THE SMALL UBIQUITIN-RELATED MODIFIER IN THE STRESS RESPONSE AND THE USE OF MASS SPECTROMETRY/SUMMON FOR IDENTIFICATION OF UBIQUITIN AND UBIQUITIN-LIKE PROTEIN CONJUGATION SITES

by

Stanley Martin Jeram

A thesis submitted in conformity with the requirements for the degree of Master of Science
Graduate Department of Medical Biophysics
University of Toronto

© Copyright by Stanley Martin Jeram 2009
The small ubiquitin-related modifier in the stress response and the use of mass spectrometry/SUMmOn for identification of ubiquitin and ubiquitin-like protein conjugation sites

Stanley Martin Jeram
Master of Science
Graduate Department of Medical Biophysics
University of Toronto
2009

Abstract

Ubiquitin (Ub) and the ubiquitin-like proteins (Ubls) are polypeptides that can be covalently conjugated to a variety of “target” molecules to modulate their turnover rate, localization and/or function. The full range of Ubl functions is only beginning to be understood. The Raught lab is using mass spectrometry and high throughput screening methods, along with standard cell biology and biochemistry approaches, to better understand Ubl function. Here, I describe the role of a Ubl called small ubiquitin-related modifier (SUMO) in the budding yeast alcohol stress response. We identified a regulatory mechanism of the SUMO system, involving modulation of the localization of a SUMO protease. Secondly, using mass spectrometry (MS), I assisted in identifying several yeast and mammalian Ubl “chain” linkages. Finally, I propose an integrated MS methodology designed to complement standard database software for the confident identification of Ub/Ubl conjugation sites.
Acknowledgments

I first wish to express my sincere gratitude to my thesis supervisor, Dr. Brian Raught, who welcomed me into his lab. I am forever indebted to your generosity, guidance, and patience. You have unknowingly taught me lessons that I will carry with me for the rest of my life. Thank you for your constant understanding, all the laughter and great memories.

I next would like to thank all past and present members of the Raught lab, for their efforts in maintaining a friendly and collaborative workspace. I would particularly like to thank Tharan, for his company during the late nights we often shared in the lab and keeping such nights entertaining, to say the least. You played an integral part of my graduate experience and I will never forget your kindness. I would additionally like to thank Yarko for all his unforgettable anecdotes, inspiration and coffee-time contemplations we shared. Special acknowledgments are in order for Anne, Sarah and Janet for all the colorful discussions and lunch trips we shared and just generally putting up with me. I would also like to salute Tharan, Yarko and Brian for all the adventures and battles we endured together. The three of you will always be my comrades in arms and in bars.

Finally, I would like to extend my deepest, heartfelt thanks to my family; my mother Dragica, my sisters Christina and Monica, and their respective husbands Nelson and Steve, for all their love, support, and understanding through the good times and the hard times. Words alone cannot express my gratitude. I look forward to sharing years of excitement, laughter and memories with all of you. Finally, I would like to dedicate this thesis to my father, Stanislav, who bravely fought a battle against a cancer that he eventually succumbed to. You have taught me so much and continue to be an integral part of my life. You will be forever in my heart.
This thesis contains material originally published in the following review:

Reproduced with permission.
## Contents

1. **INTRODUCTION**  
   
1.1. **Ubiquitin and the ubiquitin-like proteins**  
   1.1.1. Ubiquitin  
   1.1.2. Ubiquitin-like proteins  
      1.1.2.1. SUMO  
      1.1.2.2. Other Ubls  

1.2. **Ub/Ubl conjugation/deconjugation cycle**  
   1.2.1. SUMO conjugation/deconjugation enzymes  

1.3. **Mass spectrometry and post-translation modifications**  
   1.3.1. Peptide identification methods  
      1.3.1.1. Sequence database searching  
      1.3.1.2. *De novo* sequencing  
      1.3.1.3. Peptide sequence tagging  
      1.3.1.4. Spectral matching  
   1.3.2. Identification of Ub/Ubl conjugation sites using MS  
   1.3.3. SUMmOn pattern recognition software  

1.4. **Thesis overview**  

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>Ubiquitin and the ubiquitin-like proteins</td>
<td>2</td>
</tr>
<tr>
<td>Ubiquitin</td>
<td>4</td>
</tr>
<tr>
<td>Ubiquitin-like proteins</td>
<td>5</td>
</tr>
<tr>
<td>SUMO</td>
<td>5</td>
</tr>
<tr>
<td>Other Ubls</td>
<td>9</td>
</tr>
<tr>
<td>Ub/Ubl conjugation/deconjugation cycle</td>
<td>12</td>
</tr>
<tr>
<td>SUMO conjugation/deconjugation enzymes</td>
<td>13</td>
</tr>
<tr>
<td>Mass spectrometry and post-translation modifications</td>
<td>15</td>
</tr>
<tr>
<td>Peptide identification methods</td>
<td>16</td>
</tr>
<tr>
<td>Sequence database searching</td>
<td>16</td>
</tr>
<tr>
<td><em>De novo</em> sequencing</td>
<td>17</td>
</tr>
<tr>
<td>Peptide sequence tagging</td>
<td>18</td>
</tr>
<tr>
<td>Spectral matching</td>
<td>19</td>
</tr>
<tr>
<td>Identification of Ub/Ubl conjugation sites using MS</td>
<td>20</td>
</tr>
<tr>
<td>SUMmOn pattern recognition software</td>
<td>24</td>
</tr>
<tr>
<td>Thesis overview</td>
<td>27</td>
</tr>
</tbody>
</table>
2. **SUMO AND THE STRESS RESPONSE** 28

2.1. **Chapter Overview** 28
   2.1.1. Attributions 29

2.2. **Experimental Procedures** 29
   2.2.1. Yeast strains 29
   2.2.2. Spot Assays 31
   2.2.3. Western blotting 32
   2.2.4. Affinity purification epitope-tagged proteins 33
   2.2.5. Confocal microscopy 33
   2.2.6. Metabolic poison assay 34

2.3. **Results** 34
   2.3.1. A role for SUMO in recovery from alcohol stress 34
   2.3.2. Soluble Ulp1p levels are correlated with SUMOylation 36
   2.3.3. Ulp1p is recruited to the nucleolus in response to alcohol stress 38
   2.3.4. Ulp1p co-localizes with mRNA processing/transport proteins in alcohol-stressed cells 39
   2.3.5. A Ulp1p region required for nucleolar recruitment 41
   2.3.6. Ulp1p nucleolar recruitment is energy-dependent 43

2.4. **Discussion** 44
   2.4.1. Regulation of SUMO conjugation 44
   2.4.2. The nucleolus, mRNA and stress 45
   2.4.3. Alcohol and the nuclear pore complex 45

3. **MS/SUMMON AND UB/UBL CONJUGATION SITE IDENTIFICATION** 47

3.1. **Chapter Overview** 47
   3.1.1. Attributions 49

3.2. **Experimental Procedures** 49
   3.2.1. Materials 49
   3.2.2. Smt3p mutant plasmid construction 50
3.2.3. *In vitro* Ub/Ubl reactions 50
3.2.4. Mass spectrometry and data analysis 51

3.3. Results 52
3.3.1. Alternative SUMO chain linkages 52
3.3.1.1. SUMO-2 and SUMO-3 52
3.3.1.2. Yeast SUMO 56
3.3.2. NEDD8 chain linkages 60
3.3.3. NEDDylation of UbcH12/UBE2M 65
3.3.4. Identification of Ub/Ubl-modified peptides in more complex mixtures 65
3.3.5. HCD fragmentation improves identification of some highly charged peptides 70

3.4. Discussion 72
3.4.1. An integrated MS/SUMmOn-based methodology for the identification of Ub/Ubl conjugation sites 72
3.4.2. New Ubl chain linkages 73
3.4.3. Limitations of the approach 75

4. CONCLUDING REMARKS 77
4.1. Future directions on SUMO and the stress response 77
4.2. Future directions for the identification of Ub/Ubl conjugation sites 79
4.3. General Conclusions 83

REFERENCES 85
List of Tables

Chapter One
Table 1.1. Human Ubls and their S. cerevisiae homologs.................................................................2
Table 1.2. S. cerevisiae consensus and non-consensus SUMOylation sites.................................8

Chapter Two
Table 2.1. Yeast strains used in this study.......................................................................................31

Chapter Three
Table 3.1. Ub/Ubl linkages identified by SUMmOn. .......................................................................55
Table 3.2. NEDD8-modified LysC and tryptic UbcH12/UBE2M peptides.......................................67
List of Figures

Chapter One

Figure 1.1. Ub and the Ubls...........................................................................................................................3
Figure 1.2. Ub and Ubl chain topology, and biological function.................................................................5
Figure 1.3. Graphical overview of the prototypical Ub/Ubl conjugation/ deconjugation cycle...................13
Figure 1.4. Theoretical (simplified) CID spectra of unmodified and modified peptides.............................22
Figure 1.5. A representative SUMmOn screenshot......................................................................................26

Chapter Two

Figure 2.1. The SUMO system is involved in recovery from alcohol stress..............................................35
Figure 2.2. Soluble Ulp1p levels are correlated with steady-state SUMOylation.........................................37
Figure 2.3. In response to stress, Ulp1p is recruited to the nucleolus..........................................................40
Figure 2.4. The region encompassing aa 150-340 is necessary and sufficient for nucleolar recruitment,
and recruitment is energy-dependent........................................................................................................42
Figure 2.5. Model: Ulp1p is sequestered in the nucleolus in response to alcohol stress...............................44

Chapter Three

Figure 3.1. K42-linked SUMO-2 is identified using MS/SUMmOn...............................................................54
Figure 3.2. K41-linked SUMO-3 is identified using MS/SUMmOn...............................................................57
Figure 3.3. Conjugation of Smt3R93A to RanGAP1.....................................................................................58
Figure 3.4. An Smt3p K54-linkage is identified using MS/SUMmOn.............................................................59
Figure 3.5. K22- and K48-linked NEDD8 is identified using MS/SUMmOn...................................................62
Figure 3.6. K54-linked NEDD8 is identified using MS/SUMmOn..............................................................63
Figure 3.7. Identification of K48- and K54-linked NEDD8-NEDD8 peptides using X!Tandem..........64
Figure 3.8. NEDD8 modification sites identified in UbcH12/UBE2M using MS/SUMmOn. ..............66
Figure 3.9. Elution profiles of K48-linked Ub and K11-linked SUMO-2 peptides.................................68
Figure 3.10. Extracted ion chromatograms of a representative MS run of a HeLa S100 lysate spiked
with SUMO-2 chains......................................................................................................................................70
Figure 3.11. LysC-digested K48-linked Ub is identified by MS/SUMmOn, using CID and HCD.......71
Figure 3.12. Improved workflow for the identification of Ub/Ubl-conjugation sites............................74
# List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>aa</td>
<td>amino acid</td>
</tr>
<tr>
<td>Aos1p</td>
<td>activation of Smt3p protein</td>
</tr>
<tr>
<td>AP</td>
<td>affinity purification</td>
</tr>
<tr>
<td>ATG12</td>
<td>autophagy-related protein 12</td>
</tr>
<tr>
<td>ATP</td>
<td>adenosine-5′-triphosphate</td>
</tr>
<tr>
<td>BPI</td>
<td>base peak intensity</td>
</tr>
<tr>
<td>Cdc48p</td>
<td>cell division cycle 48 protein</td>
</tr>
<tr>
<td>CID</td>
<td>collision-induced dissociation</td>
</tr>
<tr>
<td>cNLS</td>
<td>classical nuclear localization signal</td>
</tr>
<tr>
<td>Da</td>
<td>Dalton</td>
</tr>
<tr>
<td>E. coli</td>
<td>Escherichia coli</td>
</tr>
<tr>
<td>ECL</td>
<td>enhanced chemoluminescence</td>
</tr>
<tr>
<td>EDTA</td>
<td>ethylenediaminetetraacetic acid</td>
</tr>
<tr>
<td>FAT10</td>
<td>human leukocyte antigen F-associated transcript 10</td>
</tr>
<tr>
<td>FDR</td>
<td>false discovery rate</td>
</tr>
<tr>
<td>FT</td>
<td>Fourier transform</td>
</tr>
<tr>
<td>FUB1</td>
<td>Fau ubiquitin like 1</td>
</tr>
<tr>
<td>GFP</td>
<td>green fluorescent protein</td>
</tr>
</tbody>
</table>
GST  glutathione S-transferase
HA  hemagglutinin
HCD  higher-energy C-trap dissociation
HEPES  4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HPLC  high-performance liquid chromatography
HRP  horseradish peroxidase
Hs  Homo sapiens
HU  hydroxyurea
HUB1  homologous to ubiquitin 1
IA  iodoacetamide
IgG  immunoglobulin G
ISG15  interferon stimulated gene 15kDa protein
Kan  kanamycin
Kap  karyopherin
LC-MS  liquid chromatography-mass spectroscopy
LC-ESI  liquid chromatography-electrospray ionization
LTQ  linear ion trap
LysC  lysyl endopeptidase C
MLP3A-LC3  microtuble-associated proteins 1A/1B light chain 3A
MAT  mating type locus
MDM2  murine double minute 2
Mms21  methyl methanesulfonate sensitivity 21
MS  mass spectrometry
MS/MS  tandem mass spectrometry
MudPIT  multi-dimensional protein identification technology
NEDD8  neural precursor cell-expressed developmentally downregulated-8
Nop  nucleolar protein
NPC  nuclear pore complex
NUB1  NEDD8 ultimate buster-1
NUB1L  NUB1 splice variant
OD$_{600}$  optical density units at 600 nm
ORF  open reading frame
PAGE  polyacrylamide gel electrophoresis
PBS  phosphate buffered saline
PC2  polycomb 2 homolog
PCR  polymerase chain reaction
PIAS  protein inhibitor of activated STAT
PML  promyelocytic leukemia protein
protA  protein A
PTM  post-translational modification
RanBP2  ran-binding protein 2
RanGAP1  ran GTPase activating protein 1
RFP  red fluorescent protein
RNF4  RING finger protein 4
RUB1  related to ubiquitin 1
*S. cerevisiae*  *Saccharomyces cerevisiae*
SAE  SUMO activating enzyme
*Sc*  *Saccharomyces cerevisiae*
SCF  Skp1-Cdc53/Cul1-F-box protein
SDS  sodium dodecyl sulfate
SDS-PAGE  sodium dodecyl sulfate- polyacrylamide gel electrophoresis
SENP  SUMO-specific protease
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>SIM</td>
<td>SUMO-interacting motif</td>
</tr>
<tr>
<td>Siz</td>
<td>SAP and Miz-finger domain</td>
</tr>
<tr>
<td>SM</td>
<td>selective minimal media</td>
</tr>
<tr>
<td>Smt3p</td>
<td>suppressor of Mif23 protein</td>
</tr>
<tr>
<td>Smt3R93A</td>
<td>Smt3p with arginine 93 mutated to alanine</td>
</tr>
<tr>
<td>SNAPS</td>
<td>single nucleotide-induced amino acid polymorphisms</td>
</tr>
<tr>
<td>STUbL</td>
<td>SUMO-targeted ubiquitin ligase</td>
</tr>
<tr>
<td>SUMO</td>
<td>small ubiquitin-related modifier</td>
</tr>
<tr>
<td>TAP</td>
<td>tandem-affinity purification</td>
</tr>
<tr>
<td>Th</td>
<td>Thompson</td>
</tr>
<tr>
<td>TIC</td>
<td>total ion current</td>
</tr>
<tr>
<td>TPP</td>
<td>trans-proteomic pipeline</td>
</tr>
<tr>
<td>Tris</td>
<td>tris-’(hydroxymethyl)-aminomethane</td>
</tr>
<tr>
<td>Ub</td>
<td>ubiquitin</td>
</tr>
<tr>
<td>Uba</td>
<td>ubiquitin-activating</td>
</tr>
<tr>
<td>Ubc</td>
<td>ubiquitin-conjugating</td>
</tr>
<tr>
<td>Ubl</td>
<td>ubiquitin-like protein</td>
</tr>
<tr>
<td>UBL5</td>
<td>ubiquitin-like protein 5</td>
</tr>
<tr>
<td>UFM1</td>
<td>ubiquitin fold modifier 1</td>
</tr>
<tr>
<td>Ulp</td>
<td>ubiquitin-like protease</td>
</tr>
<tr>
<td>URM1</td>
<td>ubiquitin related modifier 1</td>
</tr>
<tr>
<td>VHL</td>
<td>Von Hippel-Lindau</td>
</tr>
<tr>
<td>WCL</td>
<td>whole cell lysate</td>
</tr>
<tr>
<td>YPD</td>
<td>yeast-peptone-dextrose media</td>
</tr>
</tbody>
</table>
Chapter One

Introduction

Proteins are tightly controlled through a variety of mechanisms, including the regulation of protein quantity, localization and activity. The regulation of proteins can occur via the processes of transcription and translation, or after proteins have been synthesized, through a variety of post-translational modifications (PTMs). As the name suggests, a PTM is the chemical alteration of a protein after it has been synthesized, and represents a rapid, versatile and energy-efficient method to regulate protein activity or function. Many PTMs are reversible and dramatically increase the functional diversity of the proteome. Many PTMs are small chemical modifications of specific amino acids, such as acetylation, methylation and phosphorylation. However, PTMs are not restricted to small molecules; large, highly complex molecules, such as carbohydrates (i.e. glycosylation) or entire polypeptides (as described below) can be covalently attached to a variety of target proteins to modulate their function.
1.1. Ubiquitin and the ubiquitin-like proteins

Ubiquitin (Ub) and a growing family of Ub-related molecules, collectively referred to as the ubiquitin-like proteins (Ubls), are ~100 aa polypeptides (Table 1.1) that are covalently conjugated to proteins or other molecules to modulate their turnover rate, localization, and/or function. The Ubls share only a low degree of primary sequence identity with one another (Figure 1.1, Table 1.1), but share a common globular three-dimensional structure, referred to as the ubiquitin (or β-grasp) fold [1]. The Ubls do not possess any native catalytic activity, but rather have been proposed to: (i) alter the tertiary structure of modified target molecules, thereby altering their function; (ii) compete with other types of PTMs for the same sites of conjugation, and; (iii) recruit binding partners that would normally not interact with the unmodified

<table>
<thead>
<tr>
<th>Ubiquitin-like Protein</th>
<th>Polypeptide Length (aa)</th>
<th>Ubiquitin Sequence Homology (%)</th>
<th>S. cerevisiae Homolog</th>
<th>Polypeptide Length (aa)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ubiquitin</td>
<td>76</td>
<td>100</td>
<td>Ubiquitin</td>
<td>76</td>
</tr>
<tr>
<td>NEDD8</td>
<td>76</td>
<td>58</td>
<td>RUB1</td>
<td>76</td>
</tr>
<tr>
<td>ISG15</td>
<td>157</td>
<td>36</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FUB1</td>
<td>74</td>
<td>36</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FAT10</td>
<td>165</td>
<td>30*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>UBL5</td>
<td>72</td>
<td>21</td>
<td>HUB1</td>
<td>72</td>
</tr>
<tr>
<td>SUMO-1</td>
<td>97</td>
<td>20</td>
<td>SMT3</td>
<td>98</td>
</tr>
<tr>
<td>SUMO-2</td>
<td>93</td>
<td>13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SUMO-3</td>
<td>92</td>
<td>13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>UFM1</td>
<td>83</td>
<td>17</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MLP3A-LC3</td>
<td>120</td>
<td>10</td>
<td>ATG8</td>
<td>116</td>
</tr>
<tr>
<td>ATG12</td>
<td>140</td>
<td>9</td>
<td>ATG12</td>
<td>186</td>
</tr>
<tr>
<td>URM1</td>
<td>101</td>
<td>16</td>
<td>URM1</td>
<td>99</td>
</tr>
</tbody>
</table>

Table 1.1. Human Ubls and their S. cerevisiae homologs.
Length (in amino acids) of the mature protein for each Ubl is listed along with sequence homology to human Ub. Asterisk (*) denotes sequence homology of FAT10 against a tandem repeat of human Ub, because FAT10 consists of two head-to-tail ubiquitin-like domains (see Section 1.1.2.2)
target molecules. Differences in primary sequence apparently confer each Ubl with a unique surface charge distribution, which is thought to allow for unique interactions with Ubl-specific binding partners. The identity of the Ubl-specific binding partners, the thousands of different Ubl conjugates, and the global function of the various Ubls are only beginning to be characterized.

**Figure 1.1. Ub and the Ubls.**

(A) Amino acid sequence alignment of Ub and Ubls. Gene name and species (*Homo sapiens* (Hs); *S. cerevisiae* (Sc)) are indicated at left. C-terminal amino acids present in the pro-protein, but absent in the mature protein, are indicated in green. Ub is synthesized as a fusion protein with head-to-tail repeats of itself or as a fusion with ribosomal proteins and FUB1 is also synthesized as a fusion with a ribosomal protein (as indicated by .....). The Ubl C-terminal “tail” that remains conjugated to a target molecule following trypsin *(bold/red text)* or LysC *(underlined text)* digestion is highlighted. Additional residues resulting from LysC digestion are indicated by blue text. Lysine residues involved in Ub/Ubl chain formation are in bold. (B) Dendrogram based on multiple sequence alignments of Ub and Ubls. Neighbor joining with Kimura’s correction for distances, ignoring positions with gaps in the alignment, are illustrated. Shown are Ub and Ubls from *S. cerevisiae* *(underlined)* and *H. sapiens*. 
1.1.1. Ubiquitin

Ub is a small polypeptide ubiquitously expressed in all eukaryotes (but not prokaryotes). Ub conjugation (or ubiquitylation) is an extremely versatile post-translational modification that has been implicated in a wide variety of critical cellular processes, including protein stability, cell cycle progression, transcriptional control, receptor transport, and the immune response [2]. Thousands of proteins are modified by ubiquitin [reviewed in 3], and deregulation of the ubiquitin system has been implicated in e.g. neurological disorders, muscle wasting, cardiomyopathy, and various cancers [4, 5].

Monoubiquitylation (i.e. conjugation of a single Ub polypeptide to a target protein lysine residue) has been implicated in processes such as endocytosis and plasma membrane receptor recycling [6]. However, Ub contains 7 internal lysine residues, all of which can also be ubiquitylated in vitro and in vivo to form poly-Ub multimers, or Ub “chains” ([reviewed in 7]; Figure 1.2). The ability to form chains dramatically increases the functional diversity of Ub, in that multimers of varying lengths and linkage types can confer very different biological outcomes to a targeted protein substrate (Figure 1.2). The best characterized function of poly-Ub chains (in this case, consisting of at least four Ub molecules linked via lysine 48; K48) is the targeting of a protein substrate for 26S proteasome-dependent degradation [8]. By contrast, K63-linked Ub chains do not target protein substrates for degradation, but instead play a role in the DNA damage response and Epsin-mediated endocytosis [9, 10]. The functions of most other Ub chain types (including mixed linkage multimers), are not well understood [7].
Figure 1.2. **Ub and Ubl chain topology, and biological function.**

The topology of Ub chains determines biological outcome. Ub chains linked via K63 play roles in the DNA damage response and endocytosis. K48-linked chains act as a signal for 26S proteasome-dependent degradation. Other chain topologies, as well as mixed chains, have also been observed *in vivo*, but their biological functions remain less well understood. All of the SUMO proteins, as well as NEDD8, can also form chains, but the functions of these multimeric structures remain to be elucidated.

### 1.1.2. Ubiquitin-like proteins

#### 1.1.2.1. SUMO

The SUMO (small ubiquitin-related modifier) proteins are a family of Ubls expressed in all eukaryotes and, like Ub, are also covalently conjugated to the epsilon amine group of lysine residues in a large number of target proteins [11-18]. The *S. cerevisiae* SUMO ortholog is encoded by the Suppressor of Mif Two 3 (*SMT3*) gene, which is an essential gene for yeast. Mammals express at least three different SUMO proteins (SUMOs-1-3) [19], with a fourth protein (SUMO-4) recently identified [20]. Considerable debate remains as to whether SUMO-4 is actually a pseudogene [21]. SUMO-1 shares less than 50% homology with SUMOs-2 and -3 (which share 95% identity with each other), and the various SUMO isoforms target both unique and overlapping protein populations [e.g. 18].
SUMO has been implicated in a variety of cellular processes. Smt3p can regulate intracellular localization, cell cycle progression, DNA repair and replication, RNA metabolism, cell signaling and stress responses [17, 22]. More than 500 putative yeast SUMO targets (~8% of all yeast open reading frames; ORFs) have been identified to date [11-16], highlighting the importance of the SUMO system in this organism.

Changes in mammalian SUMO conjugate populations have been noted in various neurodegenerative disorders and malignancies [23-26], and SUMO conjugates include many proteins relevant to human health, such as the transcription factors p53 and c-jun, the promyelocytic leukemia protein (PML), the Werner helicase, and huntingtin [25, 27, 28]. SUMO conjugation (or SUMOylation) is clearly an important cellular process, and a better understanding of its nature may provide important insight on cancer pathogenesis and novel therapies.

While SUMOylation does not directly target proteins for degradation, in some cases it can recruit specific Ub E3 ligases, and thereby act as an indirect degradation signal [29, 30]. These Ub E3 ligases, called SUMO-targeted ubiquitin ligases (STUbLs), contain one or more SUMO-interacting motifs (SIMs) [31], which allow them to specifically interact with the SUMO protein. The STUbL family of proteins are evolutionarily conserved, and include the yeast heterodimer Slx5p-Slx8p [29, 30] and mammalian RNF4 [32]. Inactivation or deletion of STUbLs can cause an accumulation of SUMOylated proteins, an increase in sensitivity to genotoxic stress, and genomic instability [33]. Rad60p has been identified as a potential STUbL target [33], further implicating the SUMO system in DNA repair.

More recently, a global map of SUMO function was constructed by integrating extensive genetic and physical interaction networks with previously published SUMO conjugate data [22]. Makhnevych et al. highlighted strong crosstalk between the SUMO pathway and the Ub system [22]. The global map
included Slx5p and Slx8p, which were previously shown to physically interact with SUMO [29, 30], and highlighted a role of SUMO in Cdc48p localization and function [22]. Cdc48p promotes the degradation of ubiquitylated ER proteins by extracting them from the ER membrane and guiding them to the 26S proteasome [34]).

SUMOylation has also been demonstrated to e.g. stabilize proteins [35], recruit histone deacetylases [36], or recruit transcriptional co-repressors [37]. The SUMO pathway has also been implicated in lipid metabolism, DNA replication and repair, and DNA damage checkpoint recovery [22]. With the construction of a global SUMO map, Makhnevych et al. highlighted a role for the SUMO pathway in 15 broad biological processes, showcasing the importance of this small polypeptide in a vast number of cellular functions. How SUMO can effect such a wide variety of different functions has not been elucidated.

A consensus sequence for SUMOylation, ψKxE/D, where ψ is a hydrophobic residue (usually I, L or V), K is the lysine residue to which SUMO is conjugated, and x is any amino acid, was first described for the yeast septin proteins [17]. This sequence was later also shown to be sufficient for SUMOylation of some mammalian proteins [e.g. 38]. However, as more SUMO conjugation sites have been identified, it has become clear that not every SUMO consensus site is SUMOylated, and that SUMO is conjugated to target proteins at a variety of “non-consensus” sites. While relatively few sites have been identified to date, more than 40% of the published yeast SUMO conjugation sites occur at non-consensus lysine residues (Table 1.2). Thus, while several different algorithms have been developed for the prediction of putative SUMOylation sites [39-42], these programs have somewhat limited predictive value, due to their heavy reliance on the SUMO consensus sequence.

Like Ub, the yeast and metazoan SUMO proteins can also form multimeric chains [43] (see Figure 1.2). The major SUMO branching sites appear to be located in unstructured N-terminal extensions: K15 in Smt3p; K7, K16 and K17 in SUMO-1;
### Table 1.2. *S. cerevisiae* consensus and non-consensus SUMOylation sites.

SUMOylation sites identified in budding yeast proteins, along with 10 flanking amino acids. Sites that conform to the degenerate consensus sequence ‘ψKx/E/D’ are shown in the upper half of the table, and non-consensus sites are displayed in the lower half.
and K11 in SUMO-2/3 [52-54]. While the Smt3p and SUMO-2/3 branch sites conform to the SUMOylation consensus motif, the SUMO-1 sites do not. Recently, a report indicated that SUMO chains may promote ubiquitylation of PML by RNF4, and thus facilitate the proteasomal-degradation of PML [25]. The effect of poly-SUMOylation on the function of a target protein may be distinct from mono-SUMOylation (as is the case with poly-Ub and mono-Ub, see Section 1.1.1), or it may be that poly-SUMO chains more efficiently recruit SIM-containing proteins, such as RNF4 [25]. Overall, although progress is being made in the SUMO field, the function of SUMO chains remains unknown.

1.1.2.2. Other Ubls

NEDD8 (neural precursor cell-expressed developmentally downregulated-8) is a 9.1 kDa polypeptide displaying 58% sequence homology with Ub. The *S. cerevisiae* NEDD8 ortholog is encoded by the Related to Ubiquitin 1 (RUB1) gene. NEDD8 was first identified as a modifier of Cdc53, a cullin component of the SCF (Skp1-Cdc53/Cul1-F-box protein) Ub E3 ligase complex [55]. Until recently, the cullins were the only known targets of NEDD8 conjugation (or NEDDylation); however, NEDD8 modification has now been observed on a variety of targets, including the transcription factors p53 and p73, the Von-Hippel Lindau (VHL) protein, the epidermal growth factor receptor and a number of ribosomal proteins [56-59]. NEDDylation appears to activate and/or stabilize multicomponent Ub E3 ligases, and has been reported to direct other targets to the 26S proteasome through interactions with two adaptor proteins, NUB1 (NEDD8 ultimate buster-1) and NUB1L (a NUB1 splice variant) [60, 61]. While cell cycle progression is the primary biological process that has been associated with NEDD8 [55], how NEDDylation
regulates this process is not understood. NEDD8 has also been reported to form chains in vitro and in vivo [56, 62], but the function(s) of NEDD8 multimers have not been characterized and remain poorly understood (see Figure 1.2).

ISG15 (interferon stimulated gene 15kDa protein) was the first Ubl to be identified [63]. As its name suggests, ISG15 expression is induced by exposure to type 1 interferons, as well as lipopolysaccharides and double-stranded RNA. More than 200 proteins have been identified as putative ISG15 targets, including the signaling molecules JAK1, STAT1 and ERK1/2, implicating this Ubl in the regulation of translation, glycolysis, stress responses, chromatin remodeling, and cell motility [64-66]. No ISG15 homologs have been detected in lower eukaryotes, and how ISGylation affects the functions of its target proteins remains largely unknown.

FAT10 (human leukocyte antigen F-associated transcript 10) is an 18kDa protein comprised of two tandem head-to-tail ubiquitin-like domains (FAT10 is also referred to as Diubiquitin or Ubiquitin D), identified with other human leukocyte antigen F-adjacent transcripts (FATs) in genetic studies of the histocompatibility complex class I region [67]. FAT10 can interact directly with the 26S proteasome through either of its Ubl domains or through the NUB1L adaptor protein [68], and is thus thought to provide a ubiquitin-independent pathway for protein degradation. FAT10 conjugates have been observed by Western blot analysis, but no targets have been identified [69].

UBL5 (ubiquitin-like protein 5) is an 8.5kDa polypeptide possessing a C-terminal dityrosine motif, rather than the typical diglycine moiety present in most other Ubls. This has led to some debate concerning whether UBL5 can be conjugated to other molecules. However, the yeast UBL5 ortholog HUB1 (Homologous to Ubiquitin 1) is apparently conjugated to proteins [70].

FUB1 (Fau ubiquitin like 1) is a 7.8kDa ubiquitin-like protein that shares 36% sequence homology with Ub. FUB1 is encoded by the Fau gene [71], and is
synthesized as a fusion with the ribosomal S30 protein. The fusion protein is cleaved to produce a Ubl with a C-terminal diglycine motif. The B-cell lymphoma-G protein is a FUB1 target [72], and this Ubl was also reported to be secreted, exerting anti-inflammatory/antiproliferative effects on immune cells through an unknown mechanism [73].

UFM1 (ubiquitin fold modifier 1) is a 9.1kDa polypeptide that shares only 17% sequence similarity, but a high degree of structural homology, with Ub. UFM1 possesses a single C-terminal glycine residue. While immunoprecipitation studies have indicated the presence of in vivo UFM1 conjugates, none have been identified to date [74].

URM1 (ubiquitin related modifier 1) is an 11kDa protein with only one known protein target, Ahp1p (alkyl hydroperoxide reductase-1) [75]. Urm1p shares 20% sequence homology, as well as a high degree of structural similarity, with the *E. coli* sulphur carrier proteins THiS and MoaD [76], and appears to play a critical role in the thio-modification of some tRNAs [77]. The bacterial sulphur transfer system is thought to be an evolutionary precursor to the Ub/Ubl conjugation system [76]. It will thus be very interesting to investigate the Urm1p system to understand how the sulphur carrier pathway may have played a role in the evolution of Ub/Ubl conjugation.

The ATG12 (autophagy-related protein 12) and ATG8 (autophagy-related protein 8) genes were identified in studies of autophagy-defective yeast mutants. The Atg12p protein is conjugated to the Atg5p protein, and this complex is required for autophagosome formation [78]. The Atg8p Ubl (the human homolog is microtubule-associated proteins 1A/1B light chain 3A; MLP3A-LC3) is conjugated to phosphatidylethanolamine to promote autophagosome membrane fusion [79, 80]. Thus, while the first Ubl targets to be identified were proteins, it is now clear that other classes of molecules are also targeted by this diverse PTM superfamily.
1.2. Ub/Ubl conjugation/deconjugation cycle

Ub and most of the Ubls are synthesized as pro-proteins that are processed by specific proteases (a.k.a. maturases) to expose a carboxyl-terminal glycine residue (Figure 1.3, step 1). The glycine residue is covalently conjugated, via an isopeptide (or amide) bond, to a reactive group in a target molecule (e.g. the epsilon amine group of a lysine side chain on a target protein, or an amine moiety on a target lipid). The conjugation process is typically effected via a three-stage enzymatic cascade specific for each Ubl. In the first stage, an E1 (activating) enzyme utilizes ATP to form a high-energy thioester bond between the E1 active cysteine residue and a C-terminal glycine residue of the mature Ubl (step 2) [81]. An E2 (conjugating) enzyme binds the E1 when both the adenylate and thioester E1 sites are occupied by Ubl. E2 binding is accompanied by transfer of the thioester bond to the E2 active cysteine residue (step 3). An E3 ligase then recruits the target molecule to the E2 for Ubl transfer (step 4). In some cases, the Ubl may be transferred to the E3 prior to substrate conjugation [2]. The Ubls may be removed from their targets via the action of specific proteases (step 5) [82, 83]. Of note, an E3 may not be required for conjugation of all Ubls, as is the case for SUMO [52]. Additionally, some E1, E2 and E3 proteins appear to interact with multiple Ubls, for example see reference [84].
Figure 1.3. Graphical overview of the prototypical Ub/Ubl conjugation/deconjugation cycle. Most Ubls are synthesized as pro-proteins. Maturation requires the removal of C-terminal amino acids by a Ubl-specific protease (step 1). Ubl conjugation is typically effected by a three step cascade involving (step 2) activation by an E1 protein, (step 3) transfer to an E2 enzyme, and (step 4) recruitment of a target molecule by an E3 ligase. Removal of the Ubl from a substrate (step 5) may be effected by Ubl-specific proteases.

1.2.1. SUMO conjugation/deconjugation enzymes

Unlike the Ub system (which is effected by a large number of E2 enzymes and by several hundred E3s), SUMOylation is coordinated by significantly fewer enzymes. For instance, in *S. cerevisiae*, the SUMO pathway consists of an E1 heterodimer...
composed of Aos1p/Uba2p, a single E2 enzyme Ubc9p, and three known E3 ligases (Siz1p, Siz2p and Mms21p), reviewed in [85]. Mammalian SUMO is conjugated to target proteins by an E1 heterodimer composed of SAE1/SAE2 (SUMO-activating enzyme 1 and 2, respectively), the E2 enzyme Ubc9, and three types of E3 ligases; the PIAS (protein inhibitor of activated STAT) family, RanBP2 (Ran-binding protein), and PC2 (polycomb 2 homolog), reviewed in [86]. Each type of E3 ligase is thought to provide distinct target selectivity for the SUMOylation machinery, however, a comprehensive list of targets and their respective E3 ligases has not been established.

SUMO is conjugated to a variety of targets to elicit a specific response, but may be deconjugated when its function is no longer required. *S. cerevisiae* expresses two SUMO proteases, Ulp1p and Ulp2p [87, 88]. Ulp1p is localized to the nuclear pore complex (NPC), and is involved in both the maturation and deconjugation of Smt3p [89]. Ulp2p is distributed within the nucleus, and functions as an additional deconjugation enzyme without maturase activity [87]. Human SUMO maturation and deconjugation are effected by six different SUMO-specific proteases (SENP1-3, 5-7) [90]. Generally speaking, Ulp1p is most closely related to SENP1, 2, 3, and 5, whereas Ulp2p is related to SENP6 and 7.

Many reports have indicated that a proper balance of SUMOylation and deSUMOylation is required for proper cell function. *ULP1* is an essential gene in yeast, and Ulp1p over-expression is lethal [91] [87]. *ULP2* is dispensable, and its deletion has no effect on yeast viability under standard laboratory conditions. However, *ULP2* deletion results in a dramatic increase in high molecular weight SUMO conjugates (presumably SUMO chains [92]), suggesting that Ulp2p may be involved in processing SUMO chains. With regards to the mammalian proteases, disruption of the SENP1 gene is associated with infantile teratoma [93], and a fusion SENP6 gene is observed in a human T cell lymphoma cell line [94].
1.3. Mass spectrometry and post-translation modifications

Current mass spectrometers, coupled with high-throughput strategies and modern software, can rapidly and efficiently identify many different types of PTMs [95, 96]. Global MS analyses using these technologies are beginning to uncover the wondrous complexity and variety of PTMs, and laboratories around the world are currently characterizing how PTMs are regulated during the cell cycle, throughout development, and in response to signaling, stress or infection. MS is also being utilized to characterize PTM patterns in many different disease states, in an effort to better understand the roles of specific intracellular signaling pathways and their targets in disease etiology and progression [97-99]. However, while standard database search algorithms (see below) can accurately identify peptides modified by relatively simple chemical PTMs such as phosphorylation, acetylation and methylation from MS data, more specialized algorithms are required for the identification of many types of Ubl conjugation sites. Ubls and the application of MS to their study have recently been reviewed in [1, 3, 100, 101].

In a typical LC-ESI-MS/MS (liquid chromatography-electrospray ionization-tandem mass spectrometry) analysis, the protein(s) to be studied are cleaved with one or more proteolytic enzymes (typically trypsin, which cleaves on the C-terminal side of arginine and lysine residues, unless followed by a proline). The peptide mixture is then most often separated via reversed-phase chromatography, and analyzed using data-dependent MS/MS acquisition. Following a precursor scan (a.k.a. a parent ion or MS scan), the most intense precursor ions are sequentially selected for collision-induced dissociation (CID). The resulting fragment ions are analyzed to generate an MS/MS spectrum (a.k.a. a tandem mass or product ion spectrum). During CID, many types of peptide fragments can be generated. At the
low collision energy settings used in typical proteomics applications, the major products are ions derived from fragmentation events occurring at consecutive peptide bonds, referred to as the $b$- and $y$-ion series. These ion series are most often used in bottom-up proteomics applications to reconstruct the peptide sequence (reviewed in [102]).

1.3.1. Peptide identification methods

Peptides are generally identified using one or more of the following methods: (i) sequence database searching, (ii) de novo sequencing, (iii) peptide sequence tagging, or (iv) spectral matching.

1.3.1.1. Sequence database searching

Sequence database searching (performed by programs such as SEQUEST [103], Mascot [104] and X!Tandem [105]) has revolutionized the field of proteomics, and is currently the most efficient and popular method of peptide identification. This search method compares an MS/MS fragment ion spectrum against theoretical fragmentation patterns derived from a user-defined protein database, and assigns scores based on the quality of the match [e.g. 106]. Database search engines thus require a reference database containing a comprehensive list of proteins, usually translated open reading frames from a fully sequenced genome. Along with increases in MS sensitivity and speed, these tools have allowed for the analysis of extremely complex biological samples. However, one disadvantage of database searching is that it cannot account for peptide variants not represented in the
database, such as uncharacterized splicing variants or sequence polymorphisms: An estimated 55,000 to 250,000 single nucleotide-induced amino acid polymorphisms (SNAPS) occur in human protein encoding genes [107]. Since the human sequence databases used in most proteomics studies are compiled from a small number of individuals, a significant proportion of peptides in a particular human sample may not be identified (or will be misidentified), due to naturally occurring variation in peptide amino acid sequence. Fortunately, this does not pose a major problem in experiments designed simply to identify the proteins in a sample of interest; as long as one or more peptides derived from a given protein is confidently identified, one can conclude that this protein was present in the original sample.

However, a separate search must often be conducted for each PTM (or for each small group of PTMs) of interest, and standard database search software does not typically search for PTMs that have not been specified in the search parameters by the user. Hundreds of different PTMs have been characterized to date (e.g. the PTM repository Unimod currently lists 560 protein modifications [108]). Searching the same dataset sequentially for a large number of PTMs leads to dramatic increases in computational time and resources [109], and may increase the false identification of modifications, highlighting the inefficiency of such an approach.

1.3.1.2. De novo sequencing

Before the existence of database searching software, de novo sequencing was used for peptide identification. In the de novo sequencing approach, the peptide sequence is determined solely from the fragmentation pattern of the MS/MS spectra, by assigning the amino acid sequence based upon the m/z difference between putative b- and y-ion fragments (for detailed reviews refer to [110, 111]). In the early days of
MS-based peptide identification, this was conducted manually. More recently, many different software programs have been developed for this task (e.g. PEAKS [112], PepNovo [113] and Lutefisk [114]). While the development of mass spectrometers with higher sensitivity and mass accuracy (e.g. the LTQ-Orbitrap, FT-MS) has helped to improve the signal-to-noise ratio of MS/MS spectra and the ability of de novo search algorithms to assign peptide sequences more accurately [115], this method typically requires a higher level of computational resources, and does not perform as well as database search algorithms. De novo sequencing is thus not currently routinely used for standard identifications. Nevertheless, in some cases a protein sequence database may not be available for a particular organism of interest, and de novo sequencing must be used.

1.3.1.3. Peptide sequence tagging

Peptide sequence tagging is a hybrid peptide identification method, combining elements of both sequence database searching and de novo sequencing. This method was first proposed by Mann & Wilm as an error-tolerant approach for peptide identification using sequence databases [116]. First, a short partial sequence or tag is inferred directly from the MS/MS spectrum, either manually or through the use of simple de novo sequencing algorithms. Using a sequence database search engine, the relevant sequence tag, along with the parent mass and other information are used to identify the full peptide sequence. Several approaches have been developed that use sequence tags as a filter for database searching [117-120]. Overall, the peptide sequence tagging approach requires less computational resources than traditional de novo sequencing, because only a short partial sequence is required. Additionally, using sequence tags to filter the number of possible matches dramatically decreases
the size of the database used for searching, allowing for a more thorough PTM search [120, 121]. However, since the peptide sequence tag contains only a partial peptide sequence, problems can arise when attempting to distinguish between closely related proteins and protein homologues. As compared to sequence database searching, peptide sequence tagging approaches identify fewer complete matches (e.g. GutenTag [117]). Recently there has been a resurgence of interest in peptide sequence tagging approaches (due to improved technology, MS accuracy), however sequence database search algorithms remain the method of choice.

1.3.1.4. Spectral matching

A fourth method for peptide identification is spectral matching. Theoretically, once many thousands of peptides from a given organism have been observed in a mass spectrometer, it may be more efficient to search against a library of actual spectra (or of an averaged representation of observed spectra), rather than comparing experimental spectra to theoretical in silico-generated spectral models. Such an approach requires significantly lower computational resources [122], and spectra of poorly fragmented peptides with few peaks may still be matched, so long as the fragmentation is characteristic for the unique peptide in question. Several spectral matching tools have been developed, including SpectraST [122], X!Hunter [123], and HMMatch [124]. This method has some of the same drawbacks as sequence database searching, in that rare peptide variants may not be included in the library. In addition, modified peptides are typically present in much lower stoichiometries than their unmodified counterparts, and would therefore be less likely to be represented in the spectrum library. Similarly, many peptides are modified at multiple sites (e.g. Ser/Thr phosphorylation combined with Ubl modification on the
same peptide). To successfully identify these types of peptides, each modified variant (the unmodified peptide, both types of singly modified peptide, and the doubly modified peptide) must be present in the spectrum library. With the hundreds of known modifications, this approach is thus not designed for PTM discovery, but rather such an approach is beneficial for the faster identification of previously observed peptides.

1.3.2. Identification of Ub/Ubl conjugation sites using MS

Database search algorithms are very effective at detecting peptides modified by many different types of PTMs. Most small chemical PTMs do not undergo significant fragmentation during the CID process, but simply alter one or more peptide fragment ions by a predictable, indivisible mass, which does not undergo fragmentation. Thus, the theoretical spectra used in the matching process may simply be recalculated to reflect an appropriate mass shift on one or more amino acids, as specified by the user. For example, acetylation of a lysine residue causes a +42.01 Da shift of the peptide fragments containing the modified amino acid, which can easily be distinguished from an unmodified peptide (Figure 1.4A and B). Database search algorithms can account for this mass shift, and successfully assign the match. Similarly, following trypsin cleavage, a Ub-derived peptide is modified at a lysine residue with a –GG (diglycine) remnant. Trypsin cleaves conjugated lysine residues inefficiently [125], resulting in GG-peptide conjugates containing an internal missed cleavage site. During CID, the –GG modification itself fragments only rarely, and thereby shifts the mass of peptide fragments containing the modified lysine residue by +114.04 Da. These characteristics are used by peptide sequencing software to identify Ub-conjugated target peptides (Figure 1.4C). Many
hundreds of ubiquitin conjugation sites have been successfully identified using standard database searching methods [e.g. 126].

Importantly, however, Ub is not the only modification that yields a GG-modified target peptide following trypsin cleavage. NEDD8, Rub1p, and ISG15 also possess C-terminal –(R)GG sequences (Figure 1.1A). Following trypsin cleavage, these Ubl-derived –GG remnants are thus indistinguishable from Ub-derived –GG remnants, resulting in a loss of modifier identity. Importantly, false positive Ub/Ubl–GG identifications can also arise following treatment of the sample with iodoacetamide (IA). IA is a universally used cysteine alkylating agent, recently demonstrated to form adducts on lysine residues with a mass identical to a –GG modification [127].

In an attempt to maintain modifier identity, several groups have expressed epitope tagged Ub or Ubls in vivo, which can then be used to purify covalently linked target proteins [56, 62, 65, 126, 128]. For example, Peng et al. identified 1075 putative Ub substrates and 110 Ub conjugation sites, using His6-tagged Ub [126]. A similar approach was utilized by Jones et al. to identify 496 NEDD8 conjugates using GST-tagged NEDD8 [62]. Hundreds of SUMO conjugates have also been identified using His6- or TAP-tagged SUMO proteins [11-15, 129]. A variety of other epitope tags have also been utilized to identify a multitude of other Ubl conjugates. These approaches have allowed large-scale MS screens to be conducted with the added confidence that any GG-modified peptide identified is most likely derived from the tagged Ubl. However, a single protein may be simultaneously modified by multiple Ubls. For example, p53 and MDM2 are modified by SUMO, Ub and NEDD8; identification of the -GG moiety therefore cannot always be uniquely assigned to Ub or any other particular Ubl [59, 130-132]. This issue was highlighted in a recent large-scale study designed to identify NEDD8 targets, in that a number of Ub peptides were identified in a GST-NEDD8 purification [62]. The epitope
Figure 1.4. Theoretical (simplified) CID spectra of unmodified and modified peptides. 

(A) Simplified CID spectrum of an unmodified peptide. Trypsin cleavage sites are indicated by green triangles. $b$- and $y$-ion series are indicated by dark and light blue peaks, respectively. 

(B) CID spectrum of an acetylated peptide. In this case, the entire $b$-ion series, as well as the $y_5$ ion, undergo a +42.01 Da mass shift (shifted ions indicated in magenta). The position of the unmodified $b_4$ ion (from panel A) is indicated with a broken blue line. 

(C) CID spectrum of a GG-modified peptide. The entire $b$-ion series, as well as the $y_5$ ion, undergo a +114.04 Da mass shift (shifted ions highlighted in orange). The position of the unmodified $b_4$ ion (from panel A) is indicated with a broken blue line. 

(D) CID spectrum of a Ubl-modified peptide. $b'$- and $y'$-ion series denote the fragmentation ion series of the modifier (in red and pink, respectively). These additional Ubl fragment ions create a highly complex CID spectrum which is uninterpretable by standard database search software. The $b_4$ ion, for example, undergoes an unknown mass shift, dependent on the mass of the modifier. Similarly, the $y_2$ ion is dependent upon the mass of the target peptide, and would remain unassigned. The $y_2$, $y_3$ and $y_4$ ions are unaffected, but would not be assigned by database search software. The position of the unmodified $b_4$ ion (from panel A) is indicated with a broken blue line.
tagging/pulldown approach could therefore lead to misassigned or missed Ub/Ubl conjugation sites. Epitope-tagged Ubls may also display altered conjugation specificity. For example, in one study, expression of a FLAG-tagged SUMO protein resulted in lower levels of SUMO-conjugated proteins than was observed with an untagged SUMO protein [12].

Another common problem encountered in identifying Ubl-modified proteins is that they are often of very low abundance, display rapid turnover rates, and are rapidly deconjugated during cell lysis, making their isolation extremely difficult. The use of proteasome inhibitors (e.g. MG132 or lactacystin), general protease inhibitors (e.g. N-ethylmaleimide), or chemically modified non-conjugatable Ub/Ubl proteins (e.g. Ub aldehyde) in the cell lysis buffer can however improve Ub/Ubl conjugate yield [92, 133]. In vitro approaches can also be used to generate larger amounts of modified protein for MS analysis. While this approach can often provide valuable information (e.g. identifying putative in vivo modification sites for follow-up studies), the identified modification sites may not always accurately reflect the in vivo situation.

Unfortunately, even if sufficient amounts of biologically relevant material can be obtained, many Ubls are not efficiently identified using standard sample preparation and database search algorithms. Trypsin cleavage of many of the Ubls (e.g. the SUMOs, URM1, FAT10, HUB1, etc.) results in a much longer peptide remnant conjugated to a target molecule (Figure 1.1A). During CID, the long Ubl C-terminal “tail” itself produces multiple fragment ions (Figure 1.4D). Combined with fragment ions derived from the target peptide, a complex overlapping fragmentation pattern is thus generated, which cannot be identified with standard search software. A related problem is that Ubls yielding longer peptide remnants also typically yield highly charged peptides (4+ and higher). Standard database
search algorithms do not typically search for peptides with higher charge states (most algorithms consider only 3+ and lower).

Due to these limitations, very few Ubl conjugation sites have been identified using MS-based methods. Some studies have utilized manual sequencing of high-quality MS/MS spectra [e.g. 134], an extremely labor-intensive practice not amenable to high-throughput analysis. Another MS-based method for the identification of SUMOylation sites relies upon the introduction of an arginine residue N-terminal to the -GG C-terminus, such that trypsin digestion yields an easily identifiable GGG-remnant, exactly like that used in the identification of Ub-modified peptides [40, 135]. However, it is now clear that many SUMOylated proteins are also ubiquitylated [25, 27]. This type of SUMO mutant could thus lead to misassigned SUMO and/or Ub conjugation sites. Mutations in SUMO could also have unknown effects on conjugation site selection. Ubl conjugation sites therefore continue to be identified primarily through mutagenesis of guessed sites on a protein(s) of interest [17], an extremely laborious approach, especially for proteins containing multiple lysine residues and/or Ubl conjugation sites.

1.3.3. SUMmOn pattern recognition software

To identify Ubl conjugation sites using MS/MS, Pedrioli et al. developed the SUMmOn pattern recognition software tool [54], which utilizes a novel type of search strategy. In a first step, SUMmOn searches MS/MS spectra for the presence of a user-defined fragment ion series, corresponding to the Ubl of interest. Since Ubl C-termini are conjugated to target peptides, the Ubl b-ion series (these fragments are termed the b’-ion series when the Ubl is conjugated to a target molecule; Figure 1.4D) is independent of the target peptide. However, the Ubl y-ion series (the
dependent y'-ions) include the entire target peptide (Figure 1.4D). SUMmOn calculates each y'-ion by subtracting the corresponding paired b'-ion of the relevant Ubl tail from the precursor ion mass of each MS/MS spectrum. This process is repeated for each y'-ion, in each charge state series. SUMmOn then searches for user defined b'-ion fragments and the calculated y'-ion fragments. The matching of each m/z intensity peak in an MS/MS spectrum is assumed to be a Bernoulli trial, and a binomial probability density function is used to calculate the probability of matching a given number of peaks in any given MS/MS scan. This probability, the intensity of the matched ion fragments, and other parameters are used to calculate a “modification score”. The mass of the target peptide is then calculated by subtracting the mass of the Ubl C-terminal “tail” from the measured precursor mass. Similar to the modification score, a “target score” is assigned by determining the quality of the match of the remaining fragment ions to theoretical peptide fragment ion masses generated via an in silico digest of a user-defined reference database. Generally, a smaller protein sequence database (a “subdatabase”), containing only proteins identified using an independent search method for unmodified peptides (e.g. a SEQUEST, Mascot or X!Tandem search of the same MS data), is used to reduce computational time, and decrease false positive identification rates. Unlike most database search software, SUMmOn is designed to handle highly charged fragment ions (up to 9+). Briefly, users can filter SUMmOn output for a specific MS/MS scan number, a particular target peptide precursor mass, a particular modification and/or target score threshold (e.g. modification score ≥ 0.8), or a combination of the above and other parameters (see Figure 1.5 for a representative software screenshot of data output). Overall, SUMmOn software can therefore successfully interpret complex MS/MS spectra that other peptide sequencing software cannot process [54].
Figure 1.5. A representative SUMmOn screenshot.

The SUMmOn header indicates MS/MS scan number, precursor charge state, relative scoring position of the assigned charge state compared to other charge states (relPos), the SUMmOn modification score (modScore), uncharged precursor mass, number of matched ions, total ion current (TIC), percentage TIC matched by the assigned ions (%TIC), and base peak intensity (BPI) for the specified MS/MS scan. Target peptide information is displayed to left of spectrum, where the target protein name and peptide sequence is listed. K(0) denotes the modified lysine residue. The SUMmOn target score is listed along with other target peptide information. The calculated C-terminal fragment ion series for several charge states is indicated below the spectrum, where matched y'-ions (dep) are colored red and matched b'-ions (ind) are colored blue. Matched peaks corresponding to b'- and y'-ions are highlighted on the spectrum in shades of red and green, respectively.
1.4. Thesis overview

While we have made much recent progress in understanding the wide variety of functions of the SUMO protein, a better understanding of the SUMO conjugation machinery, how the SUMOylation process is regulated, and the identification of SUMO targets remains to be accomplished.

Unlike mammalian cells, *S. cerevisiae* contains a single SUMO protein (Smt3p), making budding yeast an ideal model organism for the study of SUMO function. In Chapter 2, a variety of biochemical techniques and confocal microscopy are utilized to perform an in-depth analysis on the role of the yeast SUMO protease, Ulp1p, in the regulation of the SUMO pathway, specifically in response to alcohol stress. From this, a SUMO-based stress response is proposed.

Chapter 3 describes an integrated MS/SUMmOn-based methodology for the identification of Ub and Ubl conjugation sites. Using this methodology, several SUMO and NEDD8 chain assemblies were identified, in addition to several NEDDylation sites in the NEDD8 E2 enzyme.
Chapter Two

SUMO and the stress response

2.1. Chapter Overview

The following chapter proposes that modulation of SUMO protease localization represents a mechanism for the control of SUMO conjugate levels in response to alcohol stress. This chapter also implicates the SUMO system in the recovery from alcohol stress in *S. cerevisiae*.

As described earlier, the SUMO system has been implicated in a variety of critical cellular functions, such as DNA repair and replication, RNA metabolism, and stress responses [85, 136, 137]. Importantly, the SUMO system is highly dynamic, and the SUMO enzymes appear to work together to precisely control SUMO conjugate levels in the cell [85, 136, 137]. Dramatic changes in SUMO conjugate populations have been noted in response to many different types of stresses in yeasts, mammals and plants [16, 138-141]. For example, in *S. cerevisiae*, increased steady-state SUMO conjugate levels are observed in response to high (but
physiological) concentrations of ethanol (EtOH) [16]. However, how the SUMO system is regulated in response to stress, and indeed the relevance of the SUMO response to stress recovery, are not well understood.

### 2.1.1. Attributions

I would like to acknowledge Drs. Brenda J. Andrews, Charles Boone, Richard W. Wozniak and Michael Tyers for providing various strains and plasmids. I would also like to highlight contributions made by Dr. Yaroslav Sydorskyy, a post-doctoral fellow in the Raught lab who generated a significant amount of data used in Figure 2.2B and part of E, Figure 2.3, and Figure 2.4. Dr. Sydorskyy also constructed most of the yeast strains used in this work. Ph.D. candidate Tharan Srikumar kindly helped to generate data used in Figure 2.1 and Figure 2.2.

### 2.2. Experimental Procedures

#### 2.2.1. Yeast strains

*S. cerevisiae* strains used in this study were derivatives of BY4741/BY4742 haploid strains, unless otherwise specified (Table 2.1). Genetic manipulations were performed according to established procedures [142]. Yeast strains were grown at 30°C to mid-logarithmic phase in yeast-peptone-dextrose (YPD) or selective minimal (SM) media supplemented with appropriate nutrients and 2% glucose. Yeast transformations were performed as described [143].
<table>
<thead>
<tr>
<th>Strain</th>
<th>Genotype</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>BY4741</td>
<td>MATa his3-Δ1 leu2-Δ0 met15-Δ0 ura3-Δ0</td>
<td>[144]</td>
</tr>
<tr>
<td>Ulpl-TAP</td>
<td>MATa ulp1Δ::ULP1-TAP::HIS3MX6</td>
<td>[145]</td>
</tr>
<tr>
<td>Aos1-TAP</td>
<td>MATa AOS1-TAP::HIS3MX6</td>
<td>[146]</td>
</tr>
<tr>
<td>Uba2-TAP</td>
<td>MATa UBA2-TAP::HIS3MX6</td>
<td>[146]</td>
</tr>
<tr>
<td>Ubc9-TAP</td>
<td>MATa UBC9-TAP::HIS3MX6</td>
<td>[146]</td>
</tr>
<tr>
<td>Ulpl-HA</td>
<td>MATa pULP1-HA</td>
<td>This study</td>
</tr>
<tr>
<td>Ulp2-HA</td>
<td>MATa pULP2-HA</td>
<td>This study</td>
</tr>
<tr>
<td>Aos1-GFP</td>
<td>MATa AOS1-GFP::HIS3MX6</td>
<td>[147]</td>
</tr>
<tr>
<td>Uba2-GFP</td>
<td>MATa UBA2-GFP::HIS3MX6</td>
<td>[147]</td>
</tr>
<tr>
<td>Ubc9-GFP</td>
<td>MATa UBC9-GFP::HIS3MX6</td>
<td>[147]</td>
</tr>
<tr>
<td>Nup59-GFP</td>
<td>MATa NUP59-GFP::HIS3MX6</td>
<td>[147]</td>
</tr>
<tr>
<td>Nup100-GFP</td>
<td>MATa NUP100-GFP::HIS3MX6</td>
<td>[147]</td>
</tr>
<tr>
<td>Nop2-GFP</td>
<td>MATa NOP2-GFP::HIS3MX6</td>
<td>[147]</td>
</tr>
<tr>
<td>Noc3-GFP</td>
<td>MATa NOC3-GFP::HIS3MX6</td>
<td>[147]</td>
</tr>
<tr>
<td>Nop56-GFP</td>
<td>MATa NOP56-GFP::HIS3MX6</td>
<td>[147]</td>
</tr>
<tr>
<td>Noc2-GFP</td>
<td>MATa NOC2-GFP::HIS3MX6</td>
<td>[147]</td>
</tr>
<tr>
<td>Nop4-GFP</td>
<td>MATa NOP4-GFP::HIS3MX6</td>
<td>[147]</td>
</tr>
<tr>
<td>Nop6-GFP</td>
<td>MATa NOP6-GFP::HIS3MX6</td>
<td>[147]</td>
</tr>
<tr>
<td>Nop58-GFP</td>
<td>MATa NOP58-GFP::HIS3MX6</td>
<td>[147]</td>
</tr>
<tr>
<td>Nab2-GFP</td>
<td>MATa NAB2-GFP::HIS3MX6</td>
<td>[147]</td>
</tr>
<tr>
<td>Tom1-GFP</td>
<td>MATa TOM1-GFP::HIS3MX6</td>
<td>[147]</td>
</tr>
<tr>
<td>Hrb1-GFP</td>
<td>MATa HRB1-GFP::HIS3MX6</td>
<td>[147]</td>
</tr>
<tr>
<td>Dbp5-GFP</td>
<td>MATa DBP5-GFP::HIS3MX6</td>
<td>[147]</td>
</tr>
<tr>
<td>Mtr10-GFP</td>
<td>MATa MTR10-GFP::HIS3MX6</td>
<td>[147]</td>
</tr>
<tr>
<td>Hrp1-GFP</td>
<td>MATa HRP1-GFP::HIS3MX6</td>
<td>[147]</td>
</tr>
<tr>
<td>Gbp2-GFP</td>
<td>MATa GBP2-GFP::HIS3MX6</td>
<td>[147]</td>
</tr>
<tr>
<td>Sub2-GFP</td>
<td>MATa SUB2-GFP::HIS3MX6</td>
<td>[147]</td>
</tr>
<tr>
<td>Yra1-GFP</td>
<td>MATa YRA1-GFP::HIS3MX6</td>
<td>[147]</td>
</tr>
<tr>
<td>Thp1-GFP</td>
<td>MATa THP1-GFP::HIS3MX6</td>
<td>[147]</td>
</tr>
<tr>
<td>Sac3-GFP</td>
<td>MATa SAC3-GFP::HIS3MX6</td>
<td>[147]</td>
</tr>
<tr>
<td>Mex67-GFP</td>
<td>MATa MEX67-GFP::HIS3MX6</td>
<td>[147]</td>
</tr>
<tr>
<td>Mtr2-GFP</td>
<td>MATa MTR2-GFP::HIS3MX6</td>
<td>[147]</td>
</tr>
<tr>
<td>Npl3-GFP</td>
<td>MATa NPL3-GFP::HIS3MX6</td>
<td>[147]</td>
</tr>
<tr>
<td>DF5</td>
<td>MATa hisp-32 his3-Δ200 trp1-1 1leu2-3,112 lys2-801</td>
<td>[147]</td>
</tr>
<tr>
<td>Ulpl-GFP</td>
<td>MATa ulp1Δ::KANMX4 CDC3-HA::HIS5 pULP1-GFP</td>
<td>[145]</td>
</tr>
<tr>
<td>pULP1&lt;sup&gt;150-621&lt;/sup&gt;-GFP</td>
<td>MATa ulp1Δ::KANMX4 CDC3-HA::HIS5 pULP1&lt;sup&gt;150-621&lt;/sup&gt;-GFP</td>
<td>[145]</td>
</tr>
<tr>
<td>pULP1&lt;sup&gt;150-340&lt;/sup&gt;-GFP</td>
<td>MATa ulp1Δ::KANMX4 CDC3-HA::HIS5 pULP1&lt;sup&gt;150-340&lt;/sup&gt;-GFP</td>
<td>[145]</td>
</tr>
<tr>
<td>pULP1&lt;sup&gt;1-150&lt;/sup&gt;-GFP</td>
<td>MATa pULP1&lt;sup&gt;1-150&lt;/sup&gt;-GFP</td>
<td>This study</td>
</tr>
<tr>
<td>pULP1&lt;sup&gt;150-340&lt;/sup&gt;-GFP</td>
<td>MATa pULP1&lt;sup&gt;150-340&lt;/sup&gt;-GFP</td>
<td>This study</td>
</tr>
<tr>
<td>Nup60-GFP</td>
<td>MATa NUP60-GFP::HIS3MX6 pULP1-RFP</td>
<td>This study</td>
</tr>
<tr>
<td>Mlp1-GFP</td>
<td>MATa MLP1-GFP::HIS3MX6 pULP1-RFP</td>
<td>This study</td>
</tr>
<tr>
<td>Mlp2-GFP</td>
<td>MATa MLP2-GFP::HIS3MX6 pULP1-RFP</td>
<td>This study</td>
</tr>
<tr>
<td>Esc1-GFP</td>
<td>MATa ESC1-GFP::HIS3MX6 pULP1-RFP</td>
<td>This study</td>
</tr>
</tbody>
</table>
Nup2-GFP  \( \text{MATa NUP2-GFP::HIS3MX6} \) [147]
Nup53-GFP  \( \text{MATa NUP53-GFP::HIS3MX6} \) [147]
Nup84-GFP  \( \text{MATa NUP84-GFP::HIS3MX6} \) [147]
Nup133-GFP  \( \text{MATa NUP133-GFP::HIS3MX6} \) [147]
Nup145-GFP  \( \text{MATa NUP145-GFP::HIS3MX6} \) [147]
Nup159-GFP  \( \text{MATa NUP159-GFP::HIS3MX6} \) [147]
Nup170-GFP  \( \text{MATa NUP170-GFP::HIS3MX6} \) [147]
Nup188-GFP  \( \text{MATa NUP188-GFP::HIS3MX6} \) [147]
Nup192-GFP  \( \text{MATa NUP192-GFP::HIS3MX6} \) [147]
Nup2Δ pULP1-GFP  \( \text{MATa nup2Δ::KANMX4 pULP1-GFP} \) This study
Nup42Δ pULP1-GFP  \( \text{MATa nup42Δ::KANMX4 pULP1-GFP} \) This study
Nup53Δ pULP1-GFP  \( \text{MATa nup53Δ::KANMX4 pULP1-GFP} \) This study
Nup59Δ pULP1-GFP  \( \text{MATa nup59Δ::KANMX4 pULP1-GFP} \) This study
Nup60Δ pULP1-GFP  \( \text{MATa nup60Δ::KANMX4 pULP1-GFP} \) This study
Nup100Δ pULP1-GFP  \( \text{MATa nup100Δ::KANMX4 pULP1-GFP} \) This study
Nup133Δ pULP1-GFP  \( \text{MATa nup133Δ::KANMX4 pULP1-GFP} \) This study
Nup170Δ pULP1-GFP  \( \text{MATa nup170Δ::KANMX4 pULP1-GFP} \) This study
Nup188Δ pULP1-GFP  \( \text{MATa nup188Δ::KANMX4 pULP1-GFP} \) This study
EscΔ pULP1-GFP  \( \text{MATa escΔ::KANMX4 pULP1-GFP} \) This study
Asr1Δ pULP1-GFP  \( \text{MATa asr1Δ::KANMX4 pULP1-GFP} \) This study
Mlp1/2Δ pULP1-GFP  \( \text{MATa mlp1Δ::KANMX4 mlp2Δ::NATMX4 pULP1-GFP} \) This study
Nup59Δ pULP1\(^{150-621}\)-GFP  \( \text{MATa nup59Δ::KANMX4 pULP1}\(^{150-621}\)-GFP} \) This study
Nup100Δ pULP1\(^{150-621}\)-GFP  \( \text{MATa nup100Δ::KANMX4 pULP1}\(^{150-621}\)-GFP} \) This study
Nup2Δ pULP1\(^{150-621}\)-GFP  \( \text{MATa nup2Δ::KANMX4 pULP1}\(^{150-621}\)-GFP} \) This study
Nup60Δ pULP1\(^{150-621}\)-GFP  \( \text{MATa nup60Δ::KANMX4 pULP1}\(^{150-621}\)-GFP} \) This study
EscΔ pULP1\(^{150-621}\)-GFP  \( \text{MATa escΔ::KANMX4 pULP1}\(^{150-621}\)-GFP} \) This study
Mlp1/2Δ pULP1\(^{150-621}\)-GFP  \( \text{MATa mlp1Δ::KANMX4 mlp2Δ::NATMX4 pULP1}\(^{150-621}\)-GFP} \) This study
\( \text{ubc9-2} \)  \( \text{MATa ubc9-2::NATMX4 can1Δ::STE2pr-Sp_his5 lyp1-Δ} \) [145]

**Table 2.1. Yeast strains used in this study.**

Yeast strains used in this study were derived from BY4741 (MATa) or DF5 (MATa). Fluorescently tagged proteins were either GFP (green fluorescent protein) or RFP (red fluorescent protein).

### 2.2.2. Spot Assays

50 ml cultures were grown overnight in YPD at 28°C to OD\(_{600}\) 0.5. Strains were then diluted to OD\(_{600}\) 0.25 and grown at 34°C to OD\(_{600}\) 0.5 (~2 hours). The indicated alcohol was added to 20 ml aliquots of each culture, and incubated at 34°C for 1 hr. After treatment, each strain was diluted to OD\(_{600}\) 0.5 and serially diluted into sterile
water. 5 μl of four serial ten-fold dilutions, ranging from 1:10 to 1:10000, were plated onto YPD agar plates supplemented with 2% glucose, but lacking alcohol, and grown at 34°C for 36 hours. Alcohols used were: MeOH (99.8%, EMD Chemicals Inc.), EtOH (95%, Commercial Alcohols), iPrOH (99.5%, BioShop Canada), BuOH (99.8%, BioShop Canada), and HexOH (98%, Fluka Chemie).

2.2.3. Western blotting

Whole cell lysates were prepared by alkaline lysis and trichloracetic acid protein precipitation of cell pellets derived from 10 ml cultures treated with alcohol, and collected at time points as indicated. Protein pellets were resuspended in SDS-PAGE sample buffer, sonicated for 10 sec and incubated at 90°C for 5 min. Proteins were resolved on 10% or 4-12% SDS-polyacrylamide gels and transferred to nitrocellulose membranes. Specific rabbit polyclonal antibodies were used to detect yeast Smt3p and Swi6p (Rockland, Inc.). Rabbit serum (Sigma-Aldrich) was used to detect the *Staph. aureus* protein A (protA) in the TAP and HA tags. The TAP tag consists of a calmodulin binding peptide (CBD) followed by protA, with a tobacco etch virus protease (TEV protease) site between CBD and protA. The HA tag consists of a His6-HA-protA fusion, with a protease 3C site between HA and protA. Both the TAP and HA tags are expressed as fusions of the tag to the C-terminus of the protein under study. Binding of primary antibodies was detected using horseradish peroxidase (HRP)-conjugated donkey anti-rabbit secondary antibodies (Bio-Rad Laboratories) and ECL (Bio-Rad Laboratories) or ECL-plus (Millipore Corporation).
2.2.4. Affinity purification epitope-tagged proteins

Affinity purification of TAP- and HA-tagged proteins was performed using a single step purification method utilizing an IgG affinity column. Cells were lysed by bead beating with 0.5 mm Glass Beads (BioSpec Products, Inc.) in lysis buffer (10% glycerol, 50 mM HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid) pH 8.0, 100 mM KCl, 2 mM EDTA (ethylenediaminetetraacetic acid), 0.1% nonidet P40). Lysates were cleared by centrifugation at 16 000xg and incubated overnight with Pan Mouse IgG Dynabeads (Invitrogen Dynal AS, Oslo, Norway). Beads were washed 3x with PBS (phosphate buffered saline), and proteins eluted in Laemmli buffer and resolved on 10% polyacrylamide gels.

2.2.5. Confocal microscopy

Mid-log phase cells were washed briefly in 2% glucose, and mounted on glass slides. Cells were imaged 60x/1.42 N PlanApo on an Olympus IX80 inverted microscope (Olympus Canada) fitted with a Yokogawa CSU10 spinning disk confocal scanner unit (Quorum Technologies Inc., Guelph, ON, Canada) and a 512 x 512 EM-CCD camera (Hamamatsu, Japan). The system was controlled with Volocity 4.3.2 software (Improvision Ltd.). The CCD camera was operated at maximum resolution, and gain set to 1, binning to x1, and sensitivity to 100 (except for Ulp1-RFP images, in which it was set to 130). Exposure times varied by strain, ranging from 0.2 to 10 s. Settings were maintained for all subsequent images of a strain.
2.2.6. **Metabolic poison assay**

Treatment of cultures with metabolic poison was performed essentially as previously described [145]. Cell cultures were grown in SM media, washed twice in media lacking glucose, and incubated for 45 min in media lacking glucose, but supplemented with 100mM 2-deoxyglucose (Sigma-Aldrich) and 10mM sodium azide (BioShop Canada). Alcohol treatment was performed either prior to or after metabolic poisoning.

2.3. **Results**

2.3.1. **A role for SUMO in recovery from alcohol stress**

High, physiological levels of EtOH have been previously reported to dramatically increase SUMO conjugate populations in *S. cerevisiae* [16]. However, how the SUMO system is regulated in response to alcohol stress, and the relevance of the increase in SUMO conjugate levels to stress recovery, are not well understood. Thus, a temperature sensitive (ts) yeast strain in which SUMOylation can be inactivated, *ubc9-2* [148], was used to examine the role of the SUMO system in recovery from alcohol stress. Following acute exposure to alcohol, cells were washed and plated in increasing dilutions on standard YPD agarose plates. Colony formation - reflecting successful recovery from stress - was monitored. *ubc9-2* cells maintained at the permissive temperature (28°C) were unaffected by acute alcohol treatment (Figure 2.1A and data not shown). Exposure to 10% EtOH for up to 60 min also had no
effect on wild type (wt) yeast (Figure 2.1B). However, when *ubc9*-2 cells were shifted to the semi-permissive temperature (34°C) to inactivate SUMOylation, exposed to 10% EtOH for 1 h, then plated on solid media, a consistent and significant decrease in the size and number of colonies was observed (Figure 2.1B). These data suggest that Ubc9p activity (and thus SUMOylation) plays an important role in recovery from acute EtOH stress. To determine if this effect was specific for EtOH, or was instead a general response to alcohol stress, we subjected the wt and *ubc9*-2 strains to a variety of simple alcohols. As compared to EtOH treatment, the *ubc9*-2 strain was less sensitive to 10% methanol (MeOH), but more sensitive to isopropanol (iPrOH), and displayed extreme sensitivity to very low concentrations of butanol (BuOH) and hexanol (HexOH). At these low concentrations, the wt strain was completely unaffected (Figure 2.1B). SUMO conjugation thus appears to play an important role in recovery from acute alcohol exposure.

**Figure 2.1. The SUMO system is involved in recovery from alcohol stress.**
(A) *ubc9*-2 cells maintained at the indicated temperature were exposed to 10% EtOH for 60 min, washed, then plated on standard solid media. Colony formation was monitored after 2 d. (B) Wild type (wt) and *ubc9*-2 cells were shifted to the indicated temperature for 2 h, then exposed to the indicated alcohol for 60 min. Cells were washed and plated on solid media. Colony formation was monitored after 2 d.
2.3.2. Soluble Ulp1p levels are correlated with SUMOylation

To better understand how alcohol stress modulates steady-state SUMOylation levels, we characterized the effects of high alcohol concentrations on the proteins responsible for SUMO conjugation and deconjugation. Yeast strains expressing epitope-tagged SUMO pathway proteins were exposed to 10% EtOH for 1 h, and whole cell lysates (WCL; prepared by lysing yeast directly in Laemmli buffer) were analyzed via Western blotting. No apparent changes in expression level or electrophoretic mobility were observed (Figure 2.2A). However, isolation of the same proteins via immunoprecipitation (affinity purification on Flag- or ProtA-sepharose; AP) revealed a striking decrease in Ulp1p apparent solubility: *i.e.* while the E1 and E2 proteins displayed no change (or a slight decrease) in response to alcohol treatment, and Ulp2p levels displayed a ~2-fold decrease over 1 h, the level of Ulp1p protein available for immunoprecipitation was extremely sensitive to alcohol treatment (Figure 2.2A), disappearing within 5 min of EtOH exposure. This effect was not related to the epitope tag, and was not due to epitope masking in the presence of alcohol. A Ulp1-TAP protein displayed identical behavior (Figure 2.2B and C), and, using the AP bead-beating lysis procedure, but followed by TCA protein precipitation and Western analysis, also revealed a loss of Ulp1p from the supernatant (Figure 2.2B).

The loss of Ulp1p from immunoprecipitates was dose-dependent (Figure 2.2C). No effect was noted within 60 min for up to 5% EtOH, but 7.5% EtOH effected Ulp1p disappearance within 60 min, and higher EtOH concentrations led to a more rapid loss. Exposure to MeOH and iPrOH (Figure 2.2D) also effected a loss of Ulp1p in immunoprecipitates, with more rapid decreases in soluble Ulp1p levels associated with increasing alcohol hydrophobicity.
Figure 2.2. Soluble Ulp1p levels are correlated with steady-state SUMOylation.
(A) Ulp1p is insoluble in alcohol-stressed cells. Yeast strains expressing epitope-tagged versions of SUMO system proteins were treated with 10% EtOH, and cultures harvested at the indicated times. Whole cell lysates (WCL) were prepared, and affinity purification (AP) conducted, for each protein. Western blotting was conducted to monitor protein levels. Swi6p was used as a loading control. (B) Ulp1-TAP was affinity purified from untreated cells, or cells treated with 10% EtOH for 60 min (lanes 1, 2). Western analysis of supernatant (supe) and pellet from the same cultures (lanes 3-6). (C) Ulp1p solubility is dose-dependent. Ulp1-TAP was affinity purified from cells treated with various concentrations of EtOH, as indicated. (D) Ulp1p solubility is also affected by other alcohols. The Ulp1-TAP strain was treated for 60 min with 10% MeOH or 10% iPrOH, and Ulp1p apparent solubility monitored by AP and Western analysis. (E) Soluble Ulp1p levels are inversely correlated with steady-state sumoylation. A Ulp1-TAP culture was treated with 10% EtOH for 60 min, removed from EtOH-containing media, washed, and grown in fresh media for an additional 180 min. Aliquots were removed at indicated time points. WCL and affinity purified Ulp1-TAP were subjected to Western analysis, as indicated.
Consistent with a role for Ulp1p in the regulation of SUMO conjugate levels in response to alcohol stress, steady-state SUMOylation was inversely correlated with soluble Ulp1p levels (Figure 2.2E). Removal of alcohol from the culture media (after 60 min) was followed by a gradual increase in soluble Ulp1p levels, accompanied by a concomitant decrease in SUMO conjugates (Figure 2.2E). As in Figure 2.2A, total Ulp1p levels measured in WCL (bottom panel) were unchanged in response to alcohol exposure (Figure 2.2E).

2.3.3. Ulp1p is recruited to the nucleolus in response to alcohol stress

The fact that Ulp1p solubility is affected by alcohol exposure suggested that it could be recruited to an intracellular compartment or structure that is insoluble under standard lysis conditions. To explore this possibility, we characterized Ulp1p localization in cells subjected to alcohol stress. No changes in intracellular localization were observed for Aos1-GFP, Uba2-GFP or Ubc9-GFP in response to 10% EtOH treatment, but Ulp1p localization was dramatically altered, moving from the nuclear rim to large nuclear foci (Figure 2.3A). This effect was not restricted to EtOH; identical results were obtained with iPrOH and BuOH (data not shown).

To better define the location and composition of Ulp1p-containing stress foci, Ulp1p was co-expressed with a series of intracellular markers. As previously reported [149], Ulp1-RFP co-localized with NPC components in untreated cells (e.g. Nup59p and Nup100p; Figure 2.3B). However, following alcohol treatment, the bulk of Ulp1p no longer co-localized with NPC proteins, but was observed in large foci distributed around the nuclear periphery (Figure 2.3B).

Imaging of alcohol-treated cells expressing GFP-tagged nucleolar markers indicated that the nucleolus is fragmented in response to this type of stress (Figure
2.3C), with increasing alcohol exposure time or chain length correlated with increased fragmentation. Following alcohol treatment, Ulp1-RFP co-localized with Nop2p, Nop62p, Noc3p (Figure 2.3C) and several other nucleolar markers (data not shown). Ulp1p is therefore recruited to the nucleolus in response to alcohol stress, and Ulp1p-containing stress foci represent nucleolar fragments.

2.3.4. Ulp1p co-localizes with mRNA processing/transport proteins in alcohol-stressed cells

Following capping, splicing and polyadenylation, mature mRNA is transported from the nucleus to the cytoplasm via the NPC. In response to many different types of stress (including alcohol) poly(A)^+ RNA nucleocytoplasmic export is inhibited, and many proteins involved in RNA processing and transport are sequestered in the nucleus [150-152]. Ulp1p and other proteins associated with the NPC were previously implicated in mRNA surveillance [153, 154]. We thus characterized Ulp1-RFP localization in strains expressing GFP-tagged RNA processing and transport proteins. Ulp1p did not co-localize with Hrb1p, Dbp5p (Figure 2.3D), Thp1p, Sac3p, Mex67p, Gbp2p, Mtr10p, Yra1p, Hrp1p, Mtr2p, Npl3p or Sub2p (data not shown) in untreated or alcohol-treated cells. However, in alcohol-treated cells Ulp1p co-localized with Nab2p, a shuttling hnRNP protein required for nuclear mRNA export and poly(A) tail length control [151], and Tom1p, a putative HECT ubiquitin E3 ligase that regulates Nab2p nucleocytoplasmic export [155] (Figure 2.3D). Ulp1p did not co-localize with these proteins in untreated cells (Figure 2.3D). Ulp1p-containing nucleolar fragments therefore also contain a subset of proteins involved in RNA processing and transport. While we do not yet understand its full role, these data strongly implicate the nucleolus in the alcohol stress response.
Figure 2.3. In response to stress, Ulp1p is recruited to the nucleolus.
(A) Ulp1p localization is altered in response to alcohol stress. Confocal microscopy of untreated and EtOH-treated strains expressing GFP-tagged Aos1p, Uba2p, Ubc9p or Ulp1p. Scale bar, 5 μm. (B) Ulp1p does not co-localize with Nup59p or Nup100p following alcohol treatment. Cells co-expressing Ulp1-RFP and Nup59-GFP or Nup100-GFP were imaged before and after treatment with 10% EtOH for 60 min. (C) Ulp1p is recruited to the nucleolus. Cells co-expressing Ulp1-RFP and Nop2-GFP, Nop62-GFP or Noc3-GFP were imaged before and after treatment with 10% EtOH for 60 min. (D) Ulp1p co-localizes with RNA processing/transport proteins in alcohol-stressed cells. In alcohol-stressed cells, Ulp1p co-localizes with Nab2p and Tom1p, but not Dbp5p or Hrb1p. Cells co-expressing Ulp1-RFP and Nab2-GFP, Tom1-GFP, Dbp5-GFP or Hrb1-GFP were imaged before and after treatment with 10% EtOH for 60 min.
2.3.5. **A Ulp1p region required for nucleolar recruitment**

To better understand the structural requirements for Ulp1p nucleolar recruitment, we characterized the localization of a set of previously described Ulp1-GFP deletion mutants [145] in the presence and absence of alcohol (Figure 2.4A). Previous reports demonstrated that Ulp1p is tethered to the NPC via: aa 1-150, a region containing a Kap121p/Pse1p binding site, and aa 150-340, a region containing a Kap95p/Kap60p binding site [89]. As previously observed [145], Ulp1150-621-GFP is partially mislocalized in untreated cells, displaying both nuclear rim and cytoplasmic localization (Figure 2.4A). The N-terminal 150 aa are thus required to maintain normal levels of Ulp1p localization at the NPC. Surprisingly, following alcohol treatment, the cytoplasmic signal of Ulp1150-621-GFP was significantly diminished, and the protein rapidly recruited to nucleolar fragments (Figure 2.4A). Thus, even cytoplasmic Ulp1p protein is efficiently recruited to the nucleolus in response to alcohol stress, and the Kap121p binding region is not required for this process.

Ulp1150-340-GFP displayed the same properties as the Ulp1150-621-GFP protein (Figure 2.4A), indicating that the catalytic region (residues 403-621) does not play an important role in nucleolar recruitment. Ulp11-150-GFP displayed both cytoplasmic and nuclear localization (Figure 2.4A), and EtOH treatment had no apparent effect on the subcellular distribution of this protein. Finally, Ulp1p lacking only aa 150-340 (Ulp1Δ150-340-GFP) displayed localization similar to Ulp11-150-GFP. EtOH had no effect on Ulp1Δ150-340-GFP localization (Figure 2.4A). The region encompassing aa 150-340 is therefore necessary and sufficient for nucleolar recruitment in response to alcohol stress. Additionally, we have early evidence indicating the requirement of Kap95p in the nucleolar recruitment of Ulp1p (data not shown), however, the mechanism remains to be elucidated.
Figure 2.4. The region encompassing aa 150-340 is necessary and sufficient for nucleolar recruitment, and recruitment is energy-dependent.

(A) Cells expressing GFP-tagged Ulp1p truncation mutants (as indicated) were imaged before and after treatment with 10% EtOH for 60 min. Scale bar, 5 μm. (B) Ulp1p nucleolar recruitment is energy dependant. Strains expressing cNLS-RFP, Ulp1-GFP, Ulp1150-340-GFP or Ulp111-150-GFP were treated with 100mM deoxyglucose (D) and 10mM sodium azide (A) for 45 min (column DA), pretreated with DA for 45 min, then treated with 10% EtOH for 60 min (column DA EtOH), or pretreated with 10% EtOH for 60 min, then with DA for 45 min (column EtOH DA).
2.3.6. Ulp1p nucleolar recruitment is energy-dependent

We next wished to determine whether Ulp1p nucleolar recruitment is energy-dependent. As expected, treatment with 100mM deoxyglucose (D) and 10mM sodium azide (A) for 45 min resulted in equilibration of a classical nuclear localization signal (cNLS)-RFP protein throughout the nucleus and cytoplasm (excluding the vacuole, Figure 2.4B). As previously reported [145], treatment of cells expressing Ulp1-GFP with DA led to a decrease in nuclear rim signal, and an increase in cytoplasmic localization (Figure 2.4B). DA treatment of cells expressing Ulp1^{150-621}-GFP (data not shown) and Ulp1^{150-340}-GFP (Figure 2.4B) led to a complete loss of nuclear rim signal, indicating that aa 1-150 play a critical role in anchoring Ulp1p to the NPC, and that this process requires energy. DA treatment had no apparent effect on Ulp1^{1-150}-GFP localization (Figure 2.4B). In cells pre-treated with DA, then exposed to 10% EtOH, Ulp1-GFP was not recruited to nucleolar fragments (Figure 2.4B). Identical results were obtained for Ulp1^{150-621}-GFP (data not shown) and Ulp1^{150-340}-GFP (Figure 2.4B). Energy therefore appears to be required for Ulp1p recruitment to the nucleolus. In cells first exposed to 10% EtOH for 1 h, then treated with DA for 45 min, Ulp1p nucleolar foci remained mostly intact, suggesting that maintenance of Ulp1p in nucleolar fragments is less energy dependent.
2.4. Discussion

2.4.1. Regulation of SUMO conjugation

Here, we have established a paradigm for the control of SUMOylation: modulation of the intracellular localization of a SUMO protease. We propose that disengagement of Ulp1p from the NPC and recruitment to the nucleolus sequesters this SUMO protease from substrates to effect an increase in steady-state SUMO conjugate levels (Figure 2.5). While we do not yet understand the mechanism for Ulp1p nucleolar recruitment, proteins previously demonstrated to be important for Ulp1p localization at the NPC do not appear to be involved in the recruitment.

Figure 2.5. Model: Ulp1p is sequestered in the nucleolus in response to alcohol stress. (A) In untreated cells, Ulp1p is localized to the NPC, where it acts to remove SUMO from substrates. (B) In alcohol-stressed cells, Ulp1p is recruited to the nucleolus. Nucleolar sequestration limits access to substrates, resulting in a dramatic increase in steady-state SUMO conjugate levels.
2.4.2. The nucleolus, mRNA and stress

The nucleolus is a multifunctional nuclear subcompartment [156]. In addition to its roles in the assembly of ribosomal subunits, it takes part in other critical cellular processes, in part via the sequestration of regulatory proteins. For example, telomerase reverse transcriptase (TERT) is sequestered in the nucleolus until telomere replication [157], the VHL protein is sequestered in the nucleolus in response to hypoxia [158], and MDM2, a ubiquitin E3 ligase that catalyzes p53 degradation, is sequestered in the nucleolus in response to DNA damage [159]. Our data suggest that the nucleolus may also play an important role in regulation of the SUMO system.

The nucleolus has also been implicated in mRNA export and surveillance [156, 160]. Our data demonstrating that Nab2p and Tom1p are found along with Ulp1p in nucleolar fragments following alcohol stress are consistent with an important role of the nucleolus in the RNA stress response. It will be interesting to further characterize Ulp1p-containing nucleolar fragments, to better understand their possible role in RNA surveillance and the alcohol stress response.

2.4.3. Alcohol and the nuclear pore complex

Exposure to alcohol increases nuclear pore permeability [161]. It is unknown at present how our data may be related to this phenomenon, but we observed an inverse correlation between alcohol chain length and Ulp1p apparent solubility. Since NPC structure appears to be affected by alcohol, it may act as a sensor to trigger the alcohol-SUMO stress response. Further study will be required to better
understand the relationship between alcohol chain length, Ulp1p sequestration, and the role of the nucleolus as an alcohol stress sensor.
Chapter Three

MS/SUMmOn and Ub/Ubl conjugation site identification

3.1. Chapter Overview

This chapter describes an integrated MS methodology for the unambiguous identification of Ub- and Ubl-modified peptides, utilizing a recently developed software tool, SUMmOn, standard database search engines, two different types of MS/MS fragmentation methods, charge state exclusion, and a combination of proteolytic enzymes.

As described earlier, Ub and some Ubls can form multimeric chains. Ub multimers of varying lengths and linkage types are structurally distinct, interact with different subsets of binding proteins, and can therefore confer very different biological outcomes to a targeted substrate [7-10]. The biological functions and topologies of Ubl chains are not well characterized.
While MS has become a powerful tool for the study of many classes of PTMs, the identification of Ub/Ubl conjugation sites presents a set of special challenges. As described in the introduction, Ub conjugation site identification can be hampered by the presence of long C-terminal “tails” that remain covalently bound to target peptides following trypsin digestion (e.g. for SUMOs, Urm1p, and Hub1p). During CID, these Ubl tails produce multiple fragment ions. When combined with target peptide fragment ions, the resultant complex MS/MS spectra are not generally identified by database searching algorithms. Additionally, Ub and several other Ubls (e.g. NEDD8, Rub1p and ISG15) all possess an –(R)GG C-terminus. Following trypsin digestion, an identical –GG motif therefore remains conjugated to a target peptide lysine, and modifier identity is lost.

The SUMmOn software tool [54], which scans MS data for the presence of diagnostic PTM fragment ion series, has allowed for the identification of several SUMO conjugation sites [54]. This chapter introduces an integrated Ub/Ubl conjugation site identification strategy utilizing MS/SUMmOn analysis as a complement to standard database searching approaches. Using this methodology, a number of Ubl linkages and conjugation sites are characterized for the first time. Such linkages include alternative yeast and mammalian SUMO chain assemblies, alternative NEDD8 chain topologies, and several putative NEDD8 conjugation sites on the NEDD8 E2 enzyme UbcH12/UBE2M.

The material presented in this chapter is in preparation for submission for publication in Molecular and Cellular Proteomics.
3.1.1. Attributions

I would like to acknowledge Drs. Yi Sheng and Cheryl H. Arrowsmith for their kind gifts of recombinant E2 and E3 proteins, and Dr. R. Rogers for creating and testing the Smt3R93A protein, as illustrated in Figure 3.3. I would also like to highlight contributions made from Ph.D. candidate Tharan Srikumar, who kindly helped to generate data used in Figure 3.1D, Figure 3.4D and Figure 3.10.

3.2. Experimental Procedures

3.2.1. Materials

Lysyl Endopeptidase (LysC) was purchased from Wako Chemical (Japan), and TPCK (L-1-tosylamido-2-phenylethyl chloromethyl ketone)-treated sequencing grade trypsin was purchased from Promega (Madison, WI). HeLa S100 protein fraction was from Boston Biochem (Boston, MA). Analytical columns and pre-columns for LC-MS analysis were made from fused silica capillaries with a 75 μm (analytical column) or 100 μm (pre-column) inner diameter from InnovaQuartz (Phoenix, AZ), and were packed in-house with C_{18}–coated silica particles (Michrom Bioresources, Auburn, CA).
3.2.2. Smt3p mutant plasmid construction

PCR-mediated mutagenesis of a yeast SUMO expression plasmid [162] was conducted using the QuikChange site-directed mutagenesis system (Stratagene, La Jolla, CA), to convert the R93 codon AGA to an alanine codon, GCG. The mutation was confirmed by DNA sequencing.

3.2.3. In vitro Ub/Ubl reactions

Wild type and K11R mutant SUMO-2/3 proteins were purchased from Boston Biochem. Poly-SUMO-2/3 chains were generated in-house, or purchased from Boston Biochem. In vitro SUMOylation reactions were conducted as in [163]. In vitro NEDDylation reactions were conducted using 50 nM GST-NEDD8 E1, 0.4 μM GST-UbcH12 (both from Boston Biochem), 0.45 μM MDM2 as the E3 (a kind gift from C. Arrowsmith), 10 μM NEDD8 (Boston Biochem), 2 mM ATP, 5 mM MgCl₂, 50 mM NaCl, 50 mM Tris, pH 7.5. In vitro ubiquitylation reactions were conducted as in [164], using 20 nM Uba1, 0.6 μM E2D2, and 0.45 μM MDM2 (gifts from C. Arrowsmith). All reactions were incubated at 37°C for two hours. Unconjugated Ub/Ubl proteins were removed from the reaction mix (and the reaction buffer exchanged for ammonium bicarbonate; ammbic) using a microcon spin unit with a nominal molecular weight cutoff of 30 kDa (Millipore, Billerica, MA). After two ammbic pH 8.3 washes (400ul), the retentant was resuspended in 250ul ammbic and subjected to protease cleavage with 1 μg trypsin or LysC, overnight at 37°C.
3.2.4. Mass spectrometry and data analysis

Peptides were subjected to LC-ESI-MS/MS, using a 120 min HPLC reversed-phase buffer gradient [as in 54] running at 250nl-400nl/min on a Proxeon nanopump in-line with a hybrid LTQ-Orbitrap mass spectrometer (Thermo Fisher Scientific, Waltham, MA). A parent ion scan was performed in the Orbitrap, using a resolving power of 60,000. The three or four most intense peaks were selected for MS/MS (with a minimum ion count of 1000 for activation), using either CID or HCD (Higher-energy C-trap Dissociation) fragmentation. CID fragment ions were detected in the LTQ. HCD peptide fragmentation was performed with a normalized collision energy of 40%, 45% or 50%, followed by detection in the Orbitrap with a resolving power of 30,000. Dynamic exclusion was activated such that MS/MS of the same m/z (within a -0.1 and +2.1 Th window; Exclusion list size = 500) 3 times (or 6 times for the charge state exclusion analyses, see below) within 45 seconds were excluded from analysis for 30 seconds.

For protein identification, Thermo .RAW files were converted to the .mzXML format using readw software [165], then searched using X!Tandem [103] against yeast ORF (Saccaromyces Genome Database December 2005) or human (Ensembl 44.36F) databases, supplemented with a curated custom database containing the mature Ubl sequences. (Ub and most of the Ubls are translated as pro-proteins, and cleaved at a C-terminal diglycine motif to form the active or “mature” Ub/Ubl (e.g. see Figure 1.1A). The mature protein sequences are not represented in protein databases based on predicted open reading frames, and therefore must be added manually). X!Tandem search parameters were: log(e) < -2; complete modifications, none; cysteine modifications, none; potential modifications, +114@K (GG remnant), +16@M and W (oxidation), +32@M and W (dioxidation), +42@N-terminus (acetylated
N-terminus), +1@N and Q (deamidation); -17@N-term Q (pyroglutamate); -18@N-term E (pyroglutamate); parent mass error ±25ppm (parts per million); fragment error 0.4 Da; maximum charge 4+; missed cleavage sites 3; semi-cleavage, no. For determination of conjugation site location, MS/MS data were subjected to SUMmOn analysis [54] using a precursor mass error (i.e. SUMmOn digestion tolerance) of 25 ppm, and a fragment mass error (i.e. SUMmOn mass tolerance) of 1000 ppm.

3.3. Results

3.3.1. Alternative SUMO chain linkages

The mammalian SUMO-2 and SUMO-3 proteins form SUMO-SUMO conjugates (i.e. SUMO chains) via K11 [43, 53]. The yeast SUMO protein (Smt3p) forms chains primarily via K15 [52]. Earlier reports have suggested that mutation of these lysine residues dramatically decreases, but does eliminate, SUMO chain formation [e.g. 43]. To determine whether, like Ub, the SUMO proteins can form additional, alternative chain linkages, we subjected the products of in vitro yeast and mammalian SUMO reactions to MS/SUMmOn analysis.

3.3.1.1. SUMO-2 and SUMO-3

SUMO-2 in vitro SUMOylation reactions (as well as purchased, purified SUMO-2 chains; Boston Biochem) were treated with trypsin, and analyzed on an LTQ-Orbitrap (see Experimental Procedures). SUMmOn analysis of these data identified
SUMO-2 chains formed primarily through K11, but also identified SUMO-2 chains linked via K42 (Figure 3.1, Table 3.1. K42-linked SUMO-2 target peptide mass = 1054.61 Da. Target peptide sequence (R)HTPLSKLMK(A), corresponding to SUMO-2 amino acids 37-45). The number of K42-linked SUMO-2 spectra detected by SUMmOn accounted for ~3% of all chain spectra, with the K11-linked SUMO-2 spectra making up the remainder. (For example, in n=2 individual in vitro reactions, 2 MS runs/reaction, SUMmOn Modification score ≥0.8 Target score ≥0.8, number of K11 spectra = 1336, number of K42 spectra = 36.). Similar data were obtained for purchased, purified SUMO-2 chains (n=2 separate tryptic digests, two MS runs/reaction). Assuming up to one missed cleavage, and variable modifications at methionine (oxidation) and glutamine (pyroglutamate), no other predicted tryptic peptides in the reaction mix correspond to the m/z of the K42-linked SUMO-2-SUMO-2 peptide assigned by SUMmOn, within a parent mass error of 100 ppm. As expected, SUMmOn did not detect any K11 SUMO-2 linkages in an MS analysis of in vitro reactions containing a SUMO-2 K11R mutant protein (which cannot form K11-linked chains; Figure 3.1D, top and middle panels). However, the K11R mutation did not affect the m/z corresponding to the K42-linked SUMO-2 peptide assigned by SUMmOn (elution time ~71 min; Figure 3.1D, middle panel). While unmodified peptides derived from components of the reaction mix (e.g. the E1 SAE2 3+ peptide (R)KPPVPLDWAESQSQGEETNASDQQNPQQLGLK(D); Figure 3.1D, top, middle and bottom panels) were unchanged, SUMmOn did not detect any SUMO linkages in in vitro SUMO-2 reactions lacking only ATP (precluding the formation of SUMO conjugates; Figure 3.1D, bottom panel). We have thus identified an alternative, non-consensus SUMO-2 chain linkage.

Similarly, SUMmOn analysis of in vitro reactions containing the wild type SUMO-3 protein identified SUMO chains formed through both K11 and K41 (Figure 3.2, Table 3.1, SUMO-3 K41-linked target peptide mass = 1054.61 Da. Target peptide
Figure 3.1. K42-linked SUMO-2 is identified using MS/SUMmOn. (A) Amino acid sequence of the trypsin-digested K42-linked SUMO-2-SUMO-2 peptide, indicating the ion fragments assigned by SUMmOn. (B) SUMmOn-annotated CID spectrum of the K42-linked SUMO-2-SUMO-2 peptide highlighting b'- and y'-ions derived from the modification (i.e. the C-terminal tryptic peptide of SUMO-2, aa 61-92). (C) The same spectrum, highlighting the b- and y-ions derived from the target peptide (aa 37-45 of SUMO-2). (D) Extracted ion chromatograms (EIC) of three different representative SUMO-2 in vitro reaction mixes. The m/z corresponding to the SUMO-2 K11 and K42 linkages eluted at ~66 and ~72 min, respectively (top panel). The m/z corresponding to the SUMO-2 K11 linkage is lost when a K11R SUMO-2 mutant protein is used in the in vitro conjugation reaction (middle panel). The m/z corresponding to the K42 linkage is also lost in a reaction mix lacking ATP (bottom panel). Indicated are (i) the EIC of the SAE2 (SUMO E1) 3+ peptide (R)KPPVPLDAEVQSQGETNASDQONEFQLGLKD, eluting at ~59 min, using an m/z window of 1178.57-1178.59, (ii) an EIC of the 4+ (the most abundant charge state) SUMO-2 K11 linked peptide, using an m/z window of 1291.59-1259.61, and (iii) the EIC of the 4+ SUMO-2 K42 linked peptide (the most abundant charge state), using an m/z window of 1152.29-1152.31. For very large peptides, such as those characterized here, the monoisotopic variant represents only a very small percentage of the peptide population. The m/z windows in Figures 2, 4 and 7 therefore do not bracket the monoisotopic m/z for each linkage, but instead encompass the most abundant isotopic variants (i.e. those that were most often isolated for fragmentation).
<table>
<thead>
<tr>
<th>Chain Linkage</th>
<th>Peptide Sequence</th>
<th>Target Peptide (M+H)+</th>
<th>Observed Conjugate m/z</th>
</tr>
</thead>
<tbody>
<tr>
<td>SUMO-2 K11 (aa 8-21)</td>
<td>(K7)EGVKTENNDHINLK22(V)</td>
<td>1610.81</td>
<td>(3+) 1720.79 (4+) 1290.85 (5+) 1032.88 (6+) 860.90</td>
</tr>
<tr>
<td>SUMO-3 K11 (aa 8-20)</td>
<td>(K6)EGVKTENNDHINLK21(V)</td>
<td>1496.77</td>
<td>(3+) 1682.77 (4+) 1262.34 (5+) 1010.07 (6+) 841.89</td>
</tr>
<tr>
<td>SUMO-2 K42 (aa 37-45)</td>
<td>(R36)HTPLSKLMK22(A)</td>
<td>1054.61</td>
<td>(3+) 1535.39 (4+) 1151.79 (5+) 921.64</td>
</tr>
<tr>
<td>SUMO-3 K41 (aa 36-44)</td>
<td>(R35)HTPLSKLMK21(A)</td>
<td>1054.61</td>
<td>(3+) 1535.38 (4+) 1151.79</td>
</tr>
<tr>
<td>Smt3p* K54 (aa 48-56)</td>
<td>(R32)LMEAFAKR32(Q)</td>
<td>965.52</td>
<td>(3+) 1305.28 (4+) 979.21</td>
</tr>
<tr>
<td>NEDD8 K22 (aa 12-27)</td>
<td>(K11)EIEIDIEPTDKVERIK22(E)</td>
<td>1927.04</td>
<td>(4+) 871.50 (5+) 697.40</td>
</tr>
<tr>
<td>NEDD8 K27 (aa 23-33)</td>
<td>(K22)VERIKERVEEK23(E)</td>
<td>1414.80</td>
<td>(4+) 743.44 (5+) 594.95</td>
</tr>
<tr>
<td>NEDD8 K33 (aa 28-48)</td>
<td>(K22)ERVEEKEGIPQQQRLIYSGK34(Q)</td>
<td>2484.32</td>
<td>(4+) 1010.82 (6+) 674.22</td>
</tr>
<tr>
<td>NEDD8 K48 (aa 34-54)</td>
<td>(K33)EGIPQQQRLIYSGMKMNDEK34(T)</td>
<td>2459.24</td>
<td>(4+) 1004.55 (5+) 803.85 (6+) 670.04</td>
</tr>
<tr>
<td>NEDD8 K54 (aa 49-60)</td>
<td>(K32)Q17.02605MNDEKTAADYK16(I)</td>
<td>1396.61</td>
<td>(3+) 984.86 (4+) 738.90</td>
</tr>
<tr>
<td>Ubiquitin K48 (aa 34-63)</td>
<td>(K30)EGIPPDQQRLIFAGKQLEDGRTLSDYNIQK30(E)</td>
<td>3429.78</td>
<td>(4+) 1216.16 (5+) 973.13 (6+) 811.11 (7+) 695.38</td>
</tr>
</tbody>
</table>

**Table 3.1.** Ub/Ubl linkages identified by SUMmOn.

Observed monoisotopic m/z values (mean from four MS scans) for the identified Ub/Ubl linkages. *Bold face font* denotes modified lysine residue. *Smt3p R93A mutant. SwissProt numbers; SUMO-2 (P62988), SUMO-3 (P55854).
sequence (R)HTPLSKLMK(A), corresponding to SUMO-3 amino acids 36-44). The SUMO-3 K41 linkage spectra represented ~6% of all chain product spectra assigned by SUMmOn in \textit{in vitro} SUMOylation reactions, with K11 spectra accounting for the remainder (\textit{e.g.} in n=2 separate \textit{in vitro} reactions, two MS runs/reaction, SUMmOn Modification score $\geq 0.8$, Target score $\geq 0.8$, number of K11 spectra = 356, number of K41 spectra = 23). Similar data were obtained for purchased, purified SUMO-3 chains (n=2 separate tryptic digests, two MS runs/reaction).

In reactions containing a SUMO-3 K11R mutant protein, the $m/z$ assigned by SUMmOn as the K11-linked SUMO-3 peptide was not detected. This mutation, however, had no effect on the $m/z$ assigned by SUMmOn as the K41-linked SUMO-3 peptide. An \textit{in vitro} reaction lacking only ATP also lacked the $m/z$ assigned to the K11- and K41-linked SUMO-3 chains (data not shown). The human SUMO-3 protein can thus also form at least two types of SUMO chain linkages \textit{in vitro}. Importantly, X!Tandem was unable to identify any SUMO-2 or SUMO-3 linkage type in any of these reactions, presumably due to the long C-terminal tryptic tails.

\textbf{3.3.1.2. Yeast SUMO}

The C-terminal-most tryptic cleavage site in the 98 aa \textit{S. cerevisiae} SUMO (Smt3p) protein is at R93 (Figure 1.1A). Cleavage of a yeast SUMO-modified target protein with trypsin thus yields a 5 aa Smt3p remnant covalently bound to a lysine group within the target peptide. SUMmOn can successfully detect this 5aa "tail" [54]. However, an Smt3p R93A mutant (the C-terminal tryptic cleavage site is now at R71, and tryptic cleavage of this protein results in a 27aa "tail"; Figure 1.1A) is much more efficiently identified by SUMmOn, due to the presence of many more "reporter" fragment ions. We did not observe any difference in the ability of this
Figure 3.2. K41-linked SUMO-3 is identified using MS/SUMmOn.
(A) Amino acid sequence of the trypsin-digested K41-linked SUMO-3-SUMO-3 peptide, indicating fragments assigned by SUMmOn. (B) SUMmOn-annotated CID spectrum of the LysC-digested K41-linked SUMO-3-SUMO-3 peptide, highlighting b'- and y'-ions derived from the modification (i.e. the C-terminal SUMO-3 LysC peptide). (C) The same spectrum, highlighting the b- and y-ions derived from the target peptide (SUMO-3 amino acids 36-44).
point mutant to be conjugated to substrates: e.g. Smt3R93A was conjugated *in vitro* to a C-terminal RanGAP1 protein fragment (aa 420-589) as efficiently as the wild type protein (Figure 3.3). This protein was thus used to determine if, like the mammalian SUMOs, the yeast SUMO protein can form alternative chain topologies.

SUMmOn analysis of an Smt3R93A-containing *in vitro* reaction identified a SUMO-SUMO linkage at K54 (Figure 3.4, Table 3.1. Target peptide mass = 965.52 Da. Target peptide sequence (R)LMEAFAKR(Q), corresponding to aa 48-55 of Smt3p). As above, no other predicted tryptic peptide in the reaction mix corresponds to this mass and charge state within a parent mass error of 100 ppm. In an identical reaction mix lacking only ATP, the \( m/z \) assigned by SUMmOn as the K54-linked peptide was not detected (Figure 3.4D). We thus also identified a new non-consensus SUMO chain linkage site for the yeast SUMO protein.

---

**Table 3.1.**

<table>
<thead>
<tr>
<th></th>
<th>E1,E2 and ATP</th>
<th>Smt3p WT</th>
<th>Smt3p R93A</th>
<th>GST-NΔ419 RanGAP1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
</tbody>
</table>

---

**Figure 3.3. Conjugation of Smt3R93A to RanGAP1.**

Coomassie stained SDS-PAGE gel highlighting the conjugation of the wild type Smt3p and the Smt3R93A mutant protein to a GST-RanGAP1 NΔ419 substrate.
Figure 3.4. An Smt3p K54-linkage is identified using MS/SUMmOn.  
(A) Amino acid sequence of the K54-linked Smt3p-Smt3p peptide, indicating the fragments assigned by SUMmOn.  
(B) SUMmOn annotated CID spectrum of an Smt3p-Smt3p conjugate highlighting $b'$- and $y'$-ions derived from the modification (the C-terminal Smt3p tryptic peptide, aa 72-98).  
(C) The same spectrum, highlighting the $b$- and $y$-ions derived from the target peptide (Smt3p aa 48-55).  
(D) EIC of representative in vitro Smt3p conjugation reactions with and without ATP. The m/z windows correspond to the 3+ ($m/z = 1305.94-1305.96$, elution at ~61 min) and 4+ ($m/z = 974.71-979.72$) precursor ions of the Smt3p K54-linked peptide. An unmodified Smt3p 3+ peptide (\textit{MSDSENVQEAKPEVKPEVKPETHINLK(V)} ($m/z$ window 1026.52-1026.54, eluting at ~25 min) was detected in both reactions.
3.3.2. NEDD8 chain linkages

Trypsin is the enzyme of choice for most MS-based proteomics analyses. Following trypsin cleavage, Ub-conjugated peptides are modified at a lysine residue with a –GG (diglycine) moiety. Standard database search tools can efficiently identify peptides bearing this modification. However, several other Ubls (e.g. human NEDD8 and ISG15, and S. cerevisiae Rub1p) also possess C-terminal –(R)GG amino acid sequences (Figure 1.1A). Following trypsin proteolysis, these Ubl-derived -GG remnants are therefore indistinguishable from a Ub-derived -GG modification, resulting in a loss of modifier identity. Treatment of the same Ub/Ubl conjugates with LysC (which cleaves only at the C-terminal side of lysine) yields target peptides conjugated to longer, and therefore more informative, Ub/Ubl “tails”. LysC cleavage of Ub results in a 13 aa C-terminal peptide, while LysC cleavage of NEDD8, Rub1p or ISG15 yields distinct 16 aa, 23 aa and 14 aa C-terminal peptides, respectively (Figure 1.1A). Standard database search tools are generally unable to identify target peptides conjugated to these longer Ubl peptide fragments [14, 54]. While we have previously demonstrated that SUMmOn can efficiently identify SUMO targets and conjugation sites, it had not previously been tested on Ub, or Ubls other than SUMO.

To determine whether SUMmOn can identify other Ubl conjugation sites, we generated in vitro NEDDylation reactions, using the NEDD8 E1 (APPBP1/Uba3), E2 (UbcH12/UBE2M) and MDM2 as the E3. Half of the sample was subjected to trypsin cleavage, and the other half proteolyzed with LysC. The resulting peptides were analyzed by both SUMmOn and X!Tandem. Previous reports have indicated that in vitro NEDDylation reactions produce K48- and K22-linked NEDD8 chains [56, 62]. SUMmOn also identified both K22- and K48-linked NEDD8 chains in the LysC-
treated reaction mix (Figure 3.5, Table 3.1). The K48 linkage (Target peptide mass = 2459.24 Da. Target peptide sequence (K)EGIPQQQRLYSGKQMNDEK(T), corresponding to aa 34-54 of NEDD8) accounted for ~44% of all observed NEDD8 chain linkage spectra, while K22 linkages (Target peptide mass = 1927.04 Da. Target peptide sequence (K)EIEIDIEPTDVERIK(E), corresponding to aa 12-27 of NEDD8) accounted for ~3% of chain spectra assigned by SUMmOn (for example, in n=4 separate in vitro NEDDylation reactions, two MS runs/reaction, SUMmOn Mod and Target scores ≥0.8, number of K48 linkage spectra = 151, number of K22 spectra = 11). Interestingly, SUMmOn also identified a previously unreported NEDD8 chain linkage at K54 (Figure 3.6, Table 3.1. Target peptide mass = 1396.61 Da. Target peptide sequence (K)Q*MNDEKTAADYK(I), *pyroglutamate, corresponding to aa 49-60 of NEDD8), representing ~41% of the NEDD8 chain spectra (in the same four NEDD8 reactions, number of high scoring K54 linkage spectra = 140). The m/z assigned by SUMmOn as the K54-linked NEDD8 peptide could not be assigned to any other predicted peptide in the reaction mix, within 100ppm. An identical reaction mix lacking ATP also lacked the m/z corresponding to the NEDD8 K54 linkage, and SUMmOn did not assign any NEDD8 conjugates in this reaction (data not shown). X!Tandem was unable to identify any NEDD8 linkages in the LysC-digested samples (data not shown), but successfully identified both the NEDD8 K48 and K54 linkages in the reaction mix digested with trypsin (Figure 3.7). Thus, using a combination of SUMmOn and standard database searching, we were able to unambiguously identify an abundant, alternative NEDD8 chain linkage site.

Two less abundant, previously unreported NEDD8 linkages were also detected by SUMmOn, at K27 (Table 3.1. Target peptide mass = 1414.80 Da. Target peptide sequence (K)VERIKERFEEEK(E), corresponding to NEDD8 aa 23-33) representing ~9% of the chain products (31 spectra), and K33 (Table 3.1. Target
Figure 3.5. K22- and K48-linked NEDD8 is identified using MS/SUMmOn.
(A) Amino acid sequence of the LysC-digested K22-linked NEDD8-NEDD8 peptide, indicating fragments identified using SUMmOn. (B) SUMmOn-annotated CID spectrum of the LysC-digested K22-linked NEDD8-NEDD8 peptide highlighting the $b'$- and $y'$-ions derived from the modification (the C-terminal LysC fragment of NEDD8, aa 61-76). (C) Identical CID spectrum, highlighting the $b$- and $y$-ions derived from the target peptide (NEDD8 aa 12-27). (D-F) As described for A-C, but for K48-linked NEDD8 peptides.
Figure 3.6. K54-linked NEDD8 is identified using MS/SUMmOn.
(A) Amino acid sequence of the LysC-digested K54-linked NEDD8-NEDD8 peptide, indicating fragments identified using SUMmOn. (B) SUMmOn-annotated CID spectrum of the LysC-digested K54-linked NEDD8-NEDD8 peptide highlighting the $b'$ and $y'$-ions derived from the modification (the C-terminal LysC fragment of NEDD8, aa 61-76). (C) Identical CID spectrum, highlighting the $b$- and $y$-ions derived from the target peptide (NEDD8 aa 49-60). Asterisk (*) denotes pyroglutamate.
peptide mass = 2484.32 Da. Target peptide sequence

(K)ERVEEKEGIPQQQRLIYSKG(Q), corresponding to NEDD8 aa 28-48)
representing ~3% (12 observations) of the chain spectra. X!Tandem did not identify these linkages.

The K33-linked NEDD8 peptide was highly charged (mostly 6+), and did not give rise to highly informative spectra under standard CID conditions. This NEDD8-linked peptide fragmented preferentially at the peptide bond located between L62 and G63 of the NEDD8 C-terminal “tail”. The y₁⁵ fragment ion thus accounted for the majority of the total ion current in these CID spectra. However, CID spectra of the unconjugated NEDD8 C-terminal tail

(K)ILGGSVLHLVLALRGG(-) produce a similar fragmentation pattern, with preferential y₁⁴ fragmentation, increasing our confidence in the identification of this minor linkage type by SUMmOn. Overall, these data indicate that, like Ub, NEDD8 is capable of forming chains on many different lysine residues in vitro.
3.3.3. NEDDylation of UbcH12/UBE2M

In the same set of *in vitro* reactions, SUMmOn also identified several NEDD8 modification sites on the E2 enzyme UbcH12/UBE2M. These NEDD8 modification sites all occurred within the lysine-rich ~28aa N-terminal extension of UbcH12, at K8, K11, K12, K21, K25 and K26 (Figure 3.8). None of these conjugation sites have been previously reported, and none were detected in trypsin digests of the same *in vitro* reactions, as analyzed by X!Tandem, presumably due to the small size of the GG-modified tryptic peptides (Figure 3.8B and C, Table 3.2), which largely precludes their detection in the standard 400-1800 m/z range window used in typical ion trap instrument MS parent ion scans. The longer LysC- generated NEDD8 tail results in much larger peptides, allowing them to be trapped and analyzed by the mass spectrometer. Thus another advantage of using LysC digestion and SUMmOn analysis is that it can allow for the identification of Ubl conjugation sites on small target peptides that may otherwise go undetected.

3.3.4. Identification of Ub/Ubl-modified peptides in more complex mixtures

We next tested whether SUMmOn could distinguish between peptides modified by Ub, NEDD8 and SUMO-2/3 in the same sample. Ub and NEDD8 *in vitro* reactions were digested with LysC, and SUMO-2 chains were digested with trypsin. A mixture was prepared with relative amounts of Ub, NEDD8 and SUMO-2 chains yielding equal base peak intensities of the m/z values corresponding to the predominant linkage (i.e. Ub K48, NEDD8 K48 and SUMO-2 K11). This mixture was
Figure 3.8. NEDD8 modification sites identified in UbcH12/UBE2M using MS/SUMmOn. (A) Amino acid sequence of UbcH12/UBE2M. The 28 aa N-terminal extension is highlighted in red, and NEDD8-modified lysine residues identified in this study are in bold. (B and C) Schematic diagram of a NEDD8-modified UbcH12/UBE2M peptide (aa 4-11; NEDD8-modified K8), cleaved with (B) trypsin, or (C) LysC. Predicted monoisotopic m/z for each peptide at several different charge states are indicated. Grey m/z values fall below those used in typical ion trap mass spectrometer parent ion scans (m/z 400-1800).
<table>
<thead>
<tr>
<th>NEDD8-UbcH12 Linkage</th>
<th>Peptide Sequence</th>
<th>Predicted m/z NEDD8-modified LysC Peptide</th>
<th>Predicted m/z GG-modified Tryptic Peptide</th>
</tr>
</thead>
<tbody>
<tr>
<td>K8 (aa 4-11)</td>
<td>(K8)LFSLKQK11(K)</td>
<td>(2+) 1274.28 (3+) 849.86 (4+) 637.64</td>
<td>(2+) 553.32 (3+) 369.22 (4+) 277.16</td>
</tr>
<tr>
<td>K11 (aa 9-12)</td>
<td>(K8)QQKK12(E)</td>
<td>(2+) 1044.14 (3+) 696.43 (4+) 522.58</td>
<td>(2+) 323.19 (3+) 215.79 (4+) 162.10</td>
</tr>
<tr>
<td>K12 (aa 12-21)</td>
<td>(K11)KEESAGGTK21(G)</td>
<td>(2+) 1296.23 (3+) 864.49 (4+) 648.62</td>
<td>(2+) 575.27 (3+) 383.85 (4+) 288.14</td>
</tr>
<tr>
<td>K21 (aa 13-25)</td>
<td>(K12)EEESAGGTKGSSK25(K)</td>
<td>(2+) 1411.77 (3+) 941.52 (4+) 706.39</td>
<td>(2+) 690.81 (3+) 460.88 (4+) 345.91</td>
</tr>
<tr>
<td>K25 (aa 22-26)</td>
<td>(K21)GSSK26(A)</td>
<td>(2+) 1031.63 (3+) 688.09 (4+) 516.32</td>
<td>(2+) 310.67 (3+) 207.45 (4+) 155.84</td>
</tr>
<tr>
<td>K26 (aa 26-36 or aa 26-33)</td>
<td>(K23)KASAAQLRIQK36(D) - LysC or (K25)KASAAQLR33(I) - Trypsin</td>
<td>(2+) 1385.35 (3+) 923.90 (4+) 693.18</td>
<td>(2+) 479.77 (3+) 320.19 (4+) 240.39</td>
</tr>
</tbody>
</table>

Table 3.2. NEDD8-modified LysC and tryptic UbcH12/UBE2M peptides.  
Grey text indicates those ions that would not be detected in a precursor ion scan with a window of 400-1800 m/z. Bold text denotes NEDD8-modified lysine residues.

subjected to MS, and Ub K48-linked chains (Table 3.1. Target peptide mass 3429.78 Da. Target peptide sequence (K)EGIPPDQQRLIFAGKQLEDGRTLSDYNIQK(E) corresponding to Ub aa 34-63), NEDD8 K48- and K54-linked chains, and SUMO-2 K11-linkages were identified by SUMmOn (SUMmOn Modification score ≥0.8 Target score ≥0.8). SUMmOn was also able to differentiate between co-eluting Ub/Ubl conjugates: the Ub K48-linked peptide and the SUMO-2 K11-linked peptide elute at ~ 65 min (Figure 3.9), and both linkage types were successfully identified (SUMmOn Modification score ≥0.8, Target score ≥0.8). Thus, SUMmOn was able to successfully identify multiple Ub/Ubl linkages in a more complex sample, and
Figure 3.9. Elution profiles of K48-linked Ub and K11-linked SUMO-2 peptides.
(A) EIC of a representative MS run performed on a mixture of in vitro SUMO-2 and Ub conjugation reactions, using the indicated \( m/z \) windows for the K11-linked SUMO-2 (top panel) and K48-linked Ub (bottom panel) peptides. The SUMO-2 K11 and Ub K48 linked peptides both elute at ~65 min. (B) MS scan performed at 65.16 min, illustrates co-elution of SUMO-2 K11 and Ub K48 linkages. Indicated are the 4+, 5+, 6+ and 7+ K48-linked Ub precursor ions, along with 4+ and 5+ K11-linked SUMO-2 precursor ions.

Unlike standard trypsin/database searching approaches, the use of longer, more informative LysC-generated Ub/Ubl tails allows for unambiguous identification.

In typical shotgun proteomics approaches, only a small percentage of the total number of peptides may be sequenced in highly complex mixtures. Any
strategy to enrich for specific subsets of peptides may thus be extremely useful. Our data indicated that a majority of the Ub/Ubl-linked peptides generated by LysC cleavage are highly charged (all are $\geq 3+$, most are $\geq 4+$). We found that this property can be used to enrich for Ub/Ubl CIDs. The use of 1+, 2+ and 3+ charge state exclusion on the LTQ-Orbitrap dramatically decreases the number of CID scans taken for lower charged (and in this case, irrelevant) peptides, and increases the chance that more highly charged peptides will be selected for fragmentation. For example, analysis of an in vitro SUMO-2 reaction mix indicated a significant decrease in the number of CID scans collected when 1+, 2+ and 3+ parent ions were excluded (5785 MS scans, 1200 MS/MS scans), as compared to a standard run in which only 1+ charged ions were excluded (6019 MS scans, 5700 MS/MS scans), with no loss (or even a slight improvement) in the number of K11 (200 vs. 168) and K42 (11 vs. 6, respectively) SUMO-2 linkage identifications.

To test this approach on a more complex protein mixture, 25ng of SUMO-2 chains (corresponding to $\sim$2 pmol of the K11-linked SUMO-2 peptide) was spiked into 3ug of a HeLa S100 protein fraction. This mix was digested with trypsin and subjected to MS analysis using 1+, 2+ and 3+ charge state exclusion (n=2). The base peak intensity of the 4+ SUMO-2 K11-linked peptide was very low (4E5; 4+ was the most abundant charge state) in comparison to the average base peak intensity of the lysate fraction (7E6; Figure 3.10). However, this peptide was still selected for CID fragmentation and successfully identified by SUMmOn. Thus, SUMmOn is able to identify Ub/Ubl conjugates in more complex samples, and using a combination of various charge state exclusion settings may be advantageous for the identification of Ubl conjugates.
Figure 3.10. Extracted ion chromatograms of a representative MS run of a HeLa S100 lysate spiked with SUMO-2 chains. 
*top panel* Normalized base peak intensity chromatogram, using 1+, 2+ and 3+ charge-state exclusion (note the low number of CID taken in this run).  *middle panel* EIC of the SUMO-2 K11 peptide, using the same normalized ion intensity scale.  *bottom panel* EIC of the SUMO-2 K11 peptide, with normalized ion intensity scale set to 4E5 ion counts.

### 3.3.5. HCD fragmentation improves identification of some highly charged peptides

The LysC-derived K48-linked Ub peptide is highly charged (4+ to 7+; Figure 3.9B) under standard reversed-phase HPLC buffer conditions. The 5+, 6+ and 7+ precursor ions were predominantly selected for CID in our analyses (n=4 separate *in vitro* reactions, two MS runs/sample). While SUMmOn readily identified the 5+ K48-linked Ub peptide (Modification score ≥0.8, Target score ≥0.8), standard CID fragmentation settings resulted in poor fragmentation of the 6+ (Figure 3.11A-C) and 7+ (data not shown) precursors. The y\textsubscript{275}\textsuperscript{5+} and y\textsubscript{276}\textsuperscript{6+} ions account for a majority of the total ion current in these spectra, representing fragmentation at the peptide bond.
Figure 3.11. LysC-digested K48-linked Ub is identified by MS/SUMmOn, using CID and HCD. (A) Amino acid sequence of the K48-linked Ub-Ub peptide, indicating fragments identified by SUMmOn. (B) SUMmOn-annotated CID spectrum of a Ub-Ub conjugate highlighting $b'$- and $y'$-ions derived from the modification (the Ub C-terminal LysC peptide, aa 64-76). (C) Identical CID spectrum, highlighting the $b$- and $y$-ions derived from the target peptide (Ub aa 34-63). In this case, the majority of the total ion current is contributed by the $y_{56+}$ and $y_{55+}$ ions, resulting from the predominant fragmentation on the N-terminal side of P37 (as indicated by the arrow in A). (D-F) As described for A-C, but with the use of HCD fragmentation.
between I36 and P37. Increasing the CID normalized collision energy or activation Q did not result in any additional informative fragmentation events (data not shown). In an attempt to improve the fragmentation of this peptide at higher charge states, LysC-treated *in vitro* Ub reactions were subjected to Higher-energy C-trap Dissociation (HCD), in which the C-trap in the Orbitrap is used as a pseudo-collision cell [166]. HCD was tested using normalized collision energies of 40%, 45% and 50% (data not shown). We found that a normalized collision energy of 45% resulted in a more informative fragmentation of the LysC-derived Ub K48-linked peptide at higher charge states (Figure 3.11D-F), and SUMmOn readily identified this linkage site using HCD data. Thus, HCD and SUMmOn analysis can be used to identify Ubl-linked peptides that do not fragment well under standard CID conditions.

It should be noted that while HCD was extremely useful for the identification of the K48-linked Ub peptide, it was not as useful for the lower charged NEDD8 and SUMO-2/3 linked peptides, yielding a lower signal-to-noise ratio than CID, as a result of an increased number of fragmentation events (data not shown). An additional HCD fragmentation analysis may therefore be informative for some types of samples, or the fragmentation method may be tailored to a specific Ubl of interest.

3.4. Discussion

3.4.1. An integrated MS/SUMmOn-based methodology for the identification of Ub/Ubl conjugation sites

Taking all of the above observations into account, we propose an improved methodology for the identification of Ub and Ubl conjugates in complex biological
mixtures (Figure 3.12). A sample containing Ub/Ubl-modified peptides may be
digested in two separate reactions with trypsin and LysC (or another site-specific
endoproteinase that yields longer Ub/Ubl C-terminal tails). If a reduction/alkylation
step is required, chloroacetamide should be used as the alkylating agent [127]. Both
the trypsin- and LysC-treated samples should then be subjected to MS using
standard CID conditions (inclusion of 2+ and higher precursor ions). Standard
database searching can be used to identify putative GG-modified and unmodified
peptides, and this peptide list can be used to build a protein subdatabase for
SUMmOn analysis. Next, both the trypsin- and LysC-treated samples can be
subjected to MS analysis using charge state exclusion, where only 4+ or higher
precursor ions are fragmented. An optional HCD fragmentation may be added for
improved fragmentation of some highly charged precursor ions. The output from
standard database searches and SUMmOn analyses of these multiple datasets can
then be combined and used to unambiguously identify Ub/Ubl conjugates and
conjugation sites.

3.4.2. New Ubl chain linkages

While relatively few SUMO conjugation sites have been identified to date, more than
40% of those that have been published occur at a non-consensus lysine residue
(Table 1.2; [167]). All of the SUMO chain linkages that we have identified here occur
at non-consensus lysine residues. The use of algorithms developed for the
prediction of putative SUMOylation sites [39-42] that depend upon the consensus
sequence therefore do not identify these linkage sites (data not shown).
Figure 3.12. Improved workflow for the identification of Ub/Ubl-conjugation sites. (A) The protein sample of interest is split in two, and digested with trypsin or LysC prior to LC-MS/MS analysis. (B) Standard database searching is used to analyze a portion of the trypsin-digested sample, and to create a subdatabase for SUMmOn. (C) The remainder of the trypsin-digested sample, as well as the LysC-digested sample, are analyzed using SUMmOn. During the initial LC-MS/MS run, CIDs are collected on precursor ions with a charge of 2+ or greater. An additional LC-MS/MS run can be performed in which CIDs are collected on precursor ions with a charge of 4+ or greater to enrich for Ub/Ubl-modified peptides. HCD may also be used to obtain improved fragmentation of highly charged precursor ions. Combining the results of these multiple analyses can generate a comprehensive list of unambiguous Ub/Ubl modification sites.
Consensus sequences have not been described for the other Ubls. Identification of Ubl conjugation sites has thus generally been carried out via extensive labor-intensive mutational analysis. However, mutational analysis of a putative Ub/Ubl conjugate can be significantly hampered by the promiscuity of the E1/E2/E3 conjugation machinery: i.e. mutation of a single lysine residue may have little or no effect on overall Ubl conjugation of a given substrate, since a nearby lysine residue may be targeted instead [168]. The identification of Ub/Ubl modification sites is therefore probably best conducted by MS.

Our analysis of yeast SUMO chains revealed a K54 linkage. Bylebyl et al. [92] demonstrated that yeast strains lacking one of the two SUMO proteases (Ulp2p) accumulate high molecular weight SUMO conjugates (presumably SUMO chains), and display hypersensitivity to hydroxyurea (HU). HU hypersensitivity is alleviated by replacing the endogenous SMT3 gene with a sequence coding for a SUMO protein that is unable to form chains (Smt3p AllR). Replacing the endogenous gene with a SUMO coding sequence lacking the primary SUMO-SUMO conjugation site (Smt3p K11,15,19R) partially alleviates HU sensitivity. Interestingly, an Smt3p K54,58R mutant was also able to partially alleviate HU sensitivity. Consistent with our in vitro data, these in vivo data suggest that multiple types of SUMO chains may exist in yeast, and that the K54 linkage may play an important (but unknown) physiological role.

3.4.3. Limitations of the approach

LysC proteolysis of e.g. a K63-linked Ub chain results in a very large, multiply branched Ub structure, which cannot be identified by SUMmOn or database searching. However, standard database searching can detect trypsin-cleaved K63-
linked Ub peptides. SUMmOn analysis also relies on a protein “subdatabase” (containing e.g. all of the proteins known to be present in a given *in vitro* reaction, or a protein list generated by a standard database search of the sample in question), to decrease CPU usage, and minimize false positive identifications. Since the “T-shaped” Ubl conjugated peptides can be very large (some are >5 kDa), the monoisotopic peak normally represents only a small fraction of the peptide population: a large majority of peptides of this size contain multiple $^{13}$C atoms. Thus, while the high mass accuracy of the LTQ-Orbitrap allows for an accurate determination of the monoisotopic mass of the parent, the monoisotopic peak itself is only rarely selected for fragmentation. Since the fragments generated by CID or HCD are not derived from the monoisotopic parent, and the isotope content and/or distribution of the parent is not known *a priori*, fragment masses can vary within a wide mass window, and do not match the predicted monoisotopic fragments. The SUMmOn fragmentation mass window used in these types of analyses must therefore remain large. This fact effectively limits the practical maximum number of proteins that can be included in the subdatabase. We have found that SUMmOn can accurately identify SUMO-modified peptides using a subdatabase containing up to ~200 proteins (data not shown). An increase in false-positive rates occurs when using larger subdatabases. While this limit does not affect the analysis of simple samples such as *in vitro* Ub/Ubl reactions, or even lower complexity samples such as immunoprecipitations, it could represent a limitation for the analysis of highly complex biological samples such as serum or a cell lysate. The LysC/SUMmOn analysis strategy outlined here should therefore be used as a complement to the standard trypsin/database search approach.
Chapter Four

Concluding remarks

4.1. Future directions on SUMO and the stress response

Alcohol stress results in a dramatic change in Ulp1p localization, which can easily be detected using confocal microscopy (Chapter 2). Since this change in localization has an undefined mechanism, it would be beneficial if a large-scale screen was conducted to identify the proteins involved. Fortunately, *S. cerevisiae* is manipulated relatively easily and existing methodologies allow for high-throughput handling of this model organism. For example, a number of libraries exist for *S. cerevisiae*, including a knockout library consisting of ~4700 deletion strains [169, 170], a temperature sensitive (ts) mutant library consisting of ~500 mutants [22], and several epitope-tagged collections. We have utilized fluorescently tagged Ulp1p for detection of NPC or nucleolar localized Ulp1p, which is easily distinguishable using this technology. Therefore, introduction of fluorescently tagged Ulp1p into the ordered deletion library may help identify proteins involved in the nucleolar
sequestration of Ulp1p in response to alcohol. It would be extremely labor-intensive to image each of these ~4700 strains manually, and thus an automated imaging approach would be optimal. High-content screening technologies have been developed to provide a high-throughput methodology for the exploration of a particular biological question (e.g. OPERA by PerkinElmer Inc.). The Raught laboratory is currently exploring this methodology to image the deletion strain library expressing Ulp1-RFP, in the presence and absence of alcohol. These strains also contain Nop2-GFP for mapping of the nucleolus. An overlay of GFP and RFP signals allows for a simple co-localization test. Deletion strains displaying disruption of Ulp1p nucleolar sequestration following alcohol treatment (i.e. those in which the GFP and RFP signals do not co-localize) will allow us to identify proteins required for Ulp1p nucleolar recruitment. Further characterization of these proteins will be required to answer questions regarding how the nucleolus may act as an alcohol sensor, and how the SUMO pathway is incorporated into this stress response.

In addition to the sequestration of Ulp1p in response to alcohol stress, the nucleolus has also been implicated in mRNA export and surveillance [156, 160]. As stated in Chapter 2, Nab2p and Tom1p were found to co-localize with Ulp1p in nucleolar fragments following alcohol stress. A better understanding of the proteins sequestered in the nucleolus in response to alcohol stress may help us to better understand the role of the nucleolus in the regulation of the SUMO system. This can be conducted by introducing Ulp1-RFP into the yeast GFP library, which contains ~4000 strains containing C-terminally GFP-tagged yeast ORFs. This can then be imaged using high-content screening methodology in the absence/presence of alcohol to identify proteins that co-localize with sequestered Ulp1p in response to alcohol stress. Additionally, mass spectrometry can be used to identify the proteins in the Ulp1p-containing nucleolar fragments.
Finally, the data presented in Chapter 2 was generated using *S. cerevisiae*, therefore examination of a mammalian system should next be conducted. Ethanol is produced by yeast and is thus physiologically relevant. Therefore, one can speculate that the involvement of SUMO in the alcohol stress response in yeast may be specific to this organism. However, reports have also implicated the mammalian SUMO system in various stress responses, including heat, oxidative, genotoxic and ethanol [138, 139]. In particular, increases in SUMO-2/3 conjugate levels have been observed in response to ethanol [139], similar to that described in Chapter 2. Mammalian cells possess six different SUMO proteases (SENPs; see Section 1.2.1). It may be interesting to investigate whether one or more SENP(s) are involved in the mammalian alcohol stress response, and whether the nucleolus is involved in SENP sequestration.

### 4.2. Future directions for the identification of Ub/Ubl conjugation sites

Over the past several years significant advances have been made in MS and proteomics technology. Most notably, mass spectrometers have become more accurate and sensitive (e.g. ThermoFisher’s LTQ-Orbitrap) and improvements have been made in peptide identification software. For example, the Trans-Proteomic Pipeline (TPP; [105, 171-173]) allows for rigid statistical analysis and validation of peptides identified using MS/MS. However, with these notable developments there has been a lag in software advancements for identification of Ub/Ubl conjugation sites. As a result, the field continues to simply use a mass shift of +114.04 Da (corresponding to the mass of the C-terminal –GG remnant produced upon trypsin
digestion) for the identification of Ub, NEDD8, and ISG15 conjugation sites, even though this approach is plagued with problems (see Chapter 3). To date, the SUMmOn pattern recognition tool has been the only software developed to identify Ubl-modified peptides [54]. More recently, another attempt was made to identify SUMO conjugation sites using a “virtual peptide” manipulation [53]. Matic et al. illustrated that to calculate the m/z of a modified peptide, one can assume that the modification and target peptides are a single, linear sequence (see reference [53]). Although this approach successfully identified several SUMO-modified peptides, this method is limited by the sequence of the target peptide. For instance, if the modified lysine residue occurs in the middle of a long target peptide (as was the case for LysC-digested K48-linked Ub; see Chapter 3), fragmentation of the “virtual peptide” may not be easily detected. This, along with other limitations, illustrates the shortcomings of such an approach. Overall, this indicates that further development should be made to the MS/SUMmOn methodology, making it more amenable with in vivo samples.

The MS/SUMmOn methodology was initially tested and optimized using in vitro generated samples, which contained ample amounts of Ub/Ubl-modified peptides. Upon in vitro optimization, the next logical step is to optimize the MS/SUMmOn methodology using more relevant biological material (e.g. in vivo sample preparations) to test whether it can be potentially used for more complicated samples (e.g. tissue biopsies, etc.). Ubl-modified proteins are transient and nonstoichiometric and often represent <1% of the total protein population [85, 136], making it extremely difficult to identify these modifications and often requires some form of enrichment, in order for the mass spectrometer to select these peptides for fragmentation. Therefore, concerns regarding the use of the MS/SUMmOn methodology on in vivo samples are related to the quality of the sample and the sensitivity of the mass spectrometry and not necessarily with the SUMmOn software.
component of the methodology. SUMmOn software does indeed have limitations, especially with regards to the size of the “subdatabase” and false-discovery rates (FDR). Therefore, at present an enrichment of Ub/Ubl-modified peptides must be performed in order to keep FDR to a minimum. As described in Chapter 3, charge state exclusion is beneficial in reducing the number of uninformative CIDs performed, however, sample preparation/enrichment will most likely remain important.

Increasing the amount of Ubl-modified peptides in a sample does not guarantee detection by MS. Mass spectrometers can analyze approximately three orders of magnitude, in terms of dynamic range of peptide intensity [174]. If a Ub/Ubl-modified peptide falls below this range, it may not be identified. A complex sample (e.g. cell lysate or serum) contains peptides with a wide range of abundances, typically resulting in the missed detection of low intensity peptides. Therefore, approaches have been developed to reduce sample complexity, either off-line, or on-line with the mass spectrometer. A typical off-line approach utilizes SDS-PAGE to separate proteins based on molecular mass. The resultant gel lane is divided into several pieces and each piece is individually proteolytically digested and analyzed. Another popular approach is Multi-dimensional Protein Identification Technology (MudPIT) [175], which utilizes a biphasic column containing strong-cation exchange (SCX) media followed by standard reversed-phase C18 media. The peptides in the sample first bind to the SCX column bed, and are eluted in a stepwise fashion onto the C18 material, using increasing salt bumps. Standard reversed-phase chromatography can then be performed. Either approach can be used to reduce the complexity of the sample, allowing the mass spectrometer to identify less abundant peptides.

A single approach may not be sufficient for the identification of Ubl-modified peptides in vivo using the MS/SUMmOn methodology and it is likely that a
combination of techniques will be required. For example, reducing the complexity of a sample by using MudPIT or an off-line fractionation approach, in combination with the use of charge state exclusion, may be beneficial. As stated above, enrichment of Ub/Ubl-modified proteins or peptides will remain important for their identification using MS. Enrichment is typically performed by immunoprecipitation of epitope-tagged Ub/Ubls (described in Chapter 1). However, rather than enriching for Ub/Ubl-modified proteins, it would also be beneficial if more Ub/Ubl-modified proteins can be produced. For instance, Bylebyl et al. reported that there was an accumulation of SUMO chains in a Ulp2Δ strain [92]. Furthermore, as described in Chapter 2, the effect of alcohol, and other stressors), also results in an increase in SUMOylation levels. These approaches or combinations thereof may be fruitful for the identification of in vivo Ub/Ubl-modified peptides by MS/SUMmOn analysis.

Chapter 3 identified several alternative yeast and mammalian SUMO chain topologies in addition to several NEDD8 linkages. These linkages were identified using in vitro generated samples. Therefore, it will be important to determine whether or not these linkages exist in vivo. As described earlier in Chapter 3, there is already some evidence for the existence of yeast SUMO K54 chains in vivo [92]. However, the linkage has not been confidently identified, either through MS or immunoblot analysis. SUMO and NEDD8 chains have only been recently identified and remain poorly understood. Thus, even upon in vivo identification of these Ubl chains by MS/SUMmOn, further characterization will be required to elucidate their biological functions within the cell.
4.3. General Conclusions

This thesis outlines a SUMO pathway regulatory mechanism and provides insight into the use of MS for the identification of SUMO and other Ub/Ubl conjugation sites. The work presented in this thesis helps expand the understanding of the SUMO family of proteins and other members of the Ubl family of modifiers.

The study on the response of the SUMO system to alcohol stress resulted in a greater understanding of how the SUMO pathway can be regulated, through the modulation of the intracellular localization of the SUMO protease, Ulp1p. Furthermore, this study described Ulp1p sequestration to the nucleolus, implicating this subcompartment in an important role in the regulation of the SUMO system. Finally, this study makes a possible connection between Ulp1p-containing nucleolar fragments and RNA surveillance and the alcohol stress response, implicating the SUMO system in this process. As is often the case in science, arriving at an answer creates new questions. Linking the SUMO system with the alcohol stress response through the sequestration of Ulp1p in the nucleolus will promote further study of this interesting relationship.

MS has emerged as a powerful tool for the study of PTMs, which has expanded to include protein modification by Ub and the Ubls. Identification of certain Ubl conjugation sites (e.g. SUMOylation sites) has proven difficult, and has evaded identification through standard database software. Through the use of SUMmOn, a previously developed pattern recognition tool, several alternative SUMO chain linkages have been identified, expanding the implications of SUMO chain architecture in SUMO function. Additionally, several NEDD8 chain linkages have been identified using MS/SUMmOn, suggesting that the formation of NEDD8 chains may be involved in the function of this modifier, as is the case with Ub and
SUMO. Additionally, through the use of LysC, HCD fragmentation, and charge state exclusion, the unambiguous identification of Ub/Ubl-modified peptides through MS/SUMmOn has been made possible. Overall, an integrated-MS/SUMmOn methodology has been proposed, and has been suggested to complement standard database software for the identification of Ub/Ubl conjugation sites, which will help expand our understanding of these critical, yet poorly understood PTMs.
References


REFERENCES


REFERENCES


Saccharomyces cerevisiae and shows dynamic interplay with positive-acting histone modifications. *Genes Dev* 2006, 20, 966-976.


91


REFERENCES


REFERENCES


[115] Spengler, B., De novo sequencing, peptide composition analysis, and composition-based sequencing: a new strategy employing accurate mass


REFERENCES


REFERENCES


REFERENCES


