in size (typically 8~25 residues long), resulting in problems such as sample complexity, difficulties in

---

assigning peptides to specific gene products, and loss of combinatorial PTM

information (17). Moreover, the bottom-up strategy usually provides incomplete sequence coverage with digestion using a single protease.

In recently years, direct analysis of intact proteins by MS (often called “top-down” MS) has become more attractive and will continue to be so in the near future. This approach overcomes some of the issues that the bottom-up approach faces by introducing the intact protein directly into the mass spectrometer for analysis (18). The top-down method using high resolution mass analyzers (e.g. FT-ICR, Orbitrap, and TOF) is successful for rapid characterization of small to medium-sized (<50 kDa) proteins (19, 20). However, for larger proteins such as mAbs (150 kDa), sequence and PTM coverage is far from complete due to insufficient fragmentation, low signal-to-noise ratio (S/N) of fragment ions, and the requirement for high resolution in the high  $m/z$  region, presenting challenges for even the best performing instruments (6).

Between bottom-up and top-down analysis a “compromise” is the “middle-down” approach, in which peptides with 3-20 kDa sizes are analyzed (21-25). Recent years have seen the use of some proteases that cleave proteins at a single type of residue appearing relatively infrequently along the protein chain, such as Lys-C (cleaves at the C-terminally of Lys), Asp-N (cleaves N-terminal side of Asp), and Glu-C (cleaves at the C-terminal side of Glu). These proteases digest proteins into peptides on average larger than tryptic peptides (26-28). Large proteolytic

---

peptides have several advantages for protein characterization. Firstly, larger peptides can reveal information about protein isoforms, variants, and combinatorial PTMs (21). Secondly, they are generally more hydrophobic than short peptides and have better resolution on reversed-phase LC (25). Thirdly, compared to small tryptic peptides, large peptides carry more Arg/Lys/His residues which facilitate electrospray ionization and MS2 fragmentation. Fourthly, for protein identification with a database search, larger peptides usually have fewer potential source protein candidates in the database than small ones, leading to higher protein identification confidence (23). Finally, compared to the top-down MS method, the middle-down strategy has lower instrumental requirements (e.g. sensitivity, resolution) to achieve strong S/N of fragment ions for protein characterization.

Several studies showed that the middle-down approach is a valuable method for MS characterization of proteins. For example, Garcia *et al.* demonstrated the power of a middle-down approach to localize multiple PTMs at histone H3 tails (1-50 residues) using Glu-C digestion followed by ECD MS/MS (29). Ge *et al.* unambiguously identified phosphorylation sites in the truncated and full-length forms of cardiac myosin binding protein C (cMyBP-C) using top-down and middle-down MS (30). This combination of approaches ensures the full sequence coverage of cMyBP-C. Recently, Hess *et al.* reported that a novel strategy was developed using middle-down ECD coupled with online nanoflow-LC for the analysis of peptides

---



from various cell lines (22).

Examinations of the abundance of amino acid residues in proteins from the UniProt database show the percentages for Lys, Asp, and Glu are 4~6% (31). However, none of those amino acids are evenly distributed in protein sequences, meaning that protein digestion using proteases of single amino acid specificity can still produce many short peptides due to two cleavable amino acid residues very close to each other. On the other hand, if two cleavable amino acid residues happen to be separated far from each other by a long piece of non-cleavable amino acid sequence (e.g. >15 kDa), the generated peptide will be too long to be efficiently sequenced (just as the issue in top-down approach). Therefore sometimes multiple proteases are employed separately to digest a target protein into large peptides with overlapping sequences to cover the whole sequence of the protein. To avoid performing multiple rounds of proteolysis, new strategies other than the conventional in-solution digestion have been reported to control the size of peptides for middle-down MS analysis. The Bruening group designed a novel strategy for the generation of large tryptic peptides by rapidly pushing the protein sample solution through a trypsin-modified nylon membrane using a syringe pump (32). Precise control of the sample flow rate allows for the tuning of the residence time of the protein in the membrane pores, which determines the extent of protein proteolysis and the size of the resulting peptides. This size-controlled digestion was

---

demonstrated to be useful for protein identification via “peptide finger printing”

when equipped with ion/ion reaction MS technique. This strategy was further demonstrated by the Bruening group in the sequencing of large peptides of myoglobin and bovine serum albumin using ETD MS/MS as well as higher order structural characterization of membrane protein Arabidopsis GTPase, Root Hair Defective 3 (33).

#### **4.2.4 Designing Rationales of the Size-controlled Digestion for Middle-down nanoHPLC-MS/MS Characterization of Monoclonal Antibody**

Efficient proteolytic digestion requires full denaturation of the proteins so that the cleavage sites are easily accessible to the proteases. However, it is a challenge for mAbs due to their highly folded secondary and tertiary structures. Different denaturation methods have been used to facilitate the digestion of proteins, for example, strong acid, salt, organic solvent, and heat (34). Urea and guanidinium chloride have been widely used as chaotropic agents to denature highly folded proteins (35). However, the degree of denaturation needs to be carefully controlled in proteolysis, in order to preserve the activity of proteases.

In this work, aspergillopepsin I, often known as its commercial name Protease from *Aspergillus saitoi* Type XIII, was employed as the protease for mAb digestion due to its persistent activity in harsh conditions. This protease was isolated from *Aspergillus saitoi*, a microorganism used in fermentation of traditional

---

Japanese liquors (36). Aspergillopepsin I from *Aspergillus saitoi* is a protease with low amino acid selectivity. It favors cleavage at many hydrophobic amino acid residues in both C-terminal and N-terminal sides, as well as at C-terminal side of Lys (37, 38). Different from other widely used proteases, aspergillopepsin I tend to present continuous enzyme activity at acidic condition. It has been reported that the optimal pH of aspergillopepsin I for milk casein digestion is in the range of 2.5-3.0 and the protease is fairly stable over the pH range of 2.5-6.0 (39). Furthermore, no autolysis occurs at pH 5.5 and at 22°-25°, and the loss of enzymatic activity is only about 5% after 16 hours (40). Denaturation and inactivation of aspergillopepsin I have been reported to begin when pH is above 6-7, and the degree of inactivation in the presence of 5 M urea of an apparent pH of 5.5 is 46% after 30 minutes and 59% after 1 hour (40, 41). The high tolerance to urea makes aspergillopepsin I a candidate enzyme to digest mAbs in their fully denatured conditions.

Aspergillopepsin I has not been widely employed in proteomics studies yet, probably because of its low amino acid specificity. Recently it was used in Hydrogen/Deuterium (H/D) exchange study of large proteins (42, 43). Aspergillopepsin I digested proteins into peptides with average length of 10 amino acids at pH 2.3-2.5 and 0 °C in only 2 minutes, suggesting that this enzyme has rather high tolerance for low pH and low temperature, as well as fast digestion speed (42, 43). However, the high activity (probably partially due to the broad

---

cleavage selectivity at multiple types of amino acids) of aspergillopepsin I at low pH

also brings issues in conventional in-solution digestion. Protein proteolysis by simply mixing the protein sample with aspergillopepsin I is uncontrollable and difficult to quench. Therefore, new strategies need to be developed to control the extent of protein digestion such that aspergillopepsin I can be fully utilize to generate peptides with desired size and sample complexity.

This chapter presents a novel approach using on-column aspergillopepsin I digestion to generate large peptides for middle-down MS analysis of mAbs sequence and PTMs.

### **4.3 Materials**

#### **4.3.1 Reagents**

Bio-Rad (Hercules, CA)

Tris(hydroxymethyl)aminomethane, Tris

#### **Mallinckrodt, Inc. (Paris, KY)**

Acetonitrile, HPLC grade,  $\geq 99.8\%$  purity

Ammonium acetate,  $\text{NH}_4\text{Ac}$

Sodium sulfate,  $\text{Na}_2\text{SO}_4$

#### **New England Biolabs, Inc. (Ipswich, MA)**

---

---

Histone H2A, Human, Recombinant

Oxford Glycosystems Ltd. (Oxford, U.K.)

Tris-HCl

**PQ Corporation (Valley Forge, PA)**

Kasil® 1624

**Roche Diagnostics (Indianapolis, IN)**

Endoproteinase AspN

Endoproteinase LysC

**Sigma Aldrich (St. Louis, MO)**

Ammonium bicarbonate,  $\text{NH}_4\text{HCO}_3$

Ammonium hydroxide solution,  $\text{NH}_3 \cdot \text{H}_2\text{O}$

Angiotensin I acetate salt hydrate,  $\geq 90\%$  purity (human)

Apomyoglobin from equine skeletal muscle Azulene

Bovine serum albumin

Fluoranthene, 99% purity

Formamide

---

---

Glacial acetic acid, ≥99.9% purity

Iodoacetamide

N-(2-Aminoethyl)maleimide trifluoroacetate salt (referred to as aminoethylmaleimide and NAEM in this Chapter)

Potassium ferricyanide,  $K_3Fe(CN)_6$

Protease from *Aspergillus saitoi*, Type XIII

Sodium cyanoborohydride,  $NaCNBH_3$

Trifluoroacetic acid, TFA

Tris(2-carboxyethyl)phosphine hydrochloride, TECP•HCl

Ubiquitin human

Vasoactive intestinal peptide fragment 1-12, ≥ 97% purity

Thermo Fisher Scientific (Rockford, IL)

Formic acid

Isopropanol

Pierce® Water, LC-MS Grade

Urea

---

Intact mAb Mass Check Standard [186006552]

#### **4.3.2 Equipment & Instrumentation**

##### **Agilent Technologies (Palo Alto, CA)**

HP1100 high performance liquid chromatograph

POROSHELL 300SB-C18 (5 µm diameter)

Drummond Scientific Company (Broomall, PA)

5-µL calibrated pipet

Eppendorf (Hauppauge, NY)

5415R Benchtop Centrifuge

Eppendorf tube, 0.5 mL

Eppendorf tube, 1.5 mL

##### **Labconco Corporation (Kansas City, MO)**

Centrivap Concentrator

##### **Life Technologies Corporation (Grand Island, NY)**

POROS® AL 20 µm Self Pack® Media

---

---

**New Objective, Inc. (Woburn, MA)**

PicoClear® microcapillary union

Polymicro Technologies, LLC (Pheonix, AZ)

360 µm o.d. x 50 µm i.d. polyimide coated fused silica capillary

360 µm o.d. x 75 µm i.d. polyimide coated fused silica capillary

360 µm o.d. x 150 µm i.d. polyimide coated fused silica capillary

**Sutter Instrument Co. (Navato, CA)**

P-2000 microcapillary laser puller

Thermo Fisher Scientific (Rockford, IL)

0.8 mL Spin column from Pierce® Fab Preparation Kits

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

LTQ-Orbitrap Velos hybrid mass spectrometer

---



## 4.4 Methods

### 4.4.1 Covalent Linkage of Aspergillopepsin I to POROS 20 AL Beads

Near saturated  $\text{Na}_2\text{SO}_4$  solution was prepared by dissolving 0.28 g  $\text{Na}_2\text{SO}_4$  in 1 mL water at 80 °C, letting it cool down, and removing all the residual insoluble  $\text{Na}_2\text{SO}_4$  using centrifugation. Aspergillopepsin I (protease from *Aspergillus saitoi*, Type XIII) was dissolved in this solution at 10 mg/mL followed by centrifugation to remove the precipitations. Next, 200  $\mu\text{L}$  of the aspergillopepsin I/  $\text{Na}_2\text{SO}_4$  solution was mixed with 7 mg POROS 20 AL beads, followed by addition of 1  $\mu\text{L}$  of 80 mg/mL of  $\text{NaCNBH}_3$  in near saturated  $\text{Na}_2\text{SO}_4$  solution. The mixture was gently shaken for 19 hours at room temperature for protease conjugation, and then transferred into a spin column filter (pore size smaller than 20  $\mu\text{m}$ ). The conjugation solution was removed from the spin filter by centrifugation (2,000 rcf for 30 s, the same for the following steps). The modified POROS AL beads were washed by adding 0.5 mL of water into the spin column filter, shaking for 1 minute, and removing the water by centrifugation. This step was repeated for two more cycles. The unreacted aldehyde groups on the POROS AL 20 beads were blocked with tris(hydroxymethyl)aminomethane (Tris) by adding 300  $\mu\text{L}$  of 0.2 M Tris-HCl buffer (pH 6.5) (prepared by mixing 60 volumes of 0.2 M Tris-HCl with 5 volumes of 0.2 M Tris) and 1  $\mu\text{L}$  of 80 mg/mL  $\text{NaCNBH}_3$  (prepared in 0.2 M Tris-HCl buffer, pH 6.5) to the washed and dried beads in the spin filter, followed by shaking the beads suspension for 2 hours at room temperature. Finally, the POROS AL beads modified

---

with aspergillopepsin I (enzyme-beads) were washed with water for three times in the spin column, dried by centrifugation, and stored at 4 °C.

#### **4.4.2 Fabrication of Enzyme-column and On-column Digestion**

An enzyme-column was constructed using 360 µm o.d. X 150 µm i.d. fused silica capillary with a 2 mm Kasil® 1624 frit in the top of the capillary. The fused silica capillary was washed with first 0.1% acetic acid in 40% acetonitrile/60% water and then water by pressure-loading the corresponding solutions. The modified-POROS AL beads were suspended in 200 µL water and pressure-packed onto the fused silica capillary for 2-14 cm (depending on the desired residence time).

#### **4.4.3 Protein Reduction, Alkylation and Size-controlled On-column Digestion**

Digestion buffer (pH 3.9) was prepared by mixing 100 µL of 8 M urea in 0.5 M ammonium acetate, 8 µL of 25% formic acid, and 892 µL of 8 M urea in 0.5 % acetic acid. Prior to protein digestion, the enzyme-column was washed with the digestion buffer; the flow rate was adjusted and measured by changing the pressure applied on the enzyme-column.

**Digestion of apomyoglobin:** Apomyoglobin from equine skeletal muscle was dissolved in the digestion buffer at a concentration of 0.2 µg/µL and loaded on to the enzyme-column at various flow rates to achieve different digestion times. The digested protein was collected in an Appendorf tube (0.5 mL) and stored at -35 °C.

---

**Digestion of mAb:** A waters standard antibody was reduced, alkylated

before on-column digestion. Twenty  $\mu\text{g}$  of mAb (Intact mAb Mass Check Standard [186006552], Waters) was dissolved in 10  $\mu\text{L}$  of 10 mM tris(2-carboxyethyl)phosphine hydrochloride (TCEP•HCl)/8 M urea in 0.1% acetic acid followed by incubation at 50°C for 10 minutes for disulfide bonds reduction. The sample was then neutralized to pH 6.8 by adding 2.1  $\mu\text{L}$  of 0.2 M  $\text{NH}_3\cdot\text{H}_2\text{O}$ . Cys alkylation was performed by adding 10  $\mu\text{L}$  of 20 mM aminoethylmaleimide/8M urea in 0.5 M ammonium acetate (pH=6.8) for 10 minutes, followed by adding 1  $\mu\text{L}$  of 25% formic acid to adjust the pH to 3.9. The alkylated mAb was finally diluted to 0.2  $\mu\text{g}/\mu\text{L}$  with the digestion buffer, and on-column digested as described for apomyoglobin digestion.

**4.4.4 Protein In-solution Digestion with LysC and AspN**

Prior to in-solution digestion, 20  $\mu\text{g}$  of mAb was dissolved in 7  $\mu\text{L}$  of 2 mM TCEP•HCl/8 M urea in 0.1% acetic acid followed by incubation at 50°C for 10 minutes for disulfide bonds reduction as described before. As LysC and AspN digestion was performed in weak basic condition, which hydrolyzes aminoethylmaleimide, the reduced mAb was thus alkylated with iodoacetamide by adding 7  $\mu\text{L}$  of 20 mM iodoacetamide/8M urea in 2 M ammonium bicarbonate (final pH ~8) to the reduced sample followed by incubation in the dark at room temperature for 30 minutes. Unreacted iodoacetamide was quenched by adding 6  $\mu\text{L}$  of 30 mM dithiothreitol/8 M urea in 100 mM ammonium bicarbonate followed

---

by incubation in the dark at room temperature for 20 minutes. To prevent protease from being denatured and deactivated with 8 M urea, the sample was diluted by fivefold, followed by adding 2  $\mu\text{L}$  of 0.5  $\mu\text{g}/\mu\text{L}$  LysC or AspN (enzyme to substrate ratio 1:20). The digestion was carried out at 37 °C for 10 hours and quenched to pH 3 with acetic acid.

#### 4.4.5 HPLC Capillary Column Assembly

An analytical column was constructed using 360  $\mu\text{m}$  o.d. X 50  $\mu\text{m}$  i.d. fused silica capillary with a 2 mm Kasil® 1624 frit. The analytical column was packed with 15 cm of POROSHELL 300SB-C18 (5  $\mu\text{m}$  diameter) and equipped with a laser-pulled electrospray emitter tip. A precolumn was constructed using 360  $\mu\text{m}$  o.d. x 150  $\mu\text{m}$  i.d. fused silica capillary with a 2 mm Kasil® 1624 frit and packed with 11 cm of the same resin. Both columns were conditioned by first loading 10 pmols of angiotensin I and histone H2A tryptic digest followed by three cycles of HPLC gradients to optimize chromatographic performance. The two columns were further conditioned with 200 fmols of ubiquitin, apomyoglobin and bovine serum albumin similarly until achieving repeatable chromatographic performance (retention time within 0.3 minutes for a 2-8 kDa peptide between two adjacent gradients). The column was then ready for use.

---

#### 4.4.6 HPLC-MS/MS Analyses

One pmol of protein digests (for antibody 1 pmol corresponds to 2 pmol of reduced LC /HC) was pressure-loaded onto the precolumn on a pressure bomb at a flow rate of 1  $\mu$ L/min. Internal standard peptides (200 fmol, angiotensin and vasoactive intestinal peptide) were co-loaded onto the column for relative quantification. The sample was desalted by rinsing the precolumn with solvent A on an HPLC for 10 min at a flow rate of 1  $\mu$ L/min. The precolumn was then connected to the analytical column using a microcapillary union. Peptides were gradient eluted at a flow rate of 60 nL/min from the HPLC column and electrosprayed into the mass spectrometer. The LC gradient was as follows: 0-25% of solvent B in 5 min, 25-60% solvent B in 115 min, 60-100% of solvent B in 4 min, 100% - 0% of solvent B in 4 min, 100% of solvent A for 10 min (solvent A: 0.3% formic acid in water; solvent B: 0.3% formic acid, 72% acetonitrile, 18% isopropanol and 9.7% water).

Typical mass spectrometric analyses include a full MS scan for sample evaluation (Experiment #1), and tandem MS scans for peptide sequencing and PTM characterization (Experiment #2). Experiment #1 included high resolution MS1 scans in the Orbitrap (resolving power of 60,000 at  $m/z$  400), giving the  $m/z$  information of each peptide. For the following targeted MS2 experiments, the MS1 total ion count (TIC) chromatograms from Experiment #1 was divided into multiple segments where the division points were selected at the lowest valleys of the

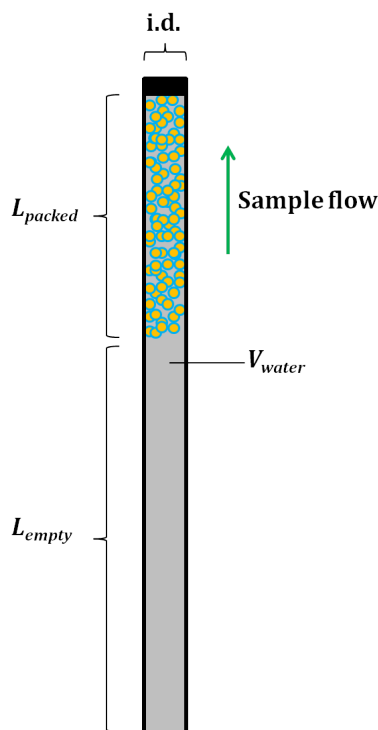
---

peptides TIC signals. To find the major large peptides, the MS1 information in each segment was simplified by converting the  $m/z$  to molecular weight using Xcalibur Xtract. The  $m/z$  and the elution times of the abundant peptides were later used in Experiment #2 where targeted MS/MS were performed for each peptide. In Experiment #2, every MS1 scan was followed by targeted ETD MS2 for the 1-6 most abundant peptides found in a 5-10 minutes segment. Both MS1 and MS2 were analyzed in the Orbitrap. For mAb analysis, CAD was utilized in a similar way in a separate experiment (Experiment #3) to generate complimentary peptide sequence information. Precursor isolation window for MS2 was typically set as 5  $m/z$ . For a few peptides with the selected  $m/z$  too close to other nearby ions (information obtained from Experiment #1), 3  $m/z$  or 4  $m/z$  isolation window was chosen. The ETD reaction time was set based on the following formula,  $t = 50 \text{ ms} \times (3/\text{charge state})^2$ . Automated gain control (AGC) was set as  $1e6$  for both FTMS and FTMSn, and  $3e5$  for ETD reagent.

---

## 4.5 Results & Discussion

### 4.5.1 Principles of Size-Controlled Digestion Method



**Figure 4. 2** A diagram of enzyme-column.

The fused silica capillary was packed with POROS AL beads modified with aspergillopepsin I. *i.d.* is the inner diameter of the capillary column (150  $\mu\text{m}$ ). Three measureable values are labeled as  $L_{\text{packed}}$ , representing the length of the portion of the column packed with enzyme-beads;  $L_{\text{empty}}$ , representing the length of the empty portion of the column; and  $V_{\text{water}}$ , representing the total volume of water trapped in the whole column including the portion packed with beads.

For the column with a given  $L_{\text{packed}}$ , the protein digestion time (i.e. protein residence time in this work) can be controlled by the flow rate as shown in

**Equation 4.1** where  $p$  is the porosity of the enzyme-beads,  $F$  is the flow rate of the protein sample flowing through the enzyme-column, and *i.d.* and  $L_{\text{packed}}$  are defined in Figure 4.2.

$$(4.1) \ t = \frac{\pi(i.d./2)^2 L_{packed} p}{F}$$

The porosity,  $p$ , of the enzyme-beads is 30%. This value was determined by first loading the enzyme-column with water, then measuring the volume of the water trapped in the whole enzyme-column (including both the packed portion and the empty portion) using a 5- $\mu$ L calibrated pipet as the water was pushed off the column by air pressure. The porosity  $p$  was calculating using **Equation 4.2**, where  $V_{water}$  is the volume of the water trapped in the whole column (shown in grey in **Figure 4.2**),  $V_{empty}$  is the volume of the empty portion of the column, and  $V_{packed}$  is the volume of the packed portion of the column.

$$(4.2) \ p = \frac{V_{water} - V_{empty}}{V_{packed}} \times 100\%$$

$$V_{empty} = \pi\left(\frac{i.d.}{2}\right)^2 L_{empty}$$

$$V_{packed} = \pi\left(\frac{i.d.}{2}\right)^2 L_{packed}$$

For practical use, **Equation 4.1** can be simplified to **Equation 4.3** when the  $i.d.$  is 0.0150 cm, and the units for  $t$ ,  $L_{packed}$  and  $F$  are s, cm, and  $\mu$ L/s, respectively. Thus, for a packed enzyme-column with known  $L_{packed}$ , the value  $t$  (residence time of the protein passing through the enzyme-beads) can be precisely controlled by adjusting the loading pressure to maintain a constant flow rate of the protein sample in the enzyme-column. Precise control of flow rate ensures repeatable

---



digestion of a sample with a fixed protein concentration and buffer composition.

That is because, as the protein flows through the enzyme-column, the most "active" digestion sites, determined by both the primary and secondary protein structures, hydrolyze first. The initial degradation products (large peptides) then undergo further hydrolysis at other "active" digestion sites to form shorter peptides. For a column with a given  $L_{packed}$ , constant flow rate  $F$  of the protein sample results in a fixed residence time for any given cross-section of the flowing stream as the stream passes through the enzyme-beads. Assuming there is no retention of proteins or peptides on the hydrophilic enzyme-beads, each single protein molecule (combining its subsequent degraded forms) should have the same residence time as it flows through the enzyme-beads. Considering the random encounters between proteins and enzymes, as well as between the resulting large peptides and enzymes, a repeatable statistical distribution of peptides with different sequences should be obtained when the flow rate is precisely controlled.

$$(4.3) \ t \ (s) = \frac{L_{packed}(cm)}{F(\mu L/s)} \times 3.19$$

#### 4.5.2 Evaluation of Size-Controlled Digestion Method Using Apomyoglobin

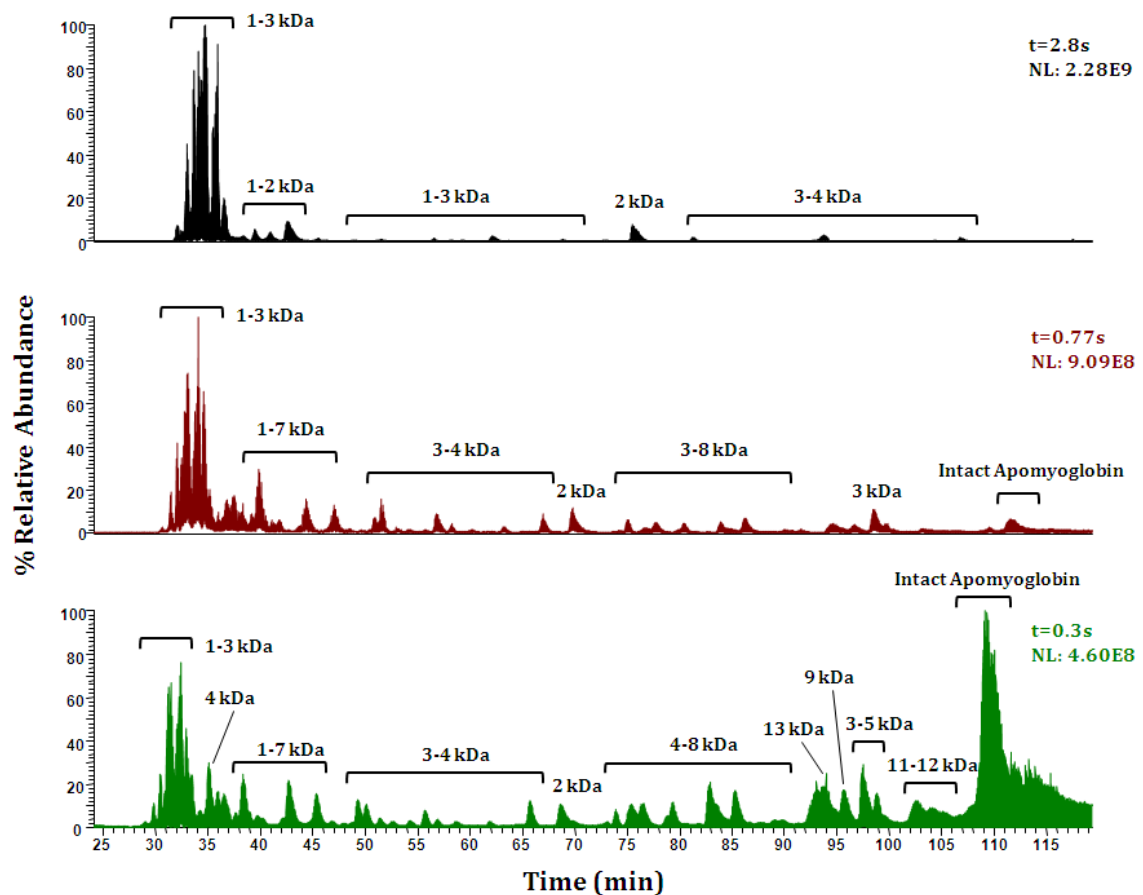
**Figure 4.3** shows the comparison of apomyoglobin size-controlled digestions with three different digestion times. Small peptides below 3 kDa dominated the chromatogram when the digestion time was 2.8 s. When the digestion time was decreased to 0.77 s, the amount of small peptides (reflected by

---

the absolute intensity of TIC decreased to ~40% and many peptides with molecular weight 3-8 kDa appeared. These large peptides eluted through 50-90 min of the gradient with base peak separation (**Figure 4.4**). These peptides are considered to be “golden-sized” peptides in this work, because (1) they typically give charge states of +5—+10 which are favorable to ETD MS2; (2) they are well separated with symmetrical chromatographic peak shapes on the Poroshell C18 column owing to proper hydrophobicity; (3) compared to top-down analysis of intact proteins, middle-down approach yields fragment ions with strong intensities in each MS2 scan, allowing for the sequencing of large peptides in chromatographic time scale. When the digestion time was further decreased to 0.3 s, the “golden-sized” peptides were still present but the undigested intact apomyoglobin was very dominant at the end of the gradient, suggesting an inefficient digestion due to short protein residence time in the enzyme-column. In addition, some peptides over 10 kDa co-elute as tailing chromatographic peaks, which could affect efficient precursor isolation for MS2. Moreover, producing a high quality MS2 spectrum from peptides over 10 kDa requires more MS2 scans than the “golden-sized” peptides. This is because, under a given AGC (1E6 in this work), the S/N of fragment ions in MS2 spectrum drops significantly as the peptide size increases over 10 kDa. Therefore, in this work, peptides over 10 kDa are considered less compatible with rapid online LC-MS/MS. The above issues are more severe with digests generated from larger proteins, as more co-eluting peptides over 10 kDa may fall into the same precursor isolation window (5 m/z in this work) and produce mixed MS2 spectra. In addition,

---

the MS2 scan time allocated to each of these large peptides may not be enough to obtain high quality MS2 spectra. In the end, we chose 0.77 s as the optimal size-controlled digestion time for middle-down analysis of apomyoglobin.



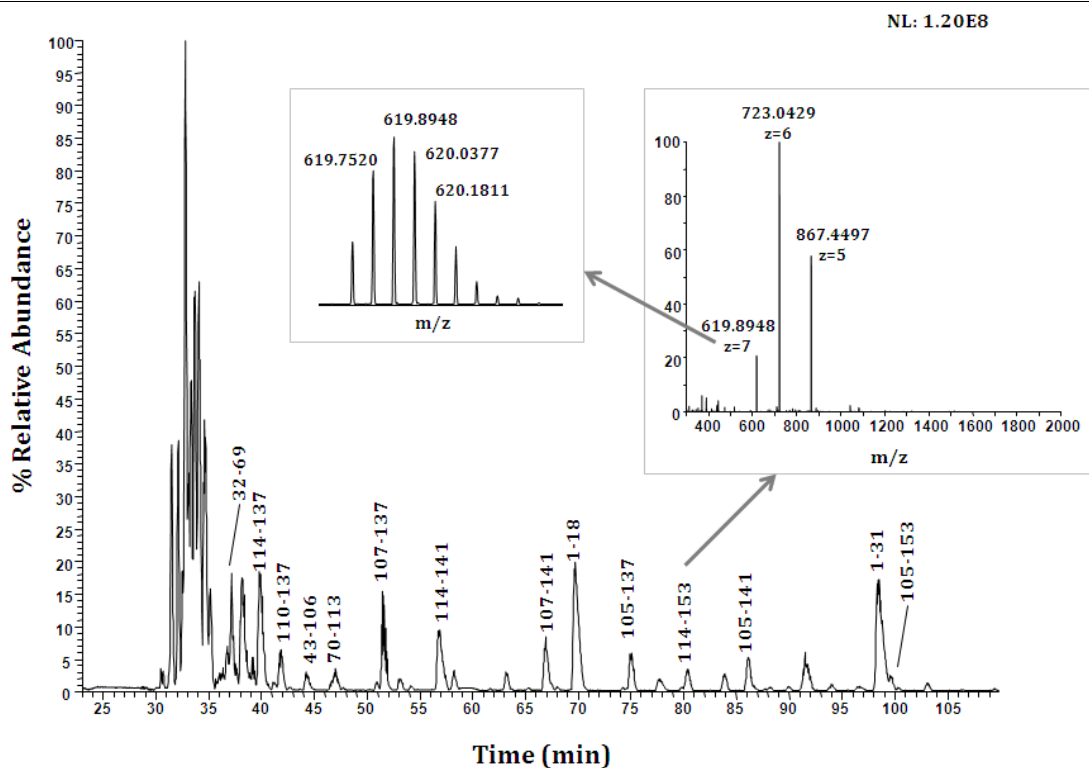
**Figure 4. 3 Comparison of different times for size-controlled digestion of apomyoglobin.**

Three chromatograms are shown with their total ion counts (TIC) as the measurement of the relative abundances. The digestion time and normalized ion count (NL) are labeled on the right of each chromatogram. Peptides with different sizes are labeled within the chromatograms.

In experiment #1, MS1 of each base peak gave the  $m/z$  information of the corresponding peptide (**Figure 4.4**). This information, along with the elution time for each base peak peptide, was later used in experiment # 2, where multiple MS/MS segments were set during the entire HPLC gradient (the same gradient as in experiment #1). In this way targeted MS2 analysis could be performed for each base peak peptide as the peptide eluted off the HPLC column and electrosprayed into the mass spectrometer.

The large peptide with “golden size” typically appears as a series of highly charged ions with a Gaussian-like distribution of different charge states (**Figure 4.4**). As high charge state ions favor ETD compared to low charge state ions, to achieve efficient ETD, typically the peptide ion with the highest charge state and the intensity above  $2 \times 10^6$  was chosen for ETD MS2 (44). In the example shown in **Figure 4.4**, MS1 spectrum was shown for the peak corresponding to apomyoglobin 114-153, in which three ions were present with different charge states. The ion with  $z=+7$  was chosen for ETD MS2, although it was not the most abundant. The  $m/z$  of the selected precursor ion should fall in the center of the isotopic peaks. As shown in **Figure 4.4**, to target the +7 species of peptide 114-153,  $m/z$  of 620.2 was chosen as the center of precursor ion with  $\pm 1.5$   $m/z$  isolation window.

---



**Figure 4. 4** Base peak chromatogram of apomyoglobin with 0.77 s of size-controlled digestion time.

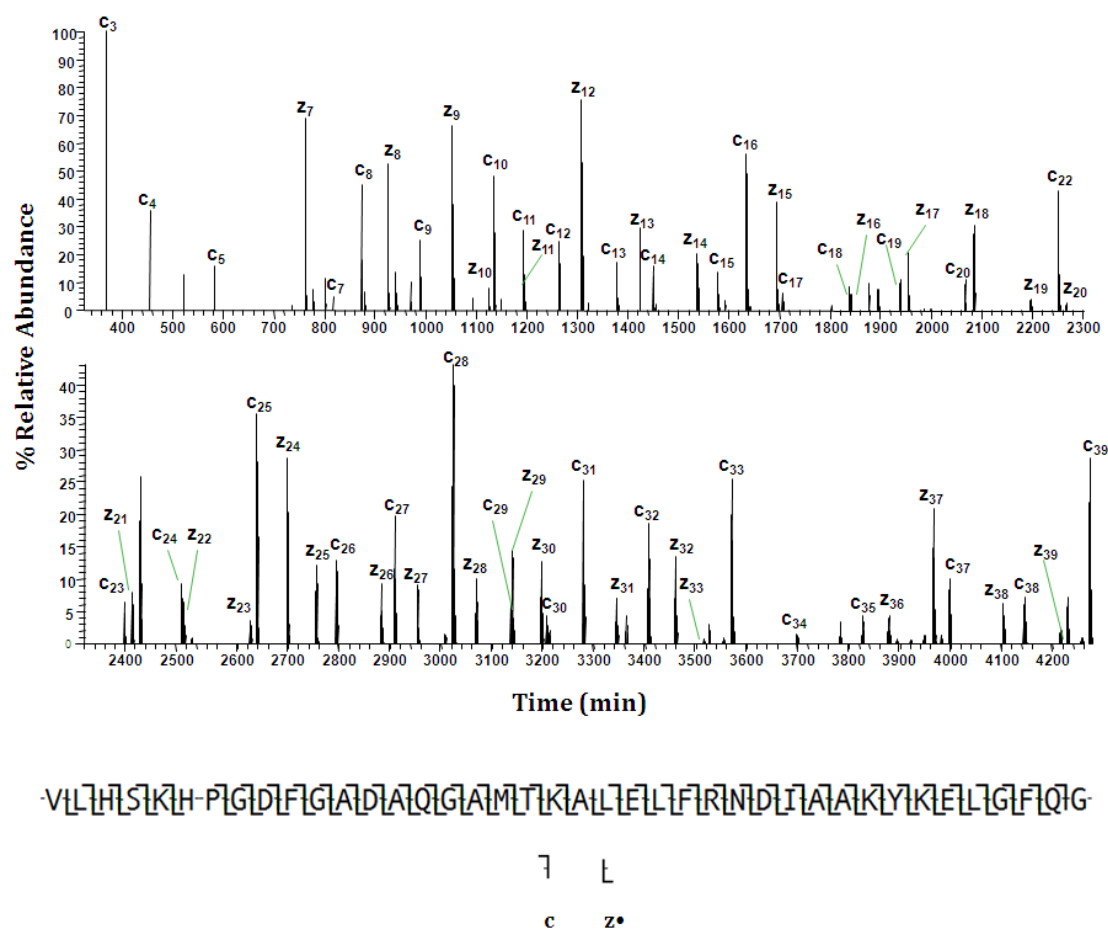
Peptides with proper sizes for middle-down analysis are labeled with the apomyoglobin sequence number based on MS/MS in experiment #2. The MS1 spectrum of peptide 114-153 and the isotopic distribution of the +7 species of peptide 114-153 are shown as the arrows indicate.

ETD reaction time was set inversely proportional to the square of the charge state of a peptide (45, 46). One advantage of performing targeted ETD in different MS/MS segments is that appropriate ETD reaction times can be set for different peptides with selected charge states. Based on this rule and experimental experience, different ETD reaction times used for different charge states are listed in **Table 4.1**. For example, ETD reaction time of 9.2 s was set for peptide 114-153 with charge state of +7.

**Table 4. 1** ETD reaction times used for peptides with different charge states.

Charge State	3	4	5	6	7	8	9	10	11	12
ETD Reaction Time (ms)	50.0	28.1	18.0	12.5	9.2	7.0	5.6	4.5	3.7	3.1

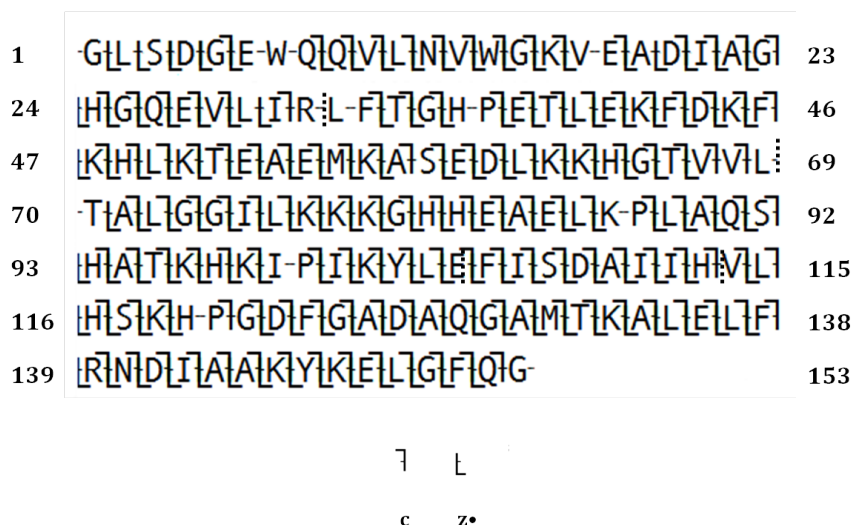
**Figure 4.5** shows the ETD MS2 spectrum of peptide 114-153 after converting the original MS2 fragment ions to charge state of +1. Regardless of the non-cleavable N-terminal side of Pro, ETD yielded 100% sequence of this peptide.



**Figure 4. 5 ETD MS2 spectrum of apomyoglobin peptide 114-153 and the sequence coverage.**

All the ions in the original MS2 spectrum were converted to +1 charge state by Xcalibur Xtract with c and z• fragments labeled. Under the spectrum the sequence coverage by c and z• fragment ions was assigned using ProSite PC based on the original MS2 spectrum and manually verified. More ions were labeled in the peptide sequence than in the spectrum because some fragment ions were not successfully converted by Xcalibur Xtract.

Among all the base peak peptides targeted in experiment #2 for MS/MS, four of them, peptides 1-31, 32-69, 70-113, and 114-153, can compose the whole sequence of apomyoglobin. Peptide 70-113 has two missing ETD cleavage sites; however, these two sites were covered by ETD fragments from another peptide, 105-153. **Figure 4.6** shows the total sequence coverage of apomyoglobin from the five peptides by ETD. Of all the 152 sites, 142 sites were covered by ETD, 4 sites were enzymatically cleaved (2 sites were also covered by ETD), and 4 sites were N-terminal to Pro. Combining both ETD and enzymatic cleavages, the sequence coverage is 97%.



**Figure 4. 6 Apomyoglobin sequence coverage with ETD.**

Dashed lines are enzymatic cleavage sites.

**Figure 4.7** shows the quantitation of the four apomyoglobin peptides that represent the whole myoglobin sequence using peak area. The peak area is the

integration of the normalized ion counts for each peptide, summarizing all the isotopic peaks of different charge states. The most abundant peptide was about three times of the least abundant peptide, suggesting a roughly equal high efficiency of protease cleavage along the apomyoglobin protein chain. This is important for comprehensive mapping of protein sequences.

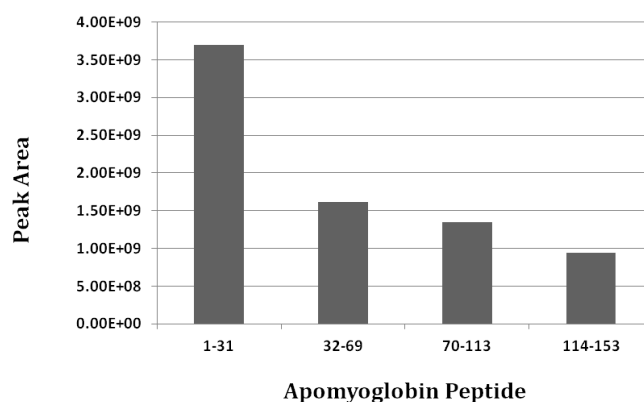


Figure 4. 7 Quantitation of apomyoglobin peptides based on normalized ion counts.

### 4.5.3 Application of Size-Controlled Digestion for Monoclonal Antibody Sequencing and Post-translational Modifications Characterization

#### 4.5.3.1 High Concentration of Urea Denatures Monoclonal Antibody

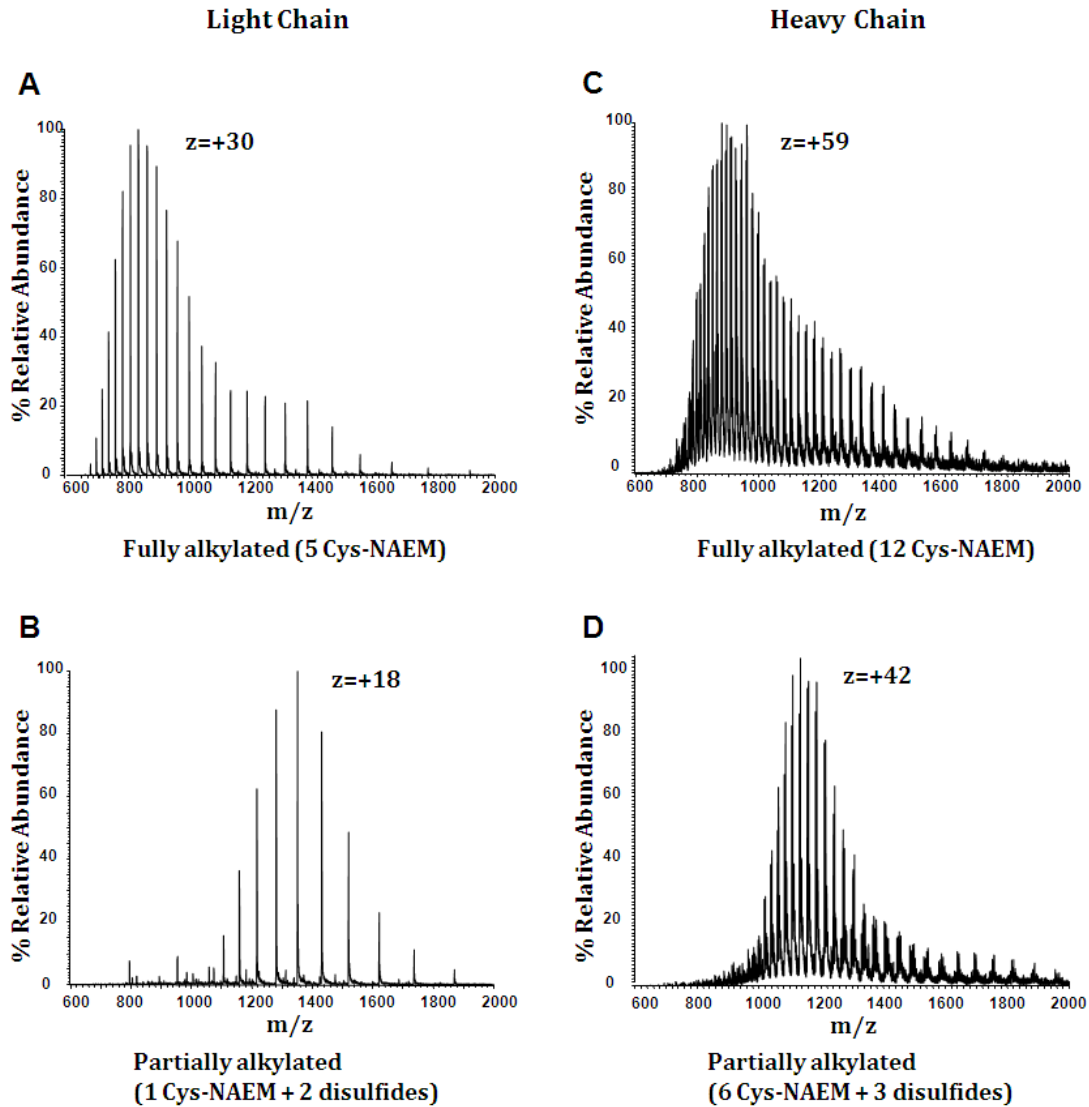
As described in the Introduction of this chapter, efficient proteolytic digestion is favored when proteins are fully denatured, so that the cleavage sites are readily accessible to proteases. In addition, protein denaturation favors complete reduction and alkylation of disulfides within compact protein structure. Complete reduction/alkylation of protein Cys residues favors protein unfolding and facilitates

---



the solvation of some of the residues that are previously shielded or buried inside the protein secondary/tertiary structure due to disulfide bonds. Due to the above reasons, the fully reduced/alkylated proteins/peptides could carry more charges than partially reduced/alkylated ones. **Figure 4.8** shows the comparison of the charge states of mAb light chain and heavy chain after complete and partial alkylation of Cys. Based on the expected sequence of the mAb in this work, LC contains 5 Cys and HC contains 12 Cys. When only one Cys was alkylated in LC, the most abundant charge state was +18. However, the most abundant charge state of fully alkylated LC increased to +30. Similarly, the most abundant charge state of fully alkylated HC was +59, while it was +42 when half of the Cys were alkylated.

---



**Figure 4. 8** Comparison of the most abundant charge states of fully and partially alkylated mAb light chain and heavy chain.

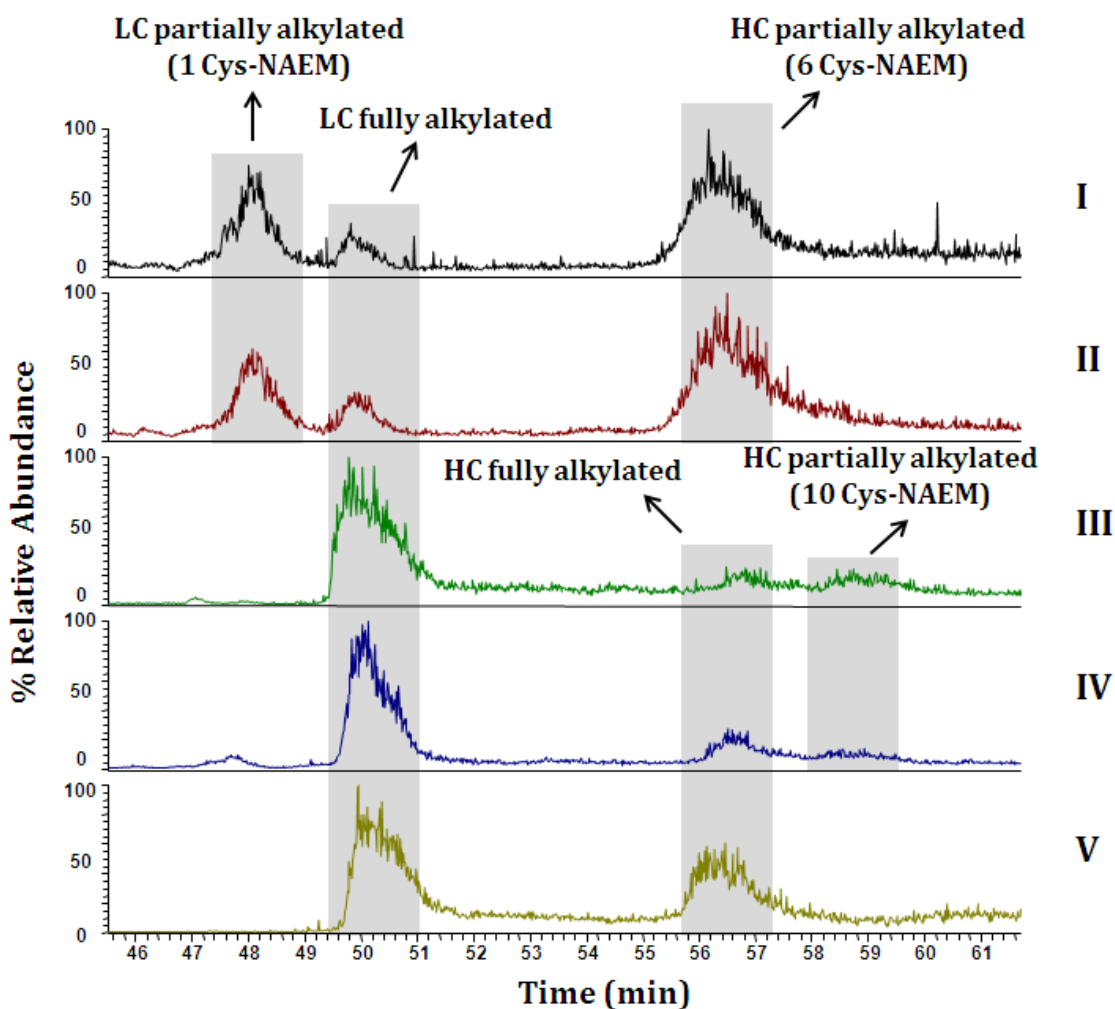
The most abundant charge state for each species is labeled. Aminoethylmaleimide (NAEM) is the alkylation reagent that is further discussed in Section 4.5.3.4.

In this work, various denaturing conditions considering heat, pH, salt, and chaotropic reagent were examined for the reduction and alkylation of Waters standard mAb, as shown in **Figure 4.9**. Maleimide derivatives (further discussed in Section 4.5.3.4) are highly efficient Cys alkylation reagents at pH 6.5-7.5 and have been used to quantify Cys quantitation; therefore, the partial alkylation should be owing to insufficient protein denaturation instead of slow reaction kinetics (47). Of all the denaturing conditions, 8 M chaotropic reagent urea was found to play a key role in fully denaturation and alkylation of LC and HC. Therefore, condition V as shown in **Figure 4.9** was utilized as the optimal denaturing condition for Waters standard mAb in this work.

Although 8 M urea allowed for the fully denaturation of Water standard mAb during reduction and alkylation, our initial studies showed that conventional in-solution digestion of fully alkylated mAb (reduction/alkylation performed in a solution containing 8 M urea) using Lys-C and Asp-N in 1.6 M urea was unsatisfactory or unsuccessful (data shown and discussed in Section 4.5.5). This is probably because that the mAb LCs and HCs became less denatured in lower concentration of urea. Therefore it would be ideal to perform the digestion in 8 M urea so that the mAb LCs and HCs can be maintained in denatured state. However, 8 M urea could denature most commonly used proteases as well. Fortunately, for aspergillopepsin I, it has been reported that the degree of inactivation in the presence of 5 M urea of an apparent pH of 5.5 was 46% after 30 minutes and 59%

---

after 1 hour's incubation (40). In this work, no noticeable drop of enzyme activity was found after 1 hour of continuous mAb digestion using immobilized aspergillopepsin I in 8 M urea at pH 3.9.



**Figure 4. 9 Chromatograms of Waters standard mAb LC and HC resulting from various denaturing conditions for Cys reduction and alkylation.**

The resulting LC and HC (shaded in grey) with different alkylated forms under different denaturing conditions are shown in the chromatograms. Aminoethylmaleimide (NAEM) is the alkylation reagent.

Denaturing conditions were:

- I. Reduction: 2 mM dithiothreitol/6 M urea/0.5 M  $\text{NH}_4\text{Ac}$  (pH 6.8) at 37°C for 30 minutes.  
Alkylation: 10 mM NAEM/6 M urea/ 0.5 M  $\text{NH}_4\text{Ac}$  (pH 6.8) at room temperature for 10 minutes.
- II. Reduction: 5 mM TECP-HCl/6 M urea/0.5 M  $\text{NH}_4\text{Ac}$  (pH 6.8) at 37°C for 30 minutes.  
Alkylation: 10 mM NAEM/6 M urea/ 0.5 M  $\text{NH}_4\text{Ac}$  (pH 6.8) at room temperature for 10 minutes.

- III. Reduction: 2 mM TECP-HCl/water (pH 2.5) at 80°C for 10 minutes.  
Alkylation: 10 mM NAEM/6 M urea/ 0.5 M NH<sub>4</sub>Ac (pH 6.8) at room temperature for 10 minutes.
- IV. Reduction: 10 mM TECP-HCl/0.1% AcOH (pH 2) at 80°C for 30 minutes.  
Alkylation: 10 mM NAEM/6 M urea/ 0.5 M NH<sub>4</sub>Ac (pH 6.8) at room temperature for 10 minutes.
- V. Reduction: 10 mM TECP-HCl/0.1% AcOH/8 M urea (pH 4) at 50°C for 10 minutes.  
Alkylation: 10 mM NAEM/8 M urea/ 0.5 M NH<sub>4</sub>Ac (pH 6.8) at room temperature for 10 minutes.

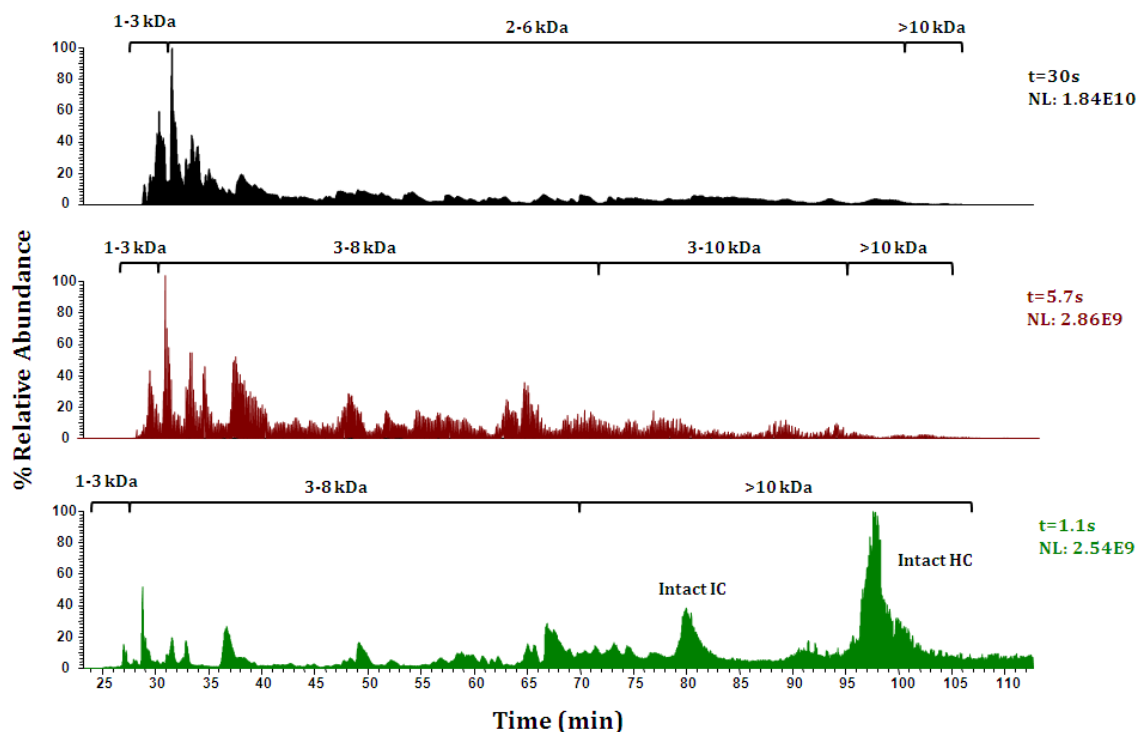
#### 4.5.3.2 Optimization of Size-Controlled Digestion Time of Monoclonal

##### Antibody

The size-controlled digestion method was applied to the sequencing and PTMs characterization of Waters standard mAb. MAbs have much larger size and more complex tertiary structure than apomyoglobin; therefore the digestion time for mAb needs to be optimized in the same way as for apomyoglobin. **Figure 4.10** showed the comparison of size-controlled digestions of Waters standard mAb with three different digestion times. When the digestion time (i.e. protein residence time in enzyme-column) was 30 s, peptides with molecular weight 2-6 kDa were dominant through 30-100 minute of the gradient. Peptides of this size range are of low efficiency for mAb sequencing because too many peptides need to be analyzed and each covers only a small portion of the protein. Decreasing the digestion time to 5.7 s resulted in the generation of mainly “golden-sized”. However, further shortening the digestion time to 1.1 s resulted in undigested and partially digested mAb light LCs and HCs near the end of the gradient, suggesting the low digestion efficiency due to the short digestion time. Therefore, 5.70 s was chosen as the optimized digestion time for middle-down analysis of Waters standard mAb. It is noteworthy that, the optimal digestion time of mAb for obtaining “golden size”

---

peptides is a much longer than that for apomyoglobin. This is not surprising considering that mAb is a much larger protein than myoglobin (150 kDa vs. 17 kDa). Even considering that mAb forms two 25 kDa LC and two 50 kDa HC after reduction and alkylation, the two polypeptides are still much larger than apomyoglobin. However, other factors such as the number of cleavable amino acid sites and the undisrupted secondary/tertiary structure of protein in 8 M urea should also affect the digestion efficiency.



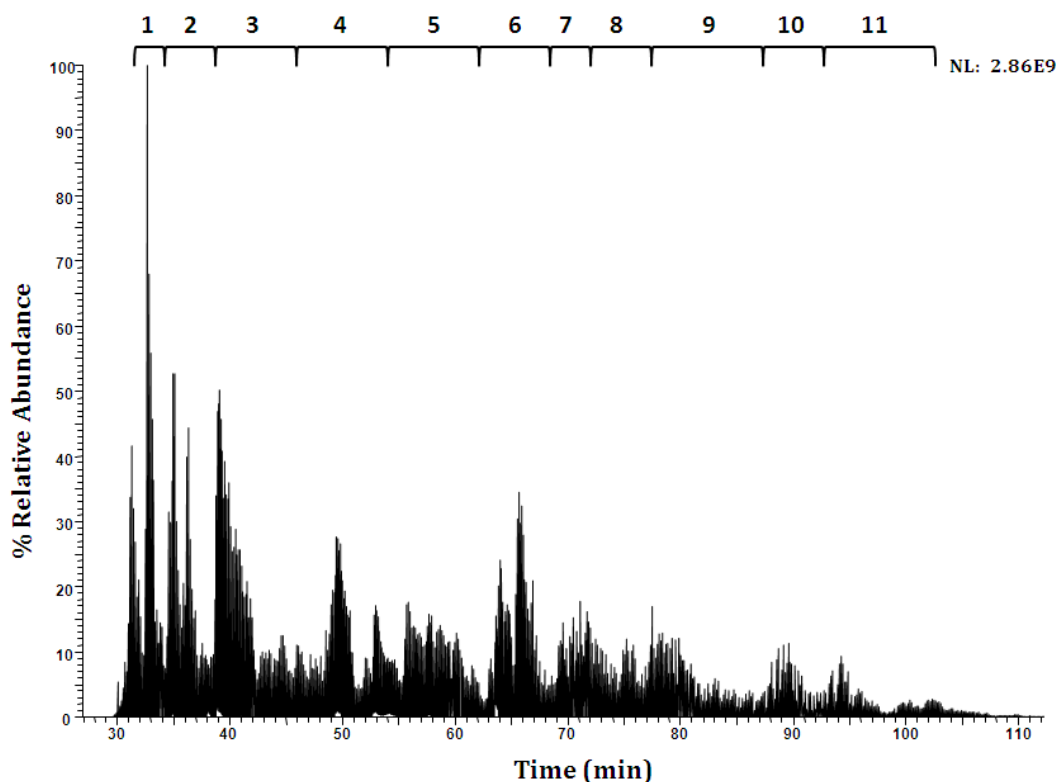
**Figure 4. 10 Comparison of different side-controlled digestion times of Waters standard mAb.**

Three chromatograms are shown with TICs as the measurement of the relative abundances. The digestion time and normalized ion count (NL) are labeled on the right of each chromatogram. Peptides with different sizes are labeled within the chromatograms.

### 4.5.3.3 Monoclonal Antibody Sequencing and Post-translational Modifications

#### Characterization

Next, based on an analysis of the peptides composition at different elution time periods, the gradient where peptides elute was divided into 11 segments as shown in **Figure 4.11**. In each segment the 1-6 most abundant peptides were targeted for MS2 analysis (depending on the complexity of peptides and the duration of that segment). ETD and CAD were utilized in two separate runs to general complimentary spectra to facilitate peptide sequencing.



**Figure 4. 11** TIC chromatogram of Waters standard mAb with 5.7 s of size-controlled digestion time.

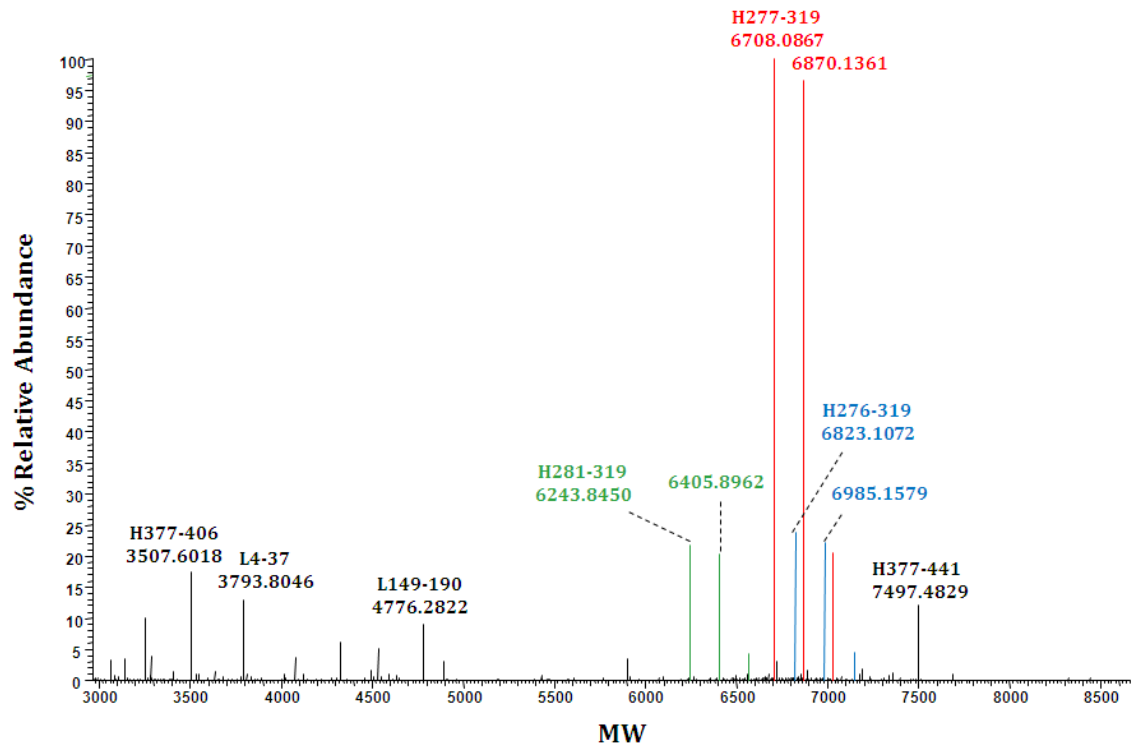
The gradient was divided into 11 segments for the following targeted MS2 analysis.

---

**Figure 4.12** shows the averaged MS1 spectrum (all the ions were converted to their monoisotopic molecular weights) of segment 3 in the gradient. The major peptides with size 3-8 kDa eluted in this segment were labeled. The three groups of colored species are peptides containing N-linked glycans. For example, the three red peaks appear to come from the same peptide but with different glycan forms. This is because the mass difference between two neighboring peaks is 162 Da, corresponding to the mass of a hexose. **Figure 4.13** shows the ETD spectrum of the most abundant glycopeptide (average molecular weight 6708.0867 Da) and its identity was confirmed to be HC299-319. From the mass differences of fragment ion  $c_{15}$  and  $c_{16}$ , and of  $z_{27}$  and  $z_{28}$ , it can be calculated that the modified site was Asn 292 and the monoisotope mass of the glycan is 1444.5338 Da, corresponding to the most common N-linked glycan form on Fc of mAbs, G0F, as shown in **Figure 4.14**. The second peak corresponds to the glycan form G1F and the third peak corresponds to G2F. This Asn site is consistent with the known antibody Fc N-glycan motif Asn-X-Ser/Thr where X can be any amino acid except for Pro (48).

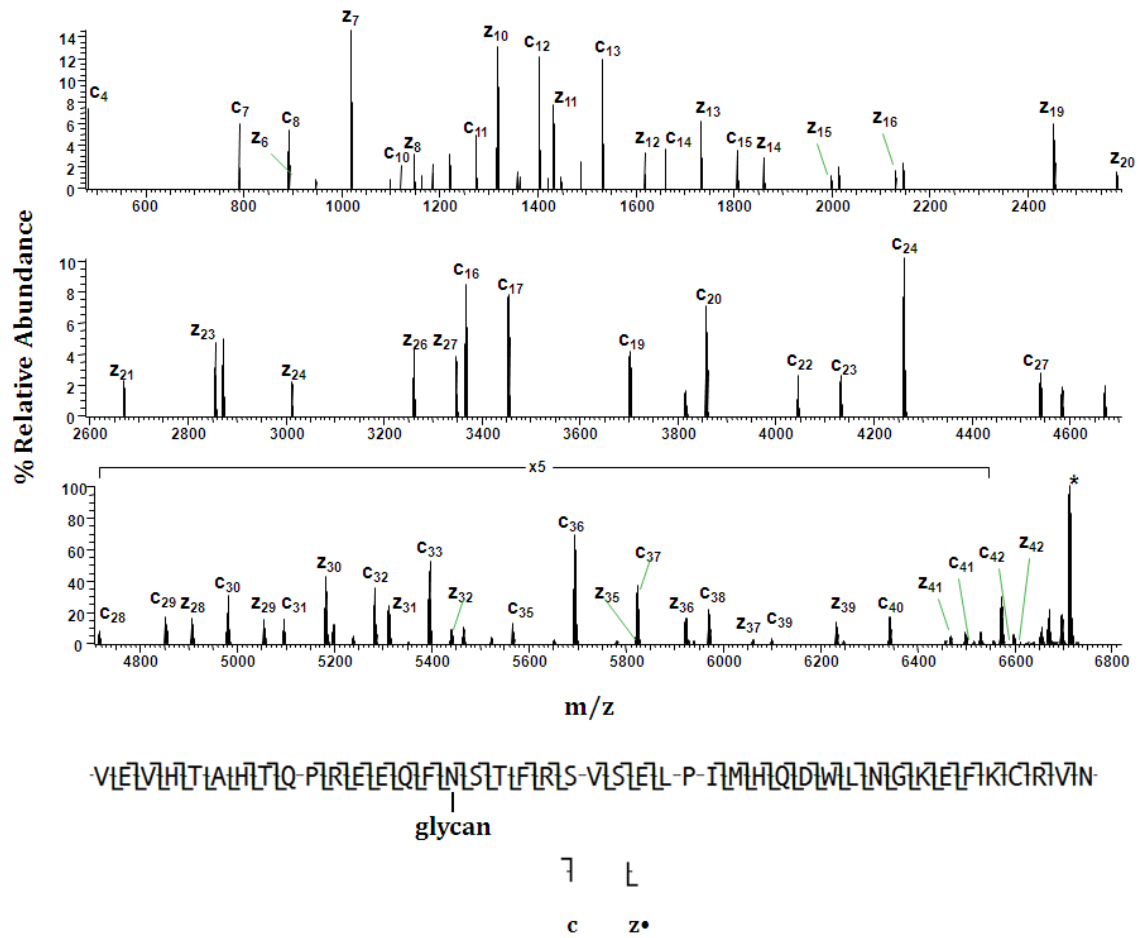
---





**Figure 4. 12 Averaged MS1 spectrum of LC gradient segment 3 from Waters standard mAb with 5.7 s of size-controlled digestion.**

All the ions were converted to their monoisotopic molecular weights (MWs) using Xcalibur Xtract. The sequences for several abundant peptides were labeled based on MS2 data. The three groups of colored species are peptides containing N-linked glycans.



**Figure 4. 13 ETD MS2 spectrum of Waters standard mAb HC 277-319 and Sequence Coverage.**

All the ions in the original MS2 spectrum were converted to +1 charge state by Xcalibur Xtract with c and z• fragments labeled. Under the spectrum the sequence coverage was shown in which the assignment of c and z• fragments was facilitated using ProSite PC based on the original MS2 spectrum and manually verified. More ions were labeled in the peptide sequence than in the spectrum because some real fragment ions were lost in Xcalibur Xtract deconvolution. The N-glycan Asn site is labeled in the sequence.

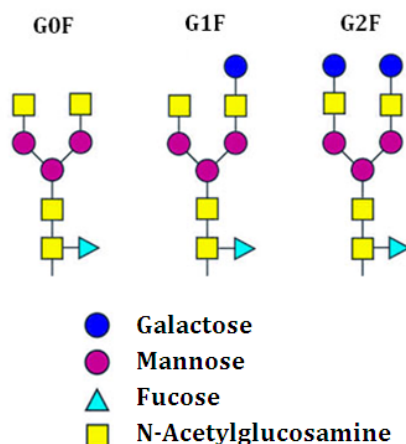


Figure 4. 14 The most common N-linked glycans on the Fc of mAbs.

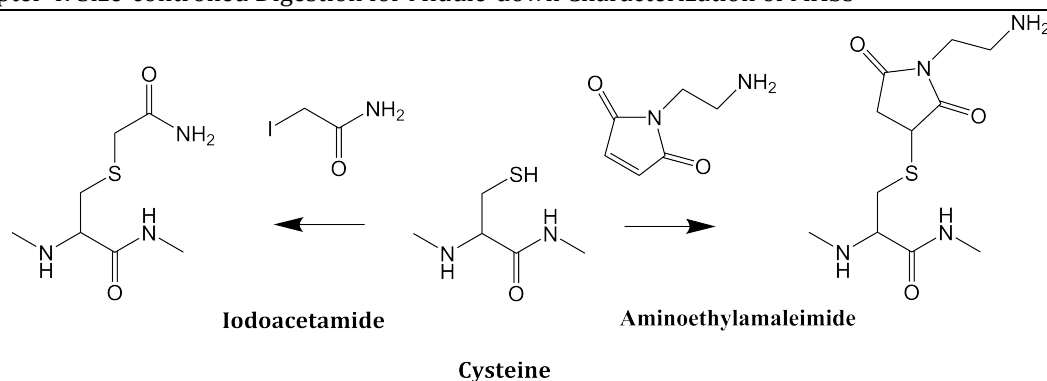
#### 4.5.3.4 Charge Enhancement of Cys-containing Peptides to Improve ETD

As introduced in Chapter 1, ETD fragmentation performs most efficiently on highly charged peptides ( $z \geq 3$ ). Moreover, since charges are neutralized by the transferred electrons, adequate remaining charges must be available to the resulting fragments to be visible to the mass spectrometer. To enhance ETD fragmentation, one of the solutions is to modify the peptide in a manner that increases the positive charges. Since mAbs contains multiple cysteines that require alkylation, appropriate alkylating reagents could be utilized as modifiers to enhance peptide charge. In addition, it was reported that CDR1 and CDR3 of both LC and HC are in close approximation to Cys (49). Therefore, modifying Cys using reagents containing basic groups can facilitate transferring proton to the amino acids nearby Cys, which may enhance ETD of peptides containing CDR sequence. As shown in **Figure 4.15**, the conventional Cys alkylation reagent iodoacetamine cannot be

protonated and carry charges. In this study, a new alkylating reagent, aminoethylmaleimide (NAEM) was employed to derivatize Cys. The primary amine of the derivatized side chain can be protonated and add charges to Cys residues. It should be noted that both NAEM reagent and the maleimide group in the derivatized Cys undergo prominent hydrolysis if the reaction time is over 30 minutes at neutral condition. Basic condition can speed up this process and cause side reactions. Therefore, the reaction was performed at pH 6.8 for only 10 minutes and acidified to pH 3.9 for the following digestion. This condition resulted in quantitative yield of alkylation of LC and HC Cys restudies without noticeable hydrolysis, as shown in **Figure 4.8 A and C**. The ETD fragmentation was improved significantly as exemplified by the sequence coverage of peptide LC1-52 shown in **Figure 4.16**.

For some large peptides such as LC1-52, however, even after enhancing its charge state by Cys-NAEM, the ETD fragments were still not enough to cover the whole peptide sequence. Therefore CAD was utilized as a complementary method to improve the sequence coverage as shown in **Figure 4.16**.

---



**Figure 4. 15** Alkylation of cysteine side chain using iodoacetamide and aminoethylmaleimide.

Cys modified with iodoacetamide, ETD					
D	V	L	M	T	Q
L	T	P	L	S	L
P	V	S	L	G	D
A	S	I	S	C	R
S	S	Q	Y	I	V
H	S	I	N	G	N
T	L	E	L	W	Y
L	Q	K	P	G	L
S	P	K	L	L	
Cys modified with aminoethylmaleimide, ETD					
D	V	L	M	T	Q
L	T	P	L	S	L
P	V	S	L	G	D
A	S	I	S	C	R
S	S	Q	Y	I	V
H	S	I	N	G	N
T	L	E	L	W	Y
L	Q	K	P	G	L
S	P	K	L	L	
Cys modified with aminoethylmaleimide, ETD and CAD					
D	V	L	M	T	Q
L	T	P	L	S	L
P	V	S	L	G	D
A	S	I	S	C	R
S	S	Q	Y	I	V
H	S	I	N	G	N
T	L	E	L	W	Y
L	Q	K	P	G	L
S	P	K	L	L	
<div style="display: flex; justify-content: space-around; align-items: center;"> <div style="text-align: center;"> <math>\downarrow</math> b         </div> <div style="text-align: center;"> <math>\downarrow</math> y         </div> <div style="text-align: center;"> <math>\downarrow</math> c         </div> <div style="text-align: center;"> <math>\downarrow</math> z•         </div> <div style="text-align: center;"> <math>\downarrow</math> b, y         </div> <div style="text-align: center;"> <math>\downarrow</math> c, z•         </div> </div>					

**Figure 4. 16** Sequence coverage of Waters standard mAb peptide LC 1-52.

#### 4.5.3.5 Summary of Total Sequence Coverage and Identified PTMs of Waters Standard MAb

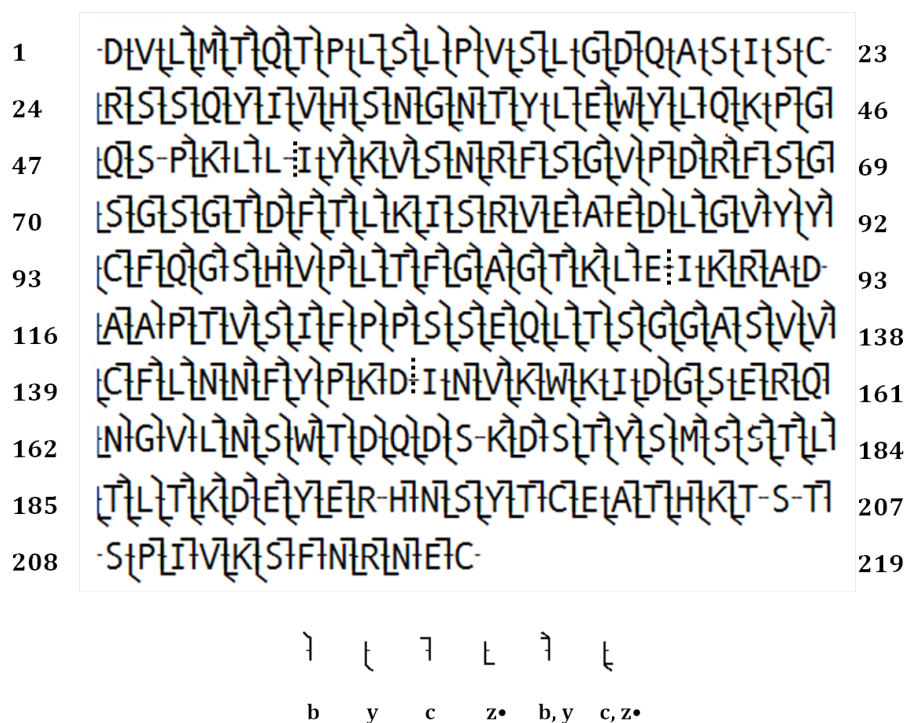
**Figure 4.17** shows the total sequence coverage of Waters standard mAb LC from 6 peptides. Of the total 218 cleavable sites, 210 sites were covered by either

ETD or CAD, or both, three of the uncovered sites were covered by enzymatic

cleavage. The total sequence coverage for LC was 98%.

**Figure 4.18** shows the total sequence coverage of Waters standard mAb HC from 14 peptides. Of the total 440 cleavable sites, 405 sites were covered by either ETD or CAD, or both, eight of the uncovered sites were covered by enzymatic cleavage. The total sequence coverage for HC was 94%.

**Table 4.2** shows a summary of all the PTMs identified on Waters standard mAb.



**Figure 4. 17** Waters standard mAb LC sequence coverage with ETD and CAD.

Dashed lines are enzymatic cleavage sites.

1	QIVQLKESGPIGLVAPSQSLSIITCTI	23
24	VSIGFSLLGYGIVNIWVLRQPPGQGLLEI	46
47	WLMGIWGDGSLTDYNSAILKSRIS-I	69
70	TKDINSKSLQVFLKMIN:SLQLTDIDT-AKI	92
93	LYYC-TIRA-PYGLKQLYFALYWGLQGLTILVTI	115
116	VSIAAKITITPPISVYPLIAPIGLSAAQTIN	138
139	SMIVTLLGLCL-VIK:GLYFPIEP-VITVITWINLS	161
162	GLSLSLSGLVHITFPALVLLQSDLLYTLLSS	184
185	SLVITVPISSTTWIPSEIT-V-TCTNVIAHPIAS	207
208	LS-TIK:VDKIKILVPIRDCLGCLK-PC-ICTIVPI	230
231	ELVLSLVFIIFPPKIPKIDVLITITLITIPKI	253
254	VITCTIVVID:ISTKDDPEVQFSLWFVIDD	276
277	:VLELVHITAHITQ-PRLEELQFLNLSLTFRLS-VLS	299
300	ELL-P-IMHQLDWLLNGKLEFKICRIVN-:LAIA	322
323	FPAPLIEKTLITSKITKGRIPKLA-PQLVYT	345
346	IPPIPKLEQMAKDKLVISLITCMITIDFIF	368
369	PEID:ITLVLEWQWNGQPLAEINLYKINTIQPI	391
392	IMIDITDGLSYFVYYS-KLNLVQKLSNWEIA	414
415	GNITFTICTSLVILHLEGLHNLHHTLEKSLTS	437
438	HIS-PIG	441

}    {    7    L    7    {  
 b    y    c    z•    b, y    c, z•

Figure 4. 18 Waters standard mAb HC sequence coverage with ETD and CAD.

Dashed lines are enzymatic cleavage sites.

**Table 4. 2 A summary of PTMs identified on Waters standard mAb.**

PTM	Site	Peptide
Oxidation	LC Met4	LC1-52
Pyru glutamate Acid Formation	HC N-terminus Gln	HC1-36
Amidation	HC Asp138	HC84-148
N-linked Glycosylation	HC Asn292	HC277-319
Oxidation	HC Met304	HC277-319

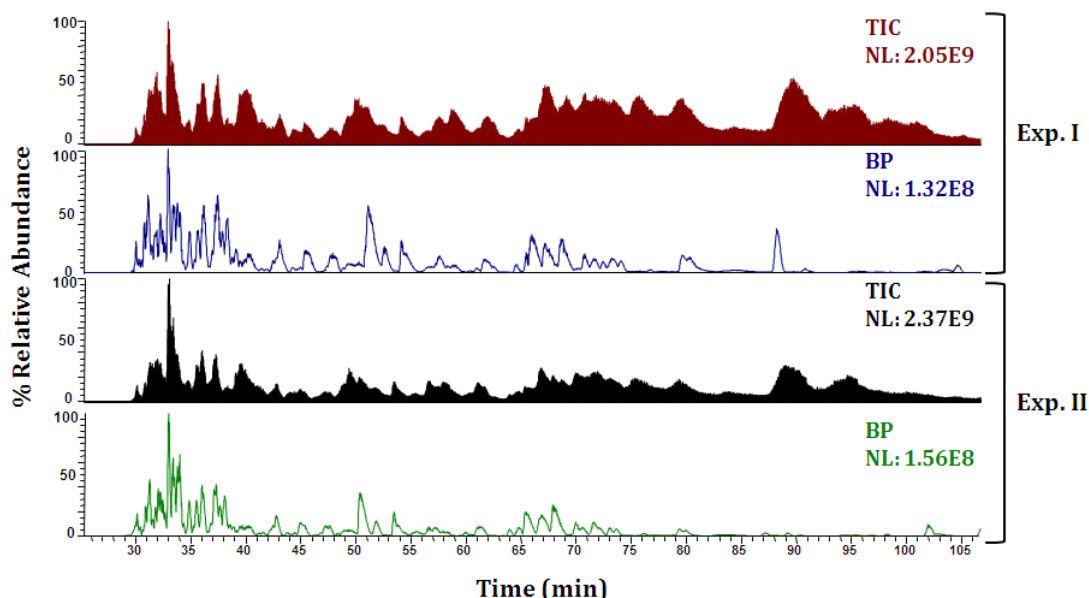
#### 4.5.4 Repeatability of Size-controlled Digestion of Monoclonal Antibody

The repeatability of the size-controlled digestion method was examined by fabricating enzyme-columns using different enzyme-beads (prepared by two different persons using the same procedures) with similar  $L_{packed}$ , performing on-column digestions with the same digestion times (5.7 s), and collecting mAb digests in two separate experiments five months apart, and then performing the same HPLC-MS/MS analysis using the same HPLC column and MS instrument on the same day. **Figure 4.19** shows the TIC and base peak (BP) chromatograms of the two experiments. The BP chromatograms showed that the most abundant peptides eluted at various times during the gradient were highly identical between the two experiments, suggesting that the digestion patterns using the two enzyme-columns were very similar. The TIC chromatograms of the two experiments also show very similar profiles, further proving that the peptides, especially those with high molecular weight, were very similar in the two experiments. In addition, comparing the NLs of the TIC and BP chromatograms can provide relative quantitation of the

---



peptides generated in the two experiments. As shown in **Figure 4.19**, the NLs of TIC chromatograms are 2.09E9 and 2.37 E9, for Exp. I and Exp. II, respectively; and the NLs of base peak chromatograms are 1.32E8 and 1.56E8, for Exp. I and Exp. II, respectively. These comparisons showed good repeatability of the size-controlled digestion in both digestion pattern and peptide yield.



**Figure 4. 19** TIC and BP chromatograms of two parallel size-control digestion and HPLC-MS/MS experiments.

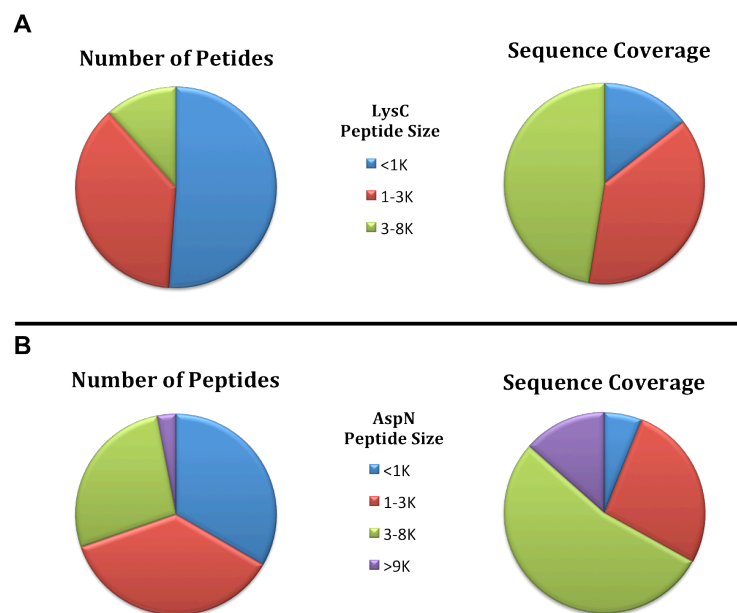
TIC: total ion count. BP: base peak. NL: normalized ion count.

#### 4.5.5 Comparison of Size-Control Digestion to Conventional In-solution Digest Using Other Proteases

To compare the size-controlled digestion method with conventional in-solution digestion methods for middle-down analysis, *in silico* analysis of LysC and AspN digestion of Waters standard mAb was performed at first. *In silico* digestion of

the Waters standard mAb generated 45 and 33 peptides with Lys and AspN,

respectively. In contrast, to represent the entire sequence with size-control digestion, only 13 peptides were required. Moreover, to achieve the sequence coverage (98% for LC and 94% for HC) reported in Section 4.5.3.5, only 20 peptides were utilized. **Figure 4.20** shows the number of different sizes of *in silico* generated peptides with LysC and AspN, and the percentage of sequence that these peptides can cover. The number of the theoretical peptides in the 3-8 kDa “golden size” range was no more than a quarter of the total number of theoretical peptides, and these “golden size” peptides only cover half of the sequence of the entire Waters standard mAb. These data suggest that with LysC or AspN digestion it was less likely to achieve a sequence coverage as high as the size-control digestion method.

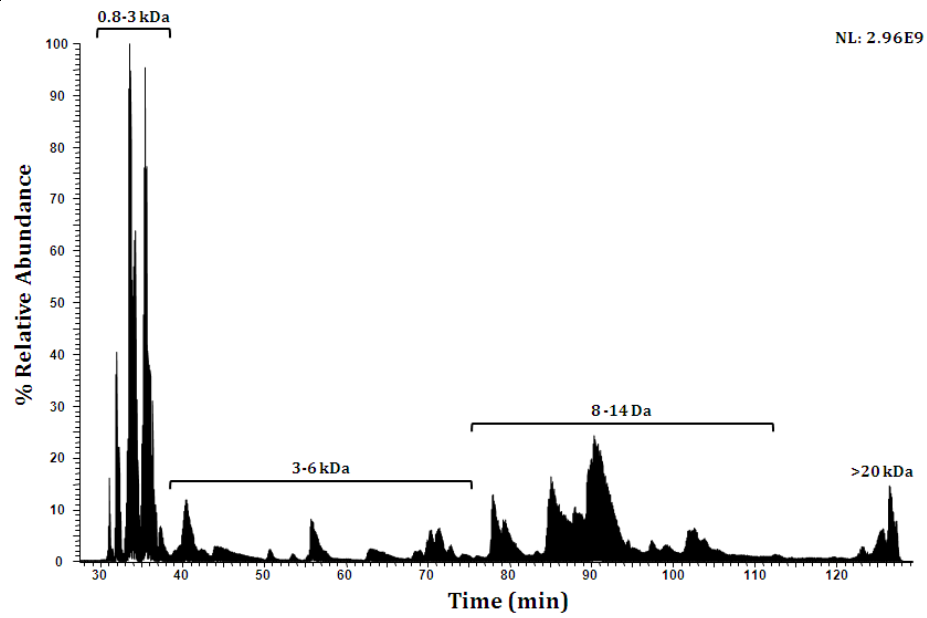


**Figure 4. 20** *In silico* digestion of Waters standard mAb with LysC and AspN

To future test this prediction, LysC and AspN digestion and LC-MS/MS

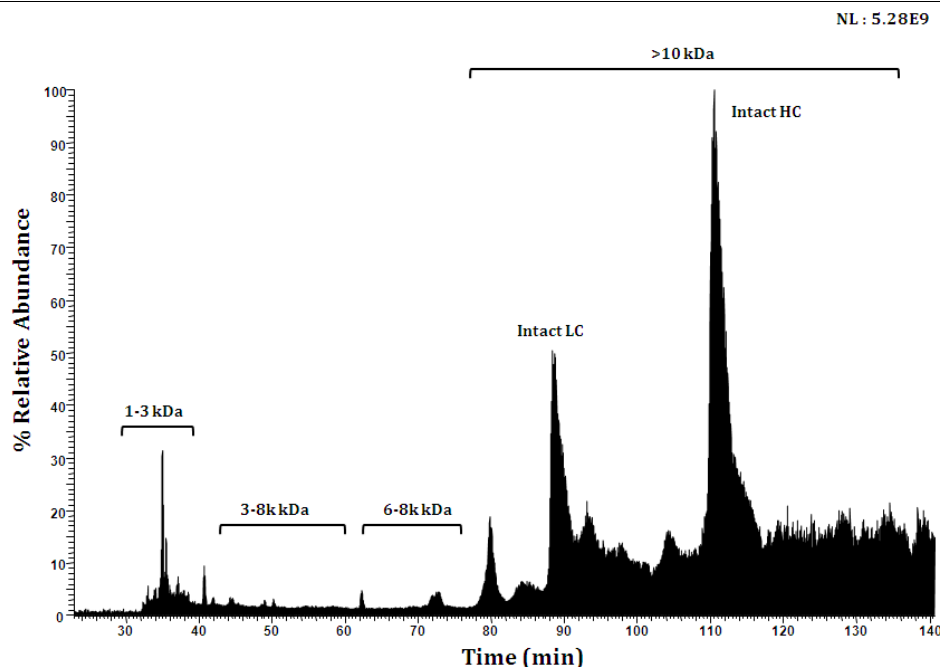
analysis were performed. Since LysC and AspN work at pH 8 condition that hydrolyzes the NAEM derivative quickly, mAbs were completely derivatized with iodoacetamide under 8 M urea in 100 mM  $\text{NH}_4\text{HCO}_3$  prior to digestion (confirmed with MS1, data not shown). To avoid the deactivation of LysC and AspN, the sample was diluted 5 fold to reduce the concentration of urea to 1.6 M for digestion. **Figure 4.21** and **4.22** show the chromatograms of peptides resulting from 10 hours of LysC and AspN digestion, respectively. It is obvious that although the digestion lasted for 10 hours, there were undigested or partially digested protein present in the sample. The low efficiency was mostly due to the use of 1.6 M urea, which was not sufficient for mAb denaturation or led to enzyme deactivation. Differently, the immobilized aspergillopepsin I for size-controlled digestion was able to maintain a consistent, reasonable activity in 8M urea and digest the denatured mAbs into mainly “golden-size” peptides.

---



**Figure 4. 21** TIC chromatogram of LysC in-solution digestion of reduced and alkylated Waters standard mAb.

Peptides with different sizes are labeled within the chromatogram. NL: normalized ion count.



**Figure 4. 22** TIC chromatogram of AspN in-solution digestion of reduced and alkylated Waters standard mAb.

Peptides with different sizes are labeled within the chromatogram. NL: normalized ion count.

## 4.6 Conclusions and Future Directions

This chapter presents a novel strategy for the generation of large peptides for middle-down MS analysis of proteins using size-controlled digestion. This strategy utilizes a column packed with porous beads modified with a low-specificity enzyme, aspergillopepsin I. Adjusting the flow rate of the substrate proteins when they flowed through the enzyme-column allowed for controlling the size of the resulting peptides. This method can digest proteins in their fully denatured state in 8 M urea at acidic condition. Selecting a proper digestion time (probably in the range of 0.1-10 s) leads to the generation of mainly 3-8 kDa “golden-sized” peptides that favor peptide sequencing when coupled with nanoLC-MS/MS. This strategy breaks the

limit of conventional in-solution protein digestion that solely relies on enzyme

specificity and extends the digestion condition to 8 M urea that favors unfolding of many compact proteins. In addition, the employment of aminoethylmaleimide as a new Cys alkylating reagent enhanced the charge states of peptides containing Cys and improved ETD MS2. This strategy shows superior ability in digesting mAb into 3-8 kDa peptides compared to in-solution digestion by LysC and AspN, and yields 98% sequence coverage for mAb LC (25 kDa) and 94% for mAb HC (50 kDa). Moreover, PTMs on mAbs including pyroglutamic acid formation, oxidation, amidation, and glycation have been identified using this novel method.

Considering the many features shown in this work, the middle-down nanoLC-MS/MS analysis based on size-controlled digestion should have wide applications in the proteomics field. The following directions can be considered for future exploration.

(1) Further studies regarding the digestion kinetics related to protein size and concentration is needed. This is especially important for digesting protein mixtures, as different proteins with a wide dynamic range of concentrations or sizes may undergo different extents of size-controlled digestion when they flow through the enzyme-column simultaneously.

(2) This strategy may work with large-scale protein samples for protein identification that requires data-dependent MS2. However, current data-dependent

---

MS2 methods are optimized for bottom-up strategies. Therefore, many parameters need to be reconsidered to optimize MS2 analysis of large peptides. For example, automatic selection of precursor ions should be “smart” enough such that for a given peptide, only its highest charge state ion (if above an intensity threshold) will be selected for ETD MS2. In addition, the time duration for ETD should be set based on the charge state of the target peptide ion, as the high-charge-state peptide ions react faster than low-charge-state ions.

(3) Protein PTMs are often in low stoichiometry. The reliability of using the middle-down strategy as a new direction for protein PTMs characterization (especially for protein mixtures) needs to be further examined. On the one hand, large peptides generally provide less complex chromatograms and more combinatorial PTM information compared to small tryptic peptides. On the other hand, however, large peptides are generally less sensitive than small peptides for MS detection, which is not favorable to the characterization of very low level PTMs. Fortunately the novel enzyme-column presented in this chapter shows better digestion efficiency than in-solution digestion using conventional enzymes. However, overall consideration is needed to make the final decision on whether to use bottom-up or middle-down for sample analysis. This is especially important for precious samples with limited amounts.

(4) Finally, the on-column mode digestion is attractive for online protein MS analysis. An ideal MS platform for protein identification from a complex protein

---

sample could include nanoHPLC protein separation followed by immediate protein digestion using an ESI emitter coupled to the end of the HPLC column, such that the protein eluate is digested right before ESI. This way the information of the intact protein (obtained from a separate LC-MS experiment at the protein level) can directly correlate with the resulting peptides. This online strategy could revolutionize the proteomics field for MS-based protein identification.

---



## 4.7 References

1. Abbas, A. K., Lichtman, A. H., and Pillai, S. (2009) *Cellular and Molecular Immunology*, 6th Ed., W.B. Saunders Co.
  2. Kohler, G. and Milstein, C. (1975) *Continuous cultures of fused cells secreting antibody of predefined specificity*. Nature 256, 495-497
  3. Falkenberg, F. W., Weichert, H., Krane, M., Bartelt, I., Palme, M., Nagels, H. O., and Fiebig, H. (1995) *In-vitro production of monoclonal-antibodies in high-concentration in a new and easy to handle modular minifermenter*. J. Immunol. Methods 179, 13-29
  4. Nissim, A. and Chernajovsky, Y. (2008) *Historical development of monoclonal antibody therapeutics*. Handb. Exp. Pharmacol. 181, 3-18
  5. Brekke, O. H. and Sandlie, I. (2003) *Therapeutic antibodies for human diseases at the dawn of the twenty-first century*. Nat. Rev. Drug Discovery 2, 52-62
  6. Zhang, Z., Pan, H., and Chen, X. (2009) *Mass spectrometry for structural characterization of therapeutic antibodies*. Mass Spectrom. Rev. 28, 147-176
  7. Ludwig, D. L., Pereira, D. S., Zhu, Z. P., Hicklin, D. J., and Bohlen, P. (2003) *Monoclonal antibody therapeutics and apoptosis*. Oncogene 22, 9097-9106
  8. Adams, G. P. and Weiner, L. M. (2005) *Monoclonal antibody therapy of cancer*. Nat. Biotechnol. 23, 1147-1157
  9. Liu, H., Gaza-Bulseco, G., Faldu, D., Chumsae, C., and Sun, J. (2008) *Heterogeneity of monoclonal antibodies*. J. Pharm. Sci. 97, 2426-2447
  10. Wang, W., Singh, S., Zeng, D. L., King, K., and Nema, S. (2007) *Antibody structure, instability, and formulation*. J. Pharm. Sci. 96, 1-26
  11. Wright, A. and Morrison, S. L. (1997) *Effect of glycosylation on antibody function: Implications for genetic engineering*. Trends Biotechnol. 15, 26-32
  12. Roberts, G. D., Johnson, W. P., Burman, S., Anumula, K. R., and Carr, S. A. (1995) *An integrated strategy for structural characterization of the protein and carbohydrate components of monoclonal-antibodies - application to anti-respiratory syncytial virus mab*. Anal. Chem. 67, 3613-3625
-

13. Fernandez, L., Kalume, D., Calvo, L., Mallo, M., Vallin, A., and Roepstorff, P. (2001) *Characterization of a recombinant monoclonal antibody by mass spectrometry combined with liquid chromatography*. J. Chromatogr. B 752, 247-261
  14. Hamuro, Y., Coales, S. J., Molnar, K. S., Tuske, S. J., and Morrow, J. A. (2008) *Specificity of immobilized porcine pepsin in H/D exchange compatible conditions*. Rapid Commun. Mass Spectrom. 22, 1041-1046
  15. Appel, W. (1986) *Chymotrypsin - molecular and catalytic properties*. Clin. Biochem. 19, 317-322
  16. Lee, T. D. and Shively, J. E. (1990) *Enzymatic and chemical digestion of proteins for mass-spectrometry*. Meth. Enzymol. 193, 361-374
  17. Nesvizhskii, A. and Aebersold, R. (2005) *Interpretation of shotgun proteomic data - the protein inference problem*. Mol. Cell. Proteomics 4, 1419-1440
  18. Kelleher, N. (2004) *Top-down proteomics*. Anal. Chem. 76, 196A-203A
  19. Kelleher, N. L. (2004) *Peer reviewed: Top-down proteomics*. Analytical Chemistry 76, 196A-203A
  20. Yates, J. R., III. and Kelleher, N. L. (2013) *Top down proteomics*. Anal. Chem. 85, 6151-6151
  21. Garcia, B. A. (2010) *What does the future hold for top down mass spectrometry?* J. Am. Soc. Mass Spectrom. 21, 193-202
  22. Kalli, A., Sweredoski, M. J., and Hess, S. (2013) *Data-dependent middle-down nano-liquid chromatography-electron capture dissociation-tandem mass spectrometry: An application for the analysis of unfractionated histones*. Anal. Chem. 85, 3501-3507
  23. Wu, S., Kim, J., Hancock, W. S., and Karger, B. (2005) *Extended range proteomic analysis (ERPA): A new and sensitive LC-MS platform for high sequence coverage of complex proteins with extensive post-translational Modifications* Comprehensive analysis of beta-casein and epidermal growth factor receptor (EGFR). J. Proteome Res. 4, 1155-1170
  24. Wu, C., Tran, J. C., Zamdborg, L., Durbin, K. R., Li, M., Ahlf, D. R., Early, B. P., Thomas, P. M., Sweedler, J. V., and Kelleher, N. L. (2012) *A protease for 'middle-down' proteomics*. Nat. Methods 9, 822-824
-

25. Cannon, J., Lohnes, K., Wynne, C., Wang, Y., Edwards, N., and Fenselau, C. (2010) *High-throughput middle-down analysis using an orbitrap*. J. Proteome Res. 9, 3886-3890
  26. Birktoft, J. J. and Breddam, K. (1994) *Glutamyl endopeptidases*. Methods Enzymol. 244, 114-126
  27. Raijmakers, R., Neerincx, P., Mohammed, S., and Heck, A. J. R. (2010) *Cleavage specificities of the brother and sister proteases lys-C and lys-N*. Chem. Commun. 46, 8827-8829
  28. Drapeau, G. R. (1980) *Substrate-specificity of a proteolytic-enzyme isolated from a mutant of pseudomonas-fragi*. J. Biol. Chem. 255, 839-840
  29. Garcia, B. A., Siuti, N., Thomas, C. E., Mizzen, C. A., and Kelleher, N. L. (2007) *Characterization of neurohistone variants and post-translational modifications by electron capture dissociation mass spectrometry*. Int. J. Mass Spectrom. 259, 184-196
  30. Ge, Y., Rybakova, I. N., Xu, Q., and Moss, R. L. (2009) *Top-down high-resolution mass spectrometry of cardiac myosin binding protein C revealed that truncation alters protein phosphorylation state*. Proc. Natl. Acad. Sci. U. S. A. 106, 12658-12663
  31. Carugo, O. (2008) *Amino acid composition and protein dimension*. Protein Sci. 17, 2187-2191
  32. Chanthamontri, C., Wang, W., Bruening, M. L., and Mcluckey, S. A. (2011) *Protein identification by ultrafast proteolysis using trypsin-immobilized membranes and ion-ion reactions on a quadrupole time-of-flight mass spectrometer*. ASMS, Denver, CO
  33. Tan, Y., Wang, W., Zheng, Y., Dong, J., Stefano, G., Brandizzi, F., Garavito, R. M., Reid, G. E., and Bruening, M. L. (2012) *Limited proteolysis via millisecond digestions in protease-modified membranes*. Anal. Chem. 84, 8357-8363
  34. Neurath, H., Greenstein, J. P., Putnam, F. W., and Erickson, J. A. (1944) *The chemistry of protein denaturation*. Chem. Rev. 34, 157-265
  35. Rossky, P. J. (2008) *Protein denaturation by urea: Slash and bond*. Proc. Natl. Acad. Sci. U. S. A. 105, 16825-16826
  36. Ichishima, E. (2012) *Chapter 27 Aspergillopepsin from Handbook of Proteolytic Enzymes, Volume 1*, 3rd Ed., Academic Press, Salt Lake City, UT
-

37. Gabeloteau, C. and Desnuelle, P. (1960) *On the activation of beef trypsinogen by a crystallized proteinase of aspergillus saitoi*. Biochim. Biophys. Acta 42, 230-237
38. Tanaka, N., Takeuchi, M., and Ichishima, E. (1977) *Purification of an acid proteinase from aspergillus-saitoi and determination of peptide-bond specificity*. Biochim. Biophys. Acta 485, 406-416
39. Ichishima, E. (1970) *Purification and mode of assay for acid proteinase of aspergillus saitoi*. Methods Enzymol. 19, 397-406
40. Ichishima, E. and Yoshida, F. (1966) *Conformation of aspergillopeptidase A in aqueous solution*. Biochim. Biophys. Acta 128, 130-135
41. Ichishima, E. and Yoshida, F. (1967) *Conformation of aspergillopeptidase A in aqueous solution. II. ultraviolet optical rotatory dispersion of aspergillopeptidase A*. Biochim. Biophys. Acta 147, 341-346
42. Zhang, H., Kazazic, S., Schaub, T. M., Tipton, J. D., Emmett, M. R., and Marshall, A. G. (2008) *Enhanced digestion efficiency, peptide ionization efficiency, and sequence resolution for protein hydrogen/deuterium exchange monitored by fourier transform ion cyclotron resonance mass spectrometry*. Anal. Chem. 80, 9034-9041
43. Cravetto, L., Lascoux, D., and Forest, E. (2003) *Use of different proteases working in acidic conditions to improve sequence coverage and resolution in hydrogen/deuterium exchange of large proteins*. Rapid Commun. Mass Spectrom. 17, 2387-2393
44. Syka, J. E. P., Coon, J. J., Schroeder, M. J., Shabanowitz, J., and Hunt, D. F. (2004) *Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry*. Proc. Natl. Acad. Sci. U. S. A. 101, 9528-9533
45. McLuckey, S. and Stephenson, J. (1998) *Ion ion chemistry of high-mass multiply charged ions*. Mass Spectrom. Rev. 17, 369-407
46. Coon, J. J., Ueberheide, B., Syka, J. E. P., Dryhurst, D. D., Ausio, J., Shabanowitz, J., and Hunt, D. F. (2005) *Protein identification using sequential ion/ion reactions and tandem mass spectrometry*. Proc. Natl. Acad. Sci. U. S. A. 102, 9463-9468
47. Hermanson, G. T. (2008) *Bioconjugate Techniques*, 2nd Ed., Academic Press, Burlington, MA
-

48. Varki, A., Cummings, R., Esko, J., Freeze, H., Hart, G., and Marth, J. (1999) *Essentials of Glycobiology*, 2nd Ed., Cold Spring Harbor Laboratory Press, New York, NY

49. Martin, A. C. R. (2010) *Chapter 3 Protein Sequence and Structure Analysis of Antibody Variable Domains from Antibody Engineering Volumn 2*, 2nd Ed., Springer, New York, NY