uOttawa

L'Université canadienne
Canada's university

FACULTÉ DES ÉTUDES SUPÉRIEURES ET POSTOCTORALES

uOttawa

L'Université canadienne
Canada's university

FACULTY OF GRADUATE AND POSDOCTORAL STUDIES

Julian Vasilescu
AUTEUR DE LA THÈSE / AUTHOR OF THESIS

Ph.D. (Biochemistry)
GRADE / DEGREE

Department of Biochemistry, Microbiology and Immunology
FACULTÉ, ÉCOLE, DÉPARTEMENT / FACULTY, SCHOOL, DEPARTMENT

Proteomic Analysis of Ubiquitinated Proteins Employing Mass Spectrometry-Based Techniques

TITRE DE LA THÈSE / TITLE OF THESIS

Daniel Figeys
DIRECTEUR (DIRECTRICE) DE LA THÈSE / THESIS SUPERVISOR

CO-DIRECTEUR (CO-DIRECTRICE) DE LA THÈSE / THESIS CO-SUPERVISOR

Alexandre Blais

Jean-François Couture

David Chen (University of British Columbia)

Douglas Gray

Gary W. Slater
Le Doyen de la Faculté des études supérieures et postdoctorales / Dean of the Faculty of Graduate and Postdoctoral Studies

# PROTEOMIC ANALYSIS OF UBIQUITINATED PROTEINS EMPLOYING MASS SPECTROMETRY-BASED TECHNIQUES

Julian Vasilescu

Thesis submitted to the Faculty of Graduate and Postdoctoral Studies in partial fulfillment of the requirements of the degree of Doctor of Philosophy (Ph.D.)

Department of Biochemistry, Microbiology and Immunology

Faculty of Medicine

University of Ottawa

Ottawa, Ontario, Canada

*Your file Votre référence*
*ISBN:* 978-0-494-66257-1
*Our file Notre référence*
*ISBN:* 978-0-494-66257-1

**Abstract**

Post-translational modification of proteins via the covalent attachment of Ubiquitin plays an important role in the regulation of protein stability and function in eukaryotic cells. In the second chapter of my thesis, a novel method for identifying ubiquitinated proteins from a complex biological sample, such as a whole cell lysate, using a combination of immunoaffinity purification and liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis is described. The applicability of this approach is demonstrated by the identification of seventy ubiquitinated proteins from the human MCF-7 breast cancer cell line after treatment with the proteasome inhibitor MG132.

In the third chapter of my thesis, an application of a microfluidic protein processing device, termed the Proteomic Reactor, for enzymatic digestion of proteins for subsequent LC-MS/MS analysis is described. Use of the Proteomic Reactor following immunoaffinity purification enabled the identification of numerous low abundance ubiquitinated proteins from the human 1299 lung cancer cell line expressing reduced amounts of the ubiquitin-dependent chaperone known as the Valosin-Containing Protein.

In the fourth chapter of my thesis, a method for determining peptide ion score cutoffs that permit the confident identification of ubiquitinated proteins by MS/MS is described. Experiments involving the analysis of gel bands containing multi-Ubiquitin chains with quadrupole time-of-flight and quadrupole ion trap mass spectrometers revealed that standard ion score cutoffs used for database searching after MS/MS analysis were not sufficiently stringent. False positive and false negative rates were also found to vary significantly depending on the cutoff scores used and that appropriate cutoffs could only be determined following a systematic evaluation of false

positive rates. When standard ion score cutoffs were used for the analysis of complex mixtures of ubiquitinated proteins, unacceptably high false positive rates were observed. Finally, false positive rates for ubiquitinated proteins were found to be affected by the size of the protein database that is searched.

In chapters five and six of my thesis, two biological applications employing the methods I developed for identifying ubiquitinated proteins are described. The first application involves the identification of specific lysine residues within the membrane-anchored transcription factor Mga2p that are targets of the E3 ubiquitin ligase Rsp5p in *Saccharomyces cerevisiae.* Ubiquitination of these lysines were found to mediate mobilization of the Mga2p transcription factor complex and its subsequent disassembly. The second application involves the identification of proteins that are ubiquitinated upon mitotic exit in *Xenopus laevis* embryos. Sequence analysis of these proteins revealed that the majority of poly-ubiquitination occurs through lysine-11 chain polymerization, which is an atypical form of ubiquitination.

**Acknowledgements**

First and foremost, I would like to thank my supervisor Dr. Daniel Figeys for giving me the opportunity to pursue a Ph.D. in his laboratory. I would like to thank him for his time, advice and encouragement over the years. I would also like to thank my parents for their help and support and their unconditional love. I thank them for providing me with all of the opportunities and teaching me the skills to be successful in life. I hope my accomplishments make them proud. Finally, I would like to thank past and present members of Dr. Figeys' lab, especially Dr. Jeffrey C. Smith and Mr. Nicholas J. Denis.

**Table of contents**

**List of abbreviations**

| | |
|---|---|
| Ab | antibody |
| BSA | bovine serum albumin |
| DMP | dimethyl pimelimidate |
| DNA | deoxyribonucleic acid |
| DTT | dithiotreitol |
| ECL | enhanced chemiluminescense |
| EDTA | ethylenediaminetetraacetic acid |
| FPR | false positive rate |
| FNR | false negative rate |
| HPLC | high performance liquid chromatography |
| IP | immunoaffinity purification |
| kDa | kilodalton (molecular mass) |
| LC | liquid chromatography |
| mAb | monoclonal antibody |
| MS | mass spectrometry |
| MS/MS | tandem mass spectrometry |

| | |
|---|---|
| m/z | mass to charge ratio |
| NP-40 | nonidet P-40 |
| PAGE | polyacrylamide gel electrophoresis |
| PBS | phosphate buffered saline |
| PTM | post-translational modification |
| QIT | quadrupole ion trap |
| QTOF | quadrupole time-of-flight |
| RT | room temperature |
| SDS | sodium dodecyl sulfate |
| TAP | tandem affinity purification |
| TFA | trifluoroacetic acid |
| Tris | tris(hydroxymethyl) aminomethane |
| Ub | ubiquitin |
| VCP | valosin-containing protein |
| WB | western blot |
| WCL | whole cell lysate |
| WT | wild-type |

**List of figures**

**List of tables**

# Chapter 1: Introduction

## 1.1 Early history of ubiquitin and the ubiquitin-proteasome pathway

In the mid 1950's, Christian de Duve discovered a membrane-bound organelle known as the lysosome, which was recognized in rat liver as a vacuolar structure that contains hydrolytic enzymes that function optimally at low pH.[1,2] This discovery, in combination with studies conducted around the same time which demonstrated that proteins are in a constant state of synthesis and degradation, led to the conclusion that the lysosome was likely responsible for mediating the degradation of all intracellular proteins.[3] Although the structure of the lysosome explained the separation required for hydrolytic enzymes and their protein substrates, and autophagy explained the mechanism of entry of proteins into the lysosomal lumen, increasing evidence accumulated over the next two decades suggested that the degradation of at least certain classes of intracellular proteins occurred in a non-lysosomal manner.[4]

Progress in identifying a non-lysosomal degradation system was hindered during this period primarily due to the lack of an *in vitro* system that could mimic proteolytic degradation in a specific and energy-dependent manner. It was not until 1977 that Etlinger and Goldberg were able to isolate and characterize an extract prepared from rabbit reticulocytes that selectively degraded abnormal hemoglobin, required energy (ATP), and functioned optimally at neutral pH.[5] Since reticulocytes are terminally differentiated red blood cells which do not contain lysosomes, the proteolytic activity was concluded to be of non-lysosomal origin. A similar system was also isolated and characterized a year later by Hershko *et al.*[6] (who was awarded the Nobel Prize in Chemistry in 2004 with Ciechanover and Rose for the discovery of the Ub-proteasome pathway).

In initial experiments, Ciechanover and Hershko fractionated a crude reticulocyte cell extract using an anion-exchange resin (DEAE-cellulose) into two components, both of which were required to reconstitute the energy-dependent proteolytic activity of the crude extract.[7] The unabsorbed column flow-through was designated as Fraction 1, while all proteins adsorbed to the resin and eluted with high salt was designated as Fraction 2. The active component from Fraction 1 was a small, 8.5 kDa heat-stable protein, that was termed ATP-dependent Proteolysis Factor 1 (APF-1). In subsequent experiments, purified APF-1 was radiolabeled and incubated with Fraction 2 in the presence or absence of ATP.[8] ATP-dependent binding of APF-1 to high molecular weight proteins was then observed by gel filtration chromatography. Treatment of these high molecular weight complexes with acid, alkali, hydroxylamine, or boiling with sodium dodecyl sulfate and mercaptoethanol did not affect their stability, thus indicating a covalent linkage between APF-1 and these proteins.

The discovery that APF-1 stimulates proteolysis in the presence of ATP and Fraction 2 and is covalently linked to protein substrates, led to the proposal of a model in 1980 by Ciechanover, Hershko and Rose that multiple moieties of APF-1 target a substrate for degradation by a downstream protease that does not recognize the unmodified substrate.[9] That same year, Wilkinson *et al.* confirmed that APF-1 was a previously characterized protein known as Ubiquitin (Ub),[10] which was purified in 1974 during the isolation of thymopoetin from the thymus and was initially thought to be present in all living cells.[11,12] Interestingly, Ub was thought to have lymphocyte-differentiating properties and was named UBiquitious Immunopoietic Polypeptide (UBIP). However, later studies revealed that it was not involved in the immune response and its expression was restricted to eukaryotes.[4]

Another important discovery in the Ub field was that a single Ub moiety was conjugated to histones, specifically to histone H2A. The structure of Ub-H2A was elucidated by Goldknopf and Busch in 1977, who found that they are linked together through an isopeptide bond between the carboxy-terminal glycine residue of Ub (G76) and the ε-$NH_2$-group of an internal lysine of histone H2A (K119).[13] Over the next few years, additional studies by Ciechanover, Hershko and Rose revealed that the isopeptide bond of the Ub-H2A conjugate was identical to the covalent linkage between a single Ub moiety and a protein substrate and between multiple Ub moieties within a poly-Ub chain that is recognized by the downstream protease.[14,15]

The identification of APF-1 as Ub and the discovery that an isopeptide bond links Ub to a protein substrate enabled the complex mechanism of isopeptide bond formation to be resolved. Ciechanover and Hershko found that Ub is ligated to a protein substrate not by a single enzyme (as was originally proposed in their 1980 model, with the putative enzyme termed APF-1 protein amide synthetase), but by the sequential action of three enzymes. These enzymes are the E1 activating enzyme, the E2 conjugating enzyme, and the E3 ligating enzyme. E1 was found to catalyze an ATP-dependent activation of the carboxy-terminal glycine of Ub by the formation of Ub adenylate, followed by the transfer of activated Ub to a thiol site of E1 with the formation of a thiolester linkage.[16,17] Activated Ub is transferred to a thiol site of E2 by transacylation, and is then further transferred to an amino group of the protein substrate in a reaction that requires E3.[18] They also found that the role of E3 is to bind specific protein substrates[19] and based on this observation, it was proposed in 1988 that the selectivity of Ub-mediated protein degradation was mainly determined by the substrate specificity of different E3 enzymes.[20] Subsequent studies verified the selectivity of numerous E3 enzymes on their substrates.[21]

The remaining component of the Ub-mediated protein degradation system to be identified was the downstream protease responsible for recognizing and degrading poly-ubiquitinated proteins. In 1986, Hough *et al.* purified and characterized a high-molecular weight alkaline protease that degraded Ub-conjugated lysozyme, but not unmodified lysozyme, in an ATP-dependent manner.[22] This discovery was confirmed a year later by Waxman *et al.* who further characterized the protease as a large, 1.5 megadalton enzyme (later termed the 26S proteasome), which differed from all other known proteases.[23] In a subsequent study, Hough *et al.* reported the existence of a smaller 20S multi-subunit protease[24] that was similar to a previously characterized protease known as the Multicatalytic Proteinase Complex (MPC), which was purified in 1980 from the bovine pituitary gland.[25] This 20S complex was found to be ATP-independent and cleaved the carboxy-terminal side of hydrophobic, basic and acidic amino acid residues. Based on this observation, it was proposed that the 20S complex was a component of the larger 26S proteasome.[24] Subsequent studies confirmed that the 20S complex is the core catalytic component of the 26S proteasome.[26,27]

In 1992, Hoffman *et al.* purified an additional component of the 26S proteasome, termed the 19S regulatory complex.[28] By mixing the catalytic 20S catalytic complex and the 19S regulatory complex, they were able to generate an active 26S proteasome that degraded poly-ubiquitinated proteins into smaller peptides. The structure of the 20S complex was found to be barrel-shaped and is composed of four stacked rings consisting of two outer $\alpha$ rings and two inner $\beta$ rings.[29] In eukaryotes, the $\alpha$ and $\beta$ rings are composed of seven subunits, giving the 20S complex the general structure of $\alpha_{1\text{-}7}$, $\beta_{1\text{-}7}$, $\beta_{1\text{-}7}$, $\alpha_{1\text{-}7}$. The catalytic activity is localized to three $\beta$ rings, the $\beta_1$, $\beta_2$ and $\beta_5$, which are described as having chymotryptic-like, tryptic-like and post-glutamyl peptidyl hydrolytic-like activities.[29,30]

Each extremity of the 20S complex is capped by a 19S regulatory complex that is composed of 17 different subunits. The 19S regulatory complex recognizes ubiquitinated proteins and opens an orifice in the α ring that allows entry of the substrate into the proteolytic chamber. Since a folded protein is unable to fit through the proteasomal channel, it is assumed that the 19S regulatory complex unfolds substrates and inserts them into the 20S complex. Both the channel opening and unfolding functions are energy-dependent, which likely accounts for the presence of six different ATPase subunits.[29,30] Following degradation of the substrate, short peptides derived from the substrate are released as well as free and reusable Ub molecules.

## 1.2 Proteasome inhibitors: Peptide aldehydes and MG132

Due to the groundbreaking work by Ciechanover, Hershko, Rose and others, it was shown that the Ub-proteasome pathway is responsible for degrading the majority of intracellular proteins, including cell cycle proteins and their regulators, transcription factors and their inhibitors, and antigenic proteins. Examples include the mitotic cyclins and the p21 and p27 inhibitors of the cyclin-dependent kinases, the p53 tumor suppressor, the IκB inhibitors of the NF-κB transcription factor, the breast cancer 1 early onset protein BRCA1, and the angiogenesis regulator HIF-1α.[31-35] Not surprisingly, the proteasome has become an attractive target for drug development for a host of human diseases and conditions, including cancer, neurodegenerative diseases and autoimmune disorders.[36-39]

The most characterized and widely-used drugs that target the proteasome belong to a class of compounds known as competitive inhibitors. These compounds function by mimicking proteasomal substrates in their structure, charge and hydrophobicity.[40] Competitive inhibitors

bind specific subunits of the 20S complex covalently or non-covalently, act in a reversible or non-reversible manner, may target a specific active site with high-affinity, or may block all three peptidase activities with a similar binding constant. Although the chymotryptic-like activity is most often targeted, competitive inhibitors have also been developed that are specific for the tryptic-like and post-glutamyl peptidyl hydrolytic-like activities.[40,41]

Peptide aldehydes were the first competitive inhibitors used for studying the proteasome. They are relatively easy to synthesize and are known to rapidly permeate the cell membrane.[41] Peptide aldehydes typically have three distinguishing features; 1) an amine protecting group on their amino-terminus, such as Carboxybenzyl (Cbz); 2) a tri- or tetra-peptide sequence and; 3) an aldehyde group on their carboxy-terminus. The amine protecting group functions as an electrophilic trap that binds the active site residues of the β subunits of the 20S complex in a manner resembling the tetrahedral transition state or the acyl-enzyme intermediate. It is for this reason that peptide aldehydes are often referred to as transition state inhibitors.[40] The aldehyde group forms a covalent hemiacetal with the active threonine hydroxyl in a reaction that is reversible under physiological conditions, and form a stable oxazolidine ring by reaction with the amino-terminal amine and an accessible side chain hydroxyl.[40,42,43]

The peptide aldehyde MG132, also known as Cbz-leu-leu-leu-al or Cbz-leu-leu-leucinal, was one of the first proteasome inhibitors used in cell culture experiments that displayed excellent potency and acceptable specificity. This tri-peptide derivative was used in several studies to demonstrate that the proteasome is responsible for degrading many important substrates, including p53, IκB, and the membrane channel protein CFTR. MG132 is a potent inhibitor of all three peptidase activities of the proteasome, with the highest affinity towards the chymotrypsin-like active site ($K_i$ = 4 nM).[40] MG132 is more potent than the tri-peptide derivative

MG115, also known as Cbz-leu-leu-norval-al or Cbz-leu-leu-norvalinal, which has a $K_i$ of 21 nM. Although MG132 also inhibits the enzymes calpain and cathepsin, they require a concentration that is an order of magnitude higher than is required for disrupting proteasomal activity *in vivo*. Despite this limitation in specificity, the availability and low cost of MG132 has made it one of the most widely-used proteasome inhibitors to date.[41,43]

## 1.3 Poly-ubiquitin chains

Ub contains seven conserved lysine residues within its amino acid sequence (K6, K11, K27, K29, K33, K48 and K63), all of which may be involved in the synthesis of poly-Ub chains. The two most common types of poly-Ub chains are those involving K48 and K63. K48-linked poly-Ub chains are typically found on substrates that are destined for proteasomal degradation, whereas poly-Ub chains linked through K63 appear to be associated with processes that do not involve proteolysis. For example, K63-linked poly-Ub chains are known to be involved in various cellular processes including the DNA damage response, the inflammatory response, protein trafficking and ribosomal protein synthesis.[44]

In the case of the DNA damage response, ubiquitination occurs on a DNA polymerase processivity factor known as the Proliferating Cell Nuclear Antigen or PCNA, which trimerizes and encircles the DNA template during replication. Conjugation of a single Ub moiety (mono-Ub) on lysine-164 of PCNA results in an error-prone mode of DNA lesion bypass through the recruitment of a translesion synthesis polymerase, which inserts nucleotides opposite damaged bases.[45] Synthesis of a K63-linked poly-Ub chain results in an error-free mode of DNA lesion bypass, which is thought to occur due to a switch in the template DNA strand.[46] In the case of the

inflammatory response, ubiquitination plays a role in the activation of the transcription factor NFκB and the IκB kinase. Synthesis of a K63-linked poly-Ub chain on TRAF6 allows the recruitment of specific kinase adaptors through their zinc finger motifs and the resulting TRAF6 complex activates IκB kinase. This in turn, leads to phosphorylation of the NFκB inhibitor IκBα and subsequent synthesis of a K48-linked poly-Ub chain on IκB. The targeting of IκB for proteasomal degradation frees NFκB, which is then able to translocate to the nucleus.[47]

Although poly-Ub chains linked through other lysine residues have been identified, their functional consequences remain less characterized than K48- and K63-linked chains. In a recent study, it was shown that the E3 ligase AIP4 synthesizes K29-linked poly-Ub chains on the Notch signaling modulator DTX, which targets it for lysosomal degradation.[48] In another study, it was shown that two AMPK-related kinases, AMPK-related kinase 5 and microtubule-affinity-regulating kinase 4, are modified with K29- and/or K33-linked poly-Ub chains, which appear to regulate their activity.[49] The BRCA1 tumor suppressor synthesizes a K6-linked poly-Ub chain on itself (auto-ubiquitination) when complexed with a protein known as BARD1. This poly-Ub chain is recognized by the 26S proteasome and is subsequently removed, but degradation of BRCA1 does not occur. This suggests a stabilization function of K6-linked poly-Ub chains rather than a role in proteasomal degradation.[50] Even less is known about the consequences of K11-linked poly-Ub chains, which have been reported to be the third most abundant type found in yeast after K48- and K63-linked chains.[51] K11-linked chains have not yet been implicated in any signaling pathways, although one study has demonstrated that it targets substrates for proteasomal degradation *in vitro*.[52] The functional consequences of mixed poly-Ub chain linkages in involving K48-, K63- and the other possible lysine branching sites is also a field that requires further study.

## 1.4 E3 ligases, E4 elongation factors and poly-ubiquitin chain synthesis

The various types of poly-Ub chains stem from the existence of a large number of enzymes and co-factors that catalyze their synthesis. It is now known that poly-ubiquitination involves at least two E1 activating enzymes, dozens of E2 conjugating enzymes, and hundreds of E3 ligases.[53] E3 ligases can be classified into at least two major types on the basis of their catalytic domains: 1) the Homologous to E6-associated protein C-terminus (HECT) domain E3s; and 2) the Really Interesting New Gene (RING) finger E3s. The HECT domain was identified in 1993 by Scheffner *et al.* in the E6-AP protein, which associates with the human papillomavirus E6 gene product to ubiquitinate and target the p53 tumor suppressor for degradation.[54] A catalytic cysteine residue in the HECT domain accepts Ub from an E2 enzyme before transferring it to protein substrates. In contrast, the RING E3s do not form covalent intermediates with Ub and function as a scaffold to position substrates in close proximity to an E2-Ub complex, which facilitates the direct transfer of Ub from E2 to the protein substrate.[55]

HECT E3 ligases have a modular structure consisting of a 350 amino acid carboxy-terminal domain that is homologous to the carboxy-terminus of E6-AP and a variable amino-terminal domain that is responsible for substrate recognition and binding.[36] Some HECT E3s have amino-terminal WW domains, which are peptide motifs characterized by two conserved tryptophan residues that form hydrophobic pockets. The WW domain binds substrates with phosphorylated serine or threonine residues as well as proline-rich motifs.[56,57] For example, the WW domain of the E3 ligase NEDD4, binds proline-rich subunits of sodium channels and ubiquitinates them. Mutations in these proline-rich motifs that prevent NEDD4 binding in epithelial cells results in a hypertensive disorder called Liddle syndrome, in which the activity of sodium channels are enhanced due to reduced proteasomal degradation.[58]

RING finger E3s contain a characteristic structure composed of conserved histidine and cysteine residues in complex with two zinc ions.[36] The RING finger E3s can be categorized into single or multi-subunit proteins. Two examples of single subunit RING finger E3s are the MDM2 oncoprotein which is known to ubiquitinate p53, and the c-CBL protein which ubiquitinates receptor tyrosine kinases such as the EGFR.[59,60] Examples of multi-subunit RING finger E3s belong to the Skp1-Cullin1-F-box (SCF) protein family that contain a RING finger domain component known as Rbx1, which recruits the E2-Ub complex, and the Cullin proteins, which act as a scaffold between Rbx1 and the proteins involved in substrate selection.[36] Substrate selection and recognition is mediated through the F-box proteins which are themselves, recruited to the SCF Ring finger E3s by the Skp1 adaptor protein.[61]

E4 elongation factors assist both HECT and RING finger E3s in the synthesis of poly-Ub chains. E4 enzymes are characterized by a conserved C-terminal U-box domain of roughly 70 amino acids that is structurally related to the RING finger motif. [62] In 1999, the first E4 enzyme known as Ubiquitin fusion degradation 2 (UFD2) was identified. It was shown to bind substrates conjugated to one to three Ub molecules and catalyses the addition of Ub moieties to yield poly-Ub substrates that are targets for the 26S proteasome.[63] In 2005, Richly *et al.* demonstrated the *in vivo* relevance of UFD2 in a study of the AAA ATPase Cdc48 (also known as p97 or VCP) and the transcription factor Mga2p (also known as Spt23), which is anchored to the endoplasmic reticulum membrane as an inactive dimer.[63] Mono-ubiquitination of inactive Mga2p by the HECT E3 ligase Rsp5p and cleavage by the 26S proteasome generates the activated 90 kDa form of Mga2p which then translocates to the nucleus. In the nucleus, Mga2p90 is a target for poly-ubiquitination by UFD2 which actives Mga2p90 and leads to the transcription of genes involved in fatty acid synthesis.

Single UFD2-homologues have been identified in *Saccharomyces cerevisiae*, *C. Elegans*, *Dictyostelium* and *Arabidopsis thaliana*, while two UFD2 related genes (UFD2a and UFD2b) have been identified in *Mus Musculus* and in humans. As mentioned above, UFD2 is involved in fatty acid synthesis but it is also known to be implicated in cell survival under stress conditions.[64] The UFD2 homologue in *Dictyostelium* known as NosA is required for cellular differentiation as NosA mutants do not develop fruiting bodies.[65] Interestingly, *in vitro* and *in vivo* experiments demonstrated that mammalian UFD2a functions as an E4 enzyme and promotes the degradation of pathological forms of ataxin-3 that are responsible for spinocerebellar ataxia type-3, which is a neurodegenerative disorder also known as Machado-Joseph disease.[66]

## 1.5 Methods of purifying ubiquitinated proteins

Previous studies with the aim of characterizing and identifying ubiquitinated proteins were largely hindered by the fact that ubiquitinated proteins have a high turnover rate and that the modification is reversed by the action of de-ubiquitinating enzymes (DUBs), which are proteases that recognize Ub and cleave the isopeptide bond between carboxy-terminal glycine residue (G76) and the amino group of a lysine residue.[67] The use of epitope tags for capturing and enriching proteins proved to be a major breakthrough that enabled researchers to overcome these obstacles. Epitope tags function as immunological binding sites for antibodies or as affinity binding sites for specific affinity resins (e.g. antibody-coupled protein G/A-agarose beads). Influenza virus hemagglutinin (HA), Myc and Flag are three short peptide epitope tags that are recognized by commercially available monoclonal antibodies.[68] HA-tagged Ub, Myc-tagged Ub and Flag-tagged Ub can be expressed in eukaryotic cells to form Ub-conjugates *in vitro* and *in*

*vivo*. Polyhistidine-tagged Ub and biotin-labeled Ub have also been used to form Ub conjugates. Polyhistidine-tagged Ub was expressed in bacteria and purified for *in vitro* ubiquitination studies and expressed in eukaryotic cells for *in vivo* studies.[69] Biotinylation of Ub as well as other proteins is known to modify the amino groups of lysine residues and hinders the formation of poly-Ub chains, making it less desirable tag for *in vivo* ubiquitination studies.[70]

Unlike the other peptide epitope tags mentioned above, proteins tagged with polyhistidine are purified using a strategy known as immobilized metal ion affinity chromatography (IMAC), which is based on the specific coordinate binding of amino acids (particularly histidine) to metal ions, such as nickel, cobalt, copper, zinc or iron.[71] IMAC resins such as nitriloacetic acid (NTA)-agarose or iminodiacetic acid (IDA)-agarose are used to load and bind the metal ions prior to purification. The use of IMAC to purify polyhistidine-tagged Ub offers an important advantage over other epitope tagging strategies in that it can be performed in the presence of denaturants (i.e. buffers containing high salt and detergent concentrations) to reduce non-specific binding to the affinity resin.[70] However, a major drawback of this strategy is the co-purification of histidine-rich proteins, which is especially problematic in mammalian cells where the population of histidine-rich proteins is significant.[72]

In 2003, Gururaja *et al.* reported a proteomics-based analysis of ubiquitinated proteins by purifying and incubating 6xHis-Ub with HeLa cell lysates in the presence of ATP. Purification of 6xHis-Ub conjugates with a nickel ($Ni^{2+}$)-charged IMAC resin enabled the identification of 80 proteins by liquid chromatography-tandem mass spectrometry (LC-MS/MS), which included 22 proteasome subunits or associated proteins, 18 Ub-proteasome pathway enzymes, 4 ubiquitin domain (UBD)-containing proteins and 36 proteins associated with redox processes, endocytosis/vessicle trafficking, the cytoskeleton, DNA repair, calcium binding and mRNA

splicing.[73] That same year, Peng *et al.* reported a proteomics-based analysis of ubiquitinated proteins from a yeast strain expressing only 6xHis-Ub.[51] 6xHis-Ub conjugates were purified, digested in solution, and analyzed by two dimensional LC-tandem mass spectrometry (LC/LC-MS/MS). A negative control was performed using an isogenic yeast strain expressing wild-type Ub. A total of 1,075 proteins were identified in the IMAC enriched sample after discarding 48 proteins that were present in the negative control. However, only 72 of these proteins were reported as being ubiquitinated. Among the proteins identified, one third were internal membrane proteins including known substrates of various E3 ligases.

Also in 2003, Hitchcock *et al.* employed a similar strategy as that described by Peng *et al.* to purify and identify 6xHis-Ub conjugates in endoplasmic reticulum (ER) membrane containing fractions from yeast cells.[74] A total of 211 ER membrane-associated proteins were identified by LC/LC-MS/MS, 29 of which were reported as being ubiquitinated. Among the 211 membrane-associated proteins, 83 were classified as potential substrates of the ER degradation pathway (ERAD). For this set of experiments, yeast strains expressing 6xHis-Ub with and without mutations in the ER resident E2 conjugating enzyme UBC7 and the ERAD-associated protein NPL4 were employed.

In early 2005, the expression of 6xHis-Ub and the purification of tagged Ub-conjugates was extended to mammalian cells by Kirkpatrick *et al.*, but with much less success compared to previous studies in yeast.[75] A total of 22 proteins from human embryonic kidney (HEK293) cells were identified by LC-MS/MS, only one of which was reported to be ubiquitinated. For this set of experiments, 6xHis-Ub and the green fluorescent protein (GFP) were co-expressed by stable transfection of HEK293 cells. The poor yield of ubiquitinated proteins in this study is thought to be the result of the low expression level of poly-histidine tagged Ub compared to that of

endogeneously expressed Ub from multiple Ub genes.[69] *In vivo* expression of polyhistidine-tagged Ub in a transgenic mouse model has also been described. Tsirigotis *et al.* successfully employed the human UbC promoter to drive the expression of 6xHis-tagged human Ub in a variety of mouse tissues as assessed by a GFP marker.[76] Expression of 6xHis-Ub was reported as early as the morula stage of embryogenesis. Following a purification step employing nickel matrix Fast Protein Liquid Chromatography (FPLC), Ub-conjugates were detected by Western blot analysis. The identities of the proteins, however, were not determined.

## 1.6 Gel-based and gel-free proteomics

The term "proteome" was coined in 1994 by Wilkins *et al.* and was defined as "the study of the full set of proteins encoded by a genome".[77,78] The traditional method for analyzing the proteome of a cell, tissue or organ involves the separation of proteins by one-dimensional (1-D) or two-dimensional (2-D) gel electrophoresis and the identification of the separated proteins by mass spectrometry analysis. 2-D gel electrophoresis is a technique that was described in 1975 by O'Farrell and since then, numerous studies have described improvements to 2-D gel based methods.[79] However, the underlying principle has remained the same; proteins are separated according to their isoelectric point (pI) in the first dimension and according to their molecular weight in the second dimension. Since these physicochemical properties are unrelated, it is possible to obtain a distribution of protein spots across a 2-D gel. The resulting map of protein spots is considered as the protein fingerprint of the sample.

Since polyacrylamide has the same ultraviolet absorbance as proteins, they must be stained to visualize them on a gel. Coomassie blue staining is the most popular staining method as it is relatively inexpensive, easy to use and has a wide linear range, making relative quantitation possible. However, coomassie blue is somewhat insensitive and a large number of proteins remain undetected using this method. In contrast, silver staining has a higher detection limit but has the disadvantage of not being very compatible with mass spectrometry analysis.[80] The use of formaldehyde or glutaraldehyde during the silver staining development step covalently modifies lysine and arginine residues of the proteins, which has the consequence of impairing the efficiency of in-gel tryptic digestion and the ability of database search engines to recognize these modified residues during MS/MS data analysis. Other mass spectrometry-compatible stains exist such as Sypro ruby.[81] However, this and other fluorescent staining methods require the use of costly and complex scanning devices for detection.

Proteins present in a sample in extremely low concentrations or proteins that cannot be separated on a gel due to their physicochemical properties (pI, hydrophobicity, molecular weight) will not be detected regardless of the staining technique employed. In practice this means that low-abundance proteins (e.g. ubiquitinated proteins), membrane proteins and small proteins/peptides present a significant challenge when it comes to their detection and analysis. Instead of using 1-D or 2-D gel electrophoresis for separating proteins, gel-free approaches for separating and digesting proteins may be employed. For example, a complex protein sample can be enzymatically digested and the peptides separated by multi-dimensional capillary liquid chromatography. Washburn *et al.* developed a method employing multi-dimensional chromatography and MS/MS analysis known as MudPIT, which is based on the sequential packing of strong cation exchange (SCX) beads and reversed phased beads on a biphasic

column.[82] In this study, MudPIT was applied for the analysis of the yeast proteome. A total of 1,484 proteins were identified from *Saccharomyces cerevisaie*, which included low-abundance proteins, such as transcription factors and protein kinases, which were rarely seen in prior studies employing standard gel-based methods.

Another gel-free approach for processing protein samples for mass spectrometry analysis was recently described by Ethier *et al.* and is based on the use of a microfluidic device known as the Proteomic Reactor.[83] This device allows for pre-concentration, derivatization and proteolytic digestion of proteins into peptides prior to MS/MS analysis. In this study, the Proteomic Reactor was configured to bind up to 10-15 μg of protein. At the lower limit, proteins were detected from a P19 cell lysate containing 0.5 μg of total protein. Online cell lysis was also performed and enabled the identification of proteins from as little as 300 cells. The Proteomic Reactor itself consists of a fused silica capillary column loaded with SCX resin, and because of the low pH environment within the reactor, proteins bind to the resin allowing for reduction and alkylation and subsequent tryptic digestion when the pH is neutralized. Comparison to other gel-free approaches revealed that the Proteomic Reactor was at least ten times more sensitive, with one protein identified per 440 pg of protein lysate loaded.

## 1.7 Protein identification by mass spectrometry

Mass spectrometry is a technique that measures the mass to charge ratios (*m/z*) of gas-phase ions. In the last 20 years, it has played an increasingly significant role in the life sciences and today it is the most sensitive method for the identification and structural characterization of biomolecules, including proteins. The successes of mass spectrometry in the life sciences are

mainly the result of the introduction of the "soft" ionization techniques known as electrospray ionization (ESI) and matrix-assisted laser desorption ionization (MALDI), which allow for the transfer of polar, thermally labile biomolecules into the gas phase for subsequent mass analysis.[84,85] In 2002, Fenn and Tanaka were awarded the Nobel Prize in Chemistry for their pioneering work on ESI and MALDI.

Modern-day mass spectrometers consist of three essential components: 1) an ion source, to produce ions from the sample; 2) one or more mass analyzers, to separate the ions according to their *m/z* ratios; and 3) a detector, to register the number of ions at each *m/z* value.[84] As mentioned above, the ionization techniques used for mass spectrometric analysis proteins and peptides are ESI and MALDI. ESI produces gaseous ions from solution phase samples at atmospheric pressure, is easily coupled to liquid-based separation methods such as liquid chromatography and is generally used to analyze complex peptide mixtures. In contrast, MALDI produces gaseous ions from a dry, crystalline matrix under vacuum conditions and is generally used to analyze simple peptide mixtures.[86]

To generate gas-phase ions by ESI, a solution phase sample is passed at low flow rates (μL/min) through a capillary to which a high voltage is applied. The solution experiences an electric field between the capillary and a counter electrode and, assuming a positive potential is applied to the capillary, positive ions in solution accumulate at the tip of the capillary and the liquid assume a conical shape known as a Taylor cone.[87] As the liquid is forced to hold more electric charge, the cone is drawn out into a filament that produces positively charged droplets via a budding process when the applied electrostatic force exceeds the surface tension. The droplets then travel towards the counter electrode and pass through a curtain of nitrogen gas which causes solvent to evaporate. As the size of the droplets decreases, the electrical charge

density at the surface of the droplets increases. When the electrostatic force exceeds the surface tension (Rayleigh limit), a Coulomb explosion occurs, emitting smaller droplets.[87] Continual depletion of droplet size eventually results in the formation of droplets containing a single ion. This is known as the Charge Residue model. An alternative mechanism for gas-phase ion production has been proposed, whereby ions are produced from highly charged droplets, with the driving force for ion formation being the repulsion between the charged ion and the other charges of the droplet.[87] This is known as the Ion Evaporation model. Although there is debate over which is the correct model, ions of the form $[M + H]^+$ or $[M + nH]^{n+}$ are produced. Nanoelectrospray ionisation (nanoESI) is simply a low flow rate (nL/min) version of ESI.

The mass analyzer is the component of the mass spectrometer that separates the ions produced by ESI. The parameters that affect the performance of the mass analyzer include its resolution, mass accuracy, dynamic range and sensitivity.[86] Currently there are five types of mass analyzers used for proteomics experiments: the quadrupole, the ion trap, the time-of-flight (TOF), the ion cyclotron resonance analyzer (ICR) and the orbitrap. They are unique in terms of their design and performance and can be used as stand-alone analyzers or in some cases, can be combined in tandem to take advantage of their different strengths.[84-86] (Only the quadrupole is discussed below as an overview of the other types of analyzers is beyond the scope of this introduction.)

Quadrupole mass analyzers consist of four parallel rods equally spaced around a central axis.[86] Opposing sets of rods have both a direct current (DC) and a radio frequency (RF) voltage component that generate electric fields, one set positive and the other set negative. The positive rods create a high-pass mass filter and the negative rods create a low pass mass filter. An area of mutual stability is created where the two mass filter regions overlap, allowing ions of a certain

*m/z* to be transmitted through the quadrupole.[86] Ions with *m/z* ratios outside this area of mutual stability cannot be transmitted and collide with the rods. The *m/z* ratio of the ions that are transmitted is proportional to the voltage that is applied the rods; the higher the voltage, the higher the *m/z* value that is transmitted. By varying the relative contributions of the DC and RF voltages, the width of the transmission area can be adjusted. The wider the area, the wider peak, resulting in lower resolution. The narrower the area, the narrower the peak, resulting in higher resolution.[86] Scanning a quadrupole involves ramping the DC and RF voltages in a constant ratio, thus changing the position of the transmission region and allowing ions of different *m/z* to be transmitted. For MS/MS analysis, three quadrupoles can be combined in succession to form a triple quad. The first and third quadrupoles are used for scanning, while the middle one is used as a collision cell to fragment the ions.[85,86] Fragmentation occurs by collision-induced dissociation (CID) via low-energy collisions with inert gas molecules, such as nitrogen or helium.

Peptide ion fragmentation is promoted by a "mobile" proton and the relative populations of the different protonated forms of a peptide depend on the internal energy content and its primary structure.[86] Collisions with gas molecules increases the energy of the protonated peptide ions. For these excited peptide ions, the proton is located at various backbone heteroatoms. Three different types of bonds can fragment along the amino acid backbone: the NH-CH; CH-CO; and CO-NH bonds.[86,88] Each bond cleavage gives rise to two species, one neutral and the other one charged, and only the charged species is detected.[88] The charge can stay on either of the two fragments depending on the chemistry and relative proton affinity of the two species. There are six possible fragment ions for each amino acid and these are denoted as *a*, *b* and *c* ions having the charge retained on the N-terminal fragment, and the *x*, *y* and *z* ions having the charge retained on the C-terminal fragment.[88] The most common cleavage site is at the CO-NH bond (peptide

bond), which gives rise to the *b* and/or the *y* ions. The mass difference between two adjacent *b* ions or *y* ions is indicative of a particular amino acid residue.[86,88]

A typical proteomics experiment using a triple quad would therefore proceed as follows. The protein sample is digested with a suitable enzyme, such as trypsin or chymotrypsin. Trypsin is useful because each proteolytic fragment contains a basic arginine or lysine residue at its C-terminus, making it suitable for positive ionization. The digest mixture is then subjected to ESI and the resulting peptide ions are transmitted through the first quadrupole. This first quadrupole is not used as an analyzer, but merely as a lens to focus the ions into the third quadrupole which functions as an analyzer and separates the ions according to their *m/z* for detection. The mass spectrometer in this case, operates in the "MS mode" and produces a rather complex MS spectrum, called a peptide map, which contains the *m/z* ratios of all of the tryptic peptides.

While the digest mixture is spraying out of the ESI needle and ions are being produced, the mass spectrometer is switched into "MS/MS" mode. The peptide ions are now independently selected and transmitted through the first quadrupole, which is now used as an analyzer to transmit the ions of interest into the collision cell. An inert gas is introduced into the collision cell and the peptide ions collide with the gas molecules, which increases their energy and causes them to fragment via CID. The fragment ions (or product ions) are then transmitted through the third quadrupole and are detected. In this way MS/MS spectra are produced which display all the fragment ions that arise directly from the chosen precursor peptide ions.

Each MS/MS spectrum needs to be interpreted either manually or using a database search engine such as Mascot or Sequest. These search engines function by comparing the MS/MS fragmentation patterns of the unknown peptides to the predicted fragmentation patterns of all

peptides of the same mass in a protein database and reporting the best matches. Although MS/MS data processing is typically automated, spectra interpretation almost always takes a longer amount of time than the data acquisition step if accurate and reliable data is generated. The amount of sequence information generated will vary from one peptide to another. Some peptide sequences will be confirmed completely while a partial sequence of, say 4-6 amino acid residues, may be obtained for others. The outcome of database searching of the MS/MS spectra is the identification of the proteins that were present in the original sample via single or multiple peptide matches.

## 1.8 Identification of Ub proteins and ubiquitination sites by mass spectrometry

The specific lysine residues on a protein substrate that are modified by Ub can be mapped by MS/MS analysis after tryptic digestion because trypsin cleaves the carboxy terminus of Ub after arginine-74 (R74), leaving a di-glycine (GG) remnant with a monoisotopic mass of 114.04292 Da on lysine residues of the substrate.[51] In addition, tryptic peptides harboring a GG remnant contain a missed cleavage site at the modified lysine residue(s). Based on these characteristics, sites of Ub modification can be identified by the unique MS/MS spectra of GG-containing peptides after database searching using search engines such as Mascot and Sequest.[89] In addition to the GG remnant, it has been speculated that a miscleavage at R74 of Ub may occur during tryptic digestion, leaving a longer LRGG remnant with a monoisotopic mass of 383.22809 Da, which can also be identified by MS/MS analysis and database searching.[90] The primary structure of Ub and the production of a signature peptide with a GG or LRGG remnant after tryptic digestion is shown in Figure 1.1.

During database searching of MS/MS data, the number of Ub modification sites may be overestimated by incorrect assignments. For example, peptides can be incorrectly assigned as being ubiquitinated if they contain an internal lysine and an adjacent amino acid residue that has a similar mass as the GG remnant. The possible residues include asparagine (114.02493 Da), aspartate (115.02694 Da), leucine (113.08406 Da) and isoleucine (113.08406 Da), provided that the mass of precursor and product ions is still within selected mass tolerance range during database searching.[89] This type of mismatch can be caught by determining the tryptic state of the peptides, as the mismatched peptides have only one tryptic end. The use of high mass accuracy mass spectrometers is also effective in minimizing errors caused by flanking amino acid residues with similar masses.

Even when high mass accuracy mass spectrometers are employed for MS/MS analysis, incorrect assignment of tryptic peptides occur as a result of random matches to the protein database that is being searched. This is because database search engines employ algorithms that compare the observed MS/MS fragmentation pattern of an unknown peptide with the predicted MS/MS fragmentation patterns of all peptides of equivalent mass within the database and return a peptide sequence whose predicted fragmentation pattern best matches the observed MS/MS fragmentation pattern.[91] The consequence of this approach is the possibility of false positive identifications.

In order to minimize the number of false positive identifications, search engines employ a set of scoring criteria so that only the most likely candidates to be correct matches are retained. Mascot and the open-source search engines X!Tandem and OMSSA use a set of scoring criteria that that are based on idealized random models. These models are typically trained on datasets consisting of peptide MS/MS spectra obtained from simple mixtures of known standard

proteins.[92] However, the specific organisms under study, growth conditions, experimental protocols, types of instrumentation and sample complexities of any particular experiment are unlikely to match the conditions under which these models are trained or the assumptions that are made by idealized peptide fragmentation models.[80] Therefore, the probabilistic estimates from these models are not always reliable and improved methods for estimating the error rates of peptide identifications are required. This is especially true for peptides harboring post-translational modifications, such as phosphorylation, glycosylation or ubiquitination.

One method for estimating the error rate or false positive rate (FPR) for peptide identifications involves the use of a randomized sequence database for database searching of experimental MS/MS data. This approach estimates the number of false positive identifications by counting the number of peptide matches to the randomized sequence database using the same set of search criteria used for performing a search with an actual protein database (i.e. peptide ion score cutoff, variable and fixed modifications, etc.). A separate randomized database can be used, where the real database is searched first, then the randomized database is searched separately. Alternatively, the real and randomized databases can be combined or appended and searched simultaneously. In 2005, Higdon *et al.* compared both strategies using MS/MS data from tryptic digests of known protein standards and from the model microorganism *Shewanella oneidensis* strain MR-1 and found that the latter provided more accurate estimates of FPR.[92]

## 1.9 Research hypotheses and objectives

Hypothesis #1: Endogenously-expressed ubiquitinated proteins can be purified using an antibody-based approach and identified by mass spectrometry.

Objectives:

1) Determine if a commercially-available anti-Ub mouse monoclonal antibody can be used to purify endogenously expressed ubiquitinated proteins.

2) Develop an antibody-based purification protocol for ubiquitinated proteins expressed in mammalian cells.

3) Profile ubiquitinated proteins in MCF-7 human breast cancer cells treated with the proteasome inhibitor MG132.

Hypothesis #2: A microfluidic processing device known as the Proteomic Reactor can be used to facilitate the analysis of low-abundance affinity-purified proteins by mass spectrometry.

Objectives:

1) Determine if the Proteomic Reactor can be used to process affinity-purified ubiquitinated proteins for mass spectrometry analysis.

2) Develop a protocol that couples affinity-purification with tryptic digestion of proteins using the Proteomic Reactor.

3) Demonstrate the applicability of the protocol using affinity-purified ubiquitinated proteins from human H1299 cells treated with siRNA targeting the ubiquitin chaperone VCP.

Hypothesis #3: False positive identification of ubiquitinated proteins by mass spectrometry can be minimized by systematically evaluating a range of peptide ion score cutoffs.

Objectives:

1) Determine if standard ion score cutoffs for identifying tryptic peptides are sufficiently stringent for identifying ubiquitinated peptides.

2) Determine if false positive rates for ubiquitinated proteins are affected by the size of the protein database that is searched.

3) Develop a systematic approach for determining appropriate ion score cutoffs for ubiquitinated proteins based on false positive rate calculations.

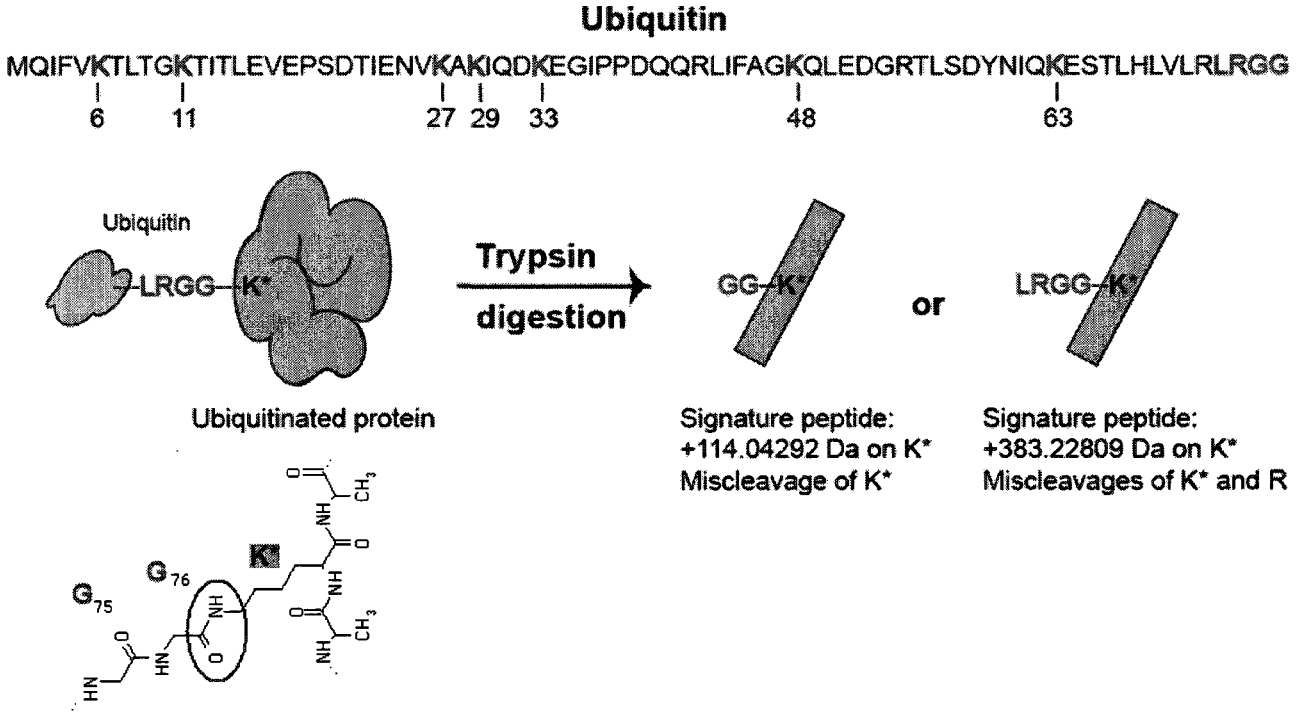
Hypothesis #4: The improved methods developed for analyzing ubiquitinated proteins (see Hypotheses 1-3) can be used to identify specific targets of ubiquitination in eukaryotic cells.

Objectives:

1) Identify the specific lysine residues on the membrane-anchored transcription factor Mga2p that are targets of the E3 ubiquitin ligase Rsp5p in *Saccharomyces cerevisiae.*

2) Identify the proteins that are ubiquitinated upon mitotic exit in the embryos of *Xenopus laevis.*

**Figure 1.1 Primary structure of Ubiquitin and production of signature peptides.** Ub is 76-amino acid protein that is covalently attached via its C-terminal glycine ($G^{76}$) to a lysine (K) on a target protein via an isopeptide bond. When ubiquitinated proteins are subjected to trypsin digestion, signature peptides with a GG or LRGG remnant are produced. Signature peptides have monoisotopic mass increases of +114.04292 or +383.22809 Da and missed cleavage sites that can be identified by tandem mass spectrometry (MS/MS) analysis.



## 1.10 References

1) de Duve C, Gianetto R, Appelmans F and Wattiaux R (1953) Enzymic content of the mitochondria fraction. Nature 172: 1143–1144.

2) Gianetto R and de Duve C (1955) Tissue fractionatin studies 4. Comparative study of the binding of acid phosphatase, b-glucoronidase and cathepsin by rat liver particles. Biochem. J. 59: 433–438.

3) Simpson MV (1953) The release of labeled amino acids from proteins in liver slices. J. Biol. Chem. 201: 143–154.

4) Ciechanover A (2005). Intracellular protein degradation: from a vague idea thru the lysosome and the ubiquitin–proteasome system and onto human diseases and drug targeting. Cell Death Diff. 12, 1178–1190.

5) Etlinger JD and Goldberg AL (1977) A soluble ATP-dependent proteolytic system responsible for the degradation of abnormal proteins in reticulocytes. Proc. Natl. Acad. Sci. USA 74: 54–58.

6) Hershko A, Heller H, Ganoth D and Ciechanover A (1978) Mode of degradation of abnormal globin chains in rabbit reticulocytes In Protein Turnover and Lysosome Function Segal HL, Doyle DJ, eds (New York: Academic Press) pp. 149–169.

7) Ciechanover A, Hod Y and Hershko A (1978) A heat-stable polypeptide component of an ATP-dependent proteolytic system from reticulocytes. Biochem. Biophys. Res. Commun. 81: 1100–1105.

8) Ciechanover A, Heller H, Elias S, Haas AL and Hershko A (1980) ATP-dependent conjugation of reticulocyte proteins with the polypeptide required for protein degradation. Proc. Natl. Acad. Sci. USA 77: 1365–1368.

9) Hershko A, Ciechanover A, Heller H, Haas AL and Rose IA (1980) Proposed role of ATP in protein breakdown: conjugation of proteins with multiple chains of the polypeptide of ATP-dependent proteolysis. Proc. Natl. Acad. Sci. USA 77: 1783–1786.

10) Wilkinson KD, Urban MK and Haas AL (1980) Ubiquitin is the ATP-dependent proteolysis factor I of rabbit reticulocytes. J. Biol. Chem. 255: 7529–7532.

11) Goldstein G (1974) Isolation of bovine thymin, a polypeptide hormone of the thymus. Nature 247: 11–14.

12) Goldstein G, Scheid M, Hammerling U, Schlesinger DH, Niall HD and Boyse EA (1975) Isolation of a polypeptide that has lymphocyte-differentiating properties and is probably represented universally in living cells. Proc. Natl. Acad. Sci. USA 72: 11–15.

13) Goldknopf IL and Busch H (1977) Isopeptide linkage between non-histone and histone 2A polypeptides of chromosome conjugate-protein A24. Proc. Natl. Acad. Sci. USA 74: 864–868.

14) Hershko A, Ciechanover A and Rose IA (1981) Identification of the active amino acid residue of the polypeptide of ATP-dependent protein breakdown. J. Biol. Chem. 256: 1525–1528.

15) Hershko A and Heller H (1985) Occurrence of a polyubiquitin structure in ubiquitin-protein conjugates. Biochem. Biophys. Res. Commun. 128: 1079–1086.

16) Ciechanover A, Heiler H, Katz-Etzion R and Hershko A (1981) Activation of the heat-stable polypeptide of the ATP-dependent proteolytic system. Proc. Natl Acad. Sci USA 78: 761–765.

17) Haas AL, Warms JV, Hershko A and Rose IA (1982) Ubiquitin-activating enzyme. Mechanism and role in protein-ubiquitin conjugation. J. Biol. Chem. 257: 2543–2548.

18) Hershko A, Heller H, Elias S and Ciechanover A (1983) Components of the ubiquitin-protein ligase system: resolution, affinity purification and role in protein breakdown. J. Biol. Chem. 258: 8206–8214.

19) Hershko A, Heiler H, Eytan E and Reiss Y (1986) The protein substrate binding site of the ubiquitin-protein ligase system. J. Biol. Chem. 261: 11992–11999.

20) Hershko A (1988) Ubiquitin-mediated protein degradation. J. Biol. Chem. 263: 15237–15240.

21) Hershko A. The ubiquitin system for protein degradation and some of its roles in the control of the cell division cycle. Cell Death and Differentiation (2005) 12, 1191–1197.

22) Hough R, Pratt G and Rechsteiner M (1986) Ubiquitin–lysozyme conjugates. Identification and characterization of an ATP-dependent protease from rabbit reticulocyte lysates. J. Biol. Chem. 261: 2400–2408.

23) Waxman L, Fagan J and Goldberg AL (1987) Demonstration of two distinct high molecular weight proteases in rabbit reticulocytes, one of which degrades ubiquitin conjugates. J. Biol. Chem. 262: 2451–2457.

24) Hough R, Pratt G and Rechsteiner M (1987) Purification of two high molecular weight proteases from rabbit reticulocyte lysate. J. Biol. Chem. 262: 8303–8313

25) Wilk S and Orlowski M (1980) Cation-sensitive neutral endopeptidase: isolation and specificity of the bovine pituitary enzyme. J. Neurochem. 35: 1172–1182.

26) Eytan E, Ganoth D, Armon T and Hershko A (1989) ATP-dependent incorporation of 20S protease into the 26S complex that degrades proteins conjugated to ubiquitin. Proc. Natl. Acad. Sci. USA 86: 7751–7755

27) Driscoll J and Goldberg AL (1990) The proteasome (multicatalytic protease) is a component of the 1500-kDa proteolytic complex which degrades ubiquitinconjugated proteins. J. Biol. Chem. 265: 4789–4792

28) Hoffman L, Pratt G and Rechsteiner M (1992) Multiple forms of the 20S multicatalytic and the 26S ubiquitin/ATP-dependent proteases from rabbit reticulocyte lysate. J. Biol. Chem. 267: 22362–22368.

29) Baumeister W, Walz J, Zühl F, Seemüller E (1998) The proteasome: paradigm of a self-compartmentalizing protease. Cell. 92(3):367-80.

30) DeMartino G.N., Slaughter C.A. (1999) The proteasome, a novel protease regulated by multiple mechanisms. J. Biol. Chem. 274(32): 22123-22126.

31) Peters JM (2002) The anaphase-promoting complex: proteolysis in mitosis and beyond. Mol Cell. 9(5): 931-43.

32) Maki CG, Huibregtse JM, Howley PM (1996) In vivo ubiquitination and proteasome-mediated degradation of p53. Cancer Res. 56(11): 2649-54.

33) Palombella VJ, Rando OJ, Goldberg AL, Maniatis T (1994) The ubiquitin-proteasome pathway is required for processing the NF-kappa B1 precursor protein and the activation of NF-kappa B. Cell. 78(5): 773-85.

34) Choi YH (2001) Proteasome-mediated degradation of BRCA1 protein in MCF-7 human breast cancer cells. Int. J. Oncol. 19(4): 687-93.

35) Semenza G (2002). Signal transduction to hypoxia-inducible factor 1 Biochem Pharmacol. 64(5-6): 993-8.

36) Mani A, Gelmann EP (2005). The ubiquitin-proteasome pathway and its role in cancer. J Clin Oncol. 23(21): 4776-89.

37) Yang Y, Kitagaki J, Wang H, Hou DX, Perantoni AO (2009). Targeting the ubiquitin-proteasome system for cancer therapy. Cancer Sci. 100(1): 24-8.

38) Petroski MD (2008) The ubiquitin system, disease, and drug discovery. BMC Biochem. Oct 21;9 Suppl 1:S7

39) Lim KL, Tan JM (2007). Role of the ubiquitin proteasome system in Parkinson's disease. BMC Biochem. 22;8 Suppl 1:S13.

40) Gaczynska M, Osmulski PA (2002). Inhibitor at the gates, inhibitor in the chamber: allosteric and competitive inhibitors of the proteasome as prospective drugs. Curr. Med. Chem. 2: 279-301.

41) Lee DH, Goldberg AL (1998). Proteasome inhibitors: valuable new tools for cell biologists. Trends Cell Biol. 8(10): 397-403.

42) Myung J, Kim KB, Crews CM (2001). The ubiquitin-proteasome pathway and proteasome inhibitors. Med. Res. Rev. 21: 245-273.

43) Kisselev AF, Goldberg AL (2001). Proteasome inhibitors: from research tools to drug candidates. Chem Biol. 8(8):739-58.

44) Pickart CM, Fushman D (2004). Polyubiquitin chains: polymeric protein signals. Curr Opin Chem Biol. 8(6): 610-6.

45) Kannouche PL, Wing J, Lehmann AR (2004). Interaction of human DNA polymerase eta with monoubiquitinated PCNA: a possible mechanism for the polymerase switch in response to DNA damage. Mol Cell. 14(4): 491-500.

46) Ulrich HD (2002). Degradation or maintenance: actions of the ubiquitin system on eukaryotic chromatin. Eukaryot Cell. 1(1):1-10.

47) Kanayama A, Seth RB, Sun L, Ea CK, Hong M, Shaito A, Chiu YH, Deng L, Chen ZJ (2004). TAB2 and TAB3 activate the NF-kappaB pathway through binding to polyubiquitin chains. Mol Cell. 15(4):535-48.

48) Chastagner P, Israël A, Brou C (2006). Itch/AIP4 mediates Deltex degradation through the formation of K29-linked polyubiquitin chains. EMBO Rep. 7(11):1147-53.

49) Al-Hakim AK, Zagorska A, Chapman L, Deak M, Peggie M, Alessi DR (2008). Control of AMPK-related kinases by USP9X and atypical Lys(29)/Lys(33)-linked polyubiquitin chains. Biochem J. 411(2): 249-60.

50) Nishikawa H, Ooka S, Sato K, Arima K, Okamoto J, Klevit RE, Fukuda M, Ohta T (2004). Mass spectrometric and mutational analyses reveal Lys-6-linked polyubiquitin chains catalyzed by BRCA1-BARD1 ubiquitin ligase. J Biol Chem. 279(6): 3916-24.

51) Peng J, Schwartz D, Elias JE, Thoreen CC, Cheng D, Marsischky G, Roelofs J, Finley D, Gygi SP (2003). A proteomics approach to understanding protein ubiquitination. Nat. Biotechnol. 21(8):921-6.

52) Baboshina OV, Haas AL (1996). Novel multiubiquitin chain linkages catalyzed by the conjugating enzymes E2EPF and RAD6 are recognized by 26S proteasome subunit 5. J. Biol. Chem. 271(5): 2823-31.

53) Li W, Ye Y (2008). Polyubiquitin chains: functions, structures, and mechanisms. Cell Mol Life Sci. 65(15): 2397-406.

54) Scheffner M, Huibregtse JM, Vierstra RD, Howley PM (1993) The HPV-16 E6 and E6-AP complex functions as a ubiquitin-protein ligase in the ubiquitination of p53. Cell. 75(3): 495-505.

55) Fang S, Weissman AM (2004). A field guide to ubiquitylation. Cell Mol Life Sci. 61(13):1546-61.

56) Sudol M, Chen HI, Bougeret C, Einbond A, Bork P (1995) Characterization of a novel protein-binding module--the WW domain. FEBS Lett. 369(1):67-71.

57) Kay BK, Williamson MP, Sudol M (2000) The importance of being proline: the interaction of proline-rich motifs in signaling proteins with their cognate domains. FASEB J. 14(2):231-41.

58) Schild L, Lu Y, Gautschi I, Schneeberger E, Lifton RP, Rossier BC (1996). Identification of a PY motif in the epithelial Na channel subunits as a target sequence for mutations causing channel activation found in Liddle syndrome. EMBO J. 15(10):2381-7.

59) Honda R, Yasuda H (2000). Activity of MDM2, a ubiquitin ligase, toward p53 or itself is dependent on the RING finger domain of the ligase. Oncogene. 19(11):1473-6.

60) Joazeiro CA, Wing SS, Huang H, Leverson JD, Hunter T, Liu YC (1999). The tyrosine kinase negative regulator c-Cbl as a RING-type, E2-dependent ubiquitin-protein ligase. Science. 286(5438):309-12.

61) Lyapina SA, Correll CC, Kipreos ET, Deshaies RJ (1998). Human CUL1 forms an evolutionarily conserved ubiquitin ligase complex (SCF) with SKP1 and an F-box protein. Proc Natl Acad Sci U S A. 95(13):7451-6.

62) Hoppe T (2005). Multiubiquitylation by E4 enzymes: 'one size' doesn't fit all. Trends Biochem Sci. 30(4):183-7.

63) Koegl M, Hoppe T, Schlenker S, Ulrich HD, Mayer TU, Jentsch S (1999). A novel ubiquitination factor, E4, is involved in multiubiquitin chain assembly. Cell. 96(5): 635-44.

64) Richly H, Rape M, Braun S, Rumpf S, Hoege C, Jentsch S (2005). A series of ubiquitin binding factors connects CDC48/p97 to substrate multiubiquitylation and proteasomal targeting. Cell. 120(1):73-84.

65) Pukatzki S, Tordilla N, Franke J, Kessin RH (1998). A novel component involved in ubiquitination is required for development of Dictyostelium discoideum. J Biol Chem. 273(37): 24131-8.

66) Matsumoto M, Yada M, Hatakeyama S, Ishimoto H, Tanimura T, Tsuji S, Kakizuka A, Kitagawa M, Nakayama KI (2004). Molecular clearance of ataxin-3 is regulated by a mammalian E4. EMBO J. 23(3):659-69.

67) Millard SM, Wood SA. Riding the DUBway: regulation of protein trafficking by deubiquitylating enzymes (2006). J Cell Biol. 173(4):463-8.

68) Vasilescu J, Figeys D (2006). Mapping protein-protein interactions by mass spectrometry. Curr Opin Biotechnol. 17(4):394-9.

69) Xu P, Peng J (2006). Dissecting the ubiquitin pathway by mass spectrometry. Biochim Biophys Acta. 1764(12):1940-7.

70) Tomlinson E, Palaniyappan N, Tooth D, Layfield R (2007). Methods for the purification of ubiquitinated proteins. Proteomics. (7):1016-22.

71) Gaberc-Porekar V, Menart V (2001). Perspectives of immobilized-metal affinity chromatography. J Biochem Biophys Methods. 49(1-3):335-60.

72) Peng J, Cheng D (2005). Proteomic analysis of ubiquitin conjugates in yeast. Methods Enzymol. 399:367-81.

73) Gururaja T, Li W, Noble WS, Payan DG, Anderson DC (2003). J Proteome Res. 2(4):394-404.

74) Hitchcock AL, Auld K, Gygi SP, Silver PA (2003). A subset of membrane-associated proteins is ubiquitinated in response to mutations in the endoplasmic reticulum degradation machinery. Proc Natl Acad Sci U S A. 100(22):12735-40.

75) Kirkpatrick DS, Weldon SF, Tsaprailis G, Liebler DC, Gandolfi AJ (2005). Proteomic identification of ubiquitinated proteins from human cells expressing His-tagged ubiquitin. Proteomics. 5(8):2104-11.

76) Tsirigotis M, Thurig S, Dubé M, Vanderhyden BC, Zhang M, Gray DA (2001). Analysis of ubiquitination in vivo using a transgenic mouse model. Biotechniques. 31(1):120-6.

77) Wilkins MR (1994). 2D Electrophoresis: from protein maps to genomes. Symposium Proceedings. University of Siena, Italy.

78) Wasinger VC, Cordwell SJ, Cerpa-Poljak A, Yan JX, Gooley AA, Wilkins MR, Duncan MW, Harris R, Williams KL, Humphery-Smith I (1995). Progress with gene-product mapping of the Mollicutes: Mycoplasma genitalium. Electrophoresis. 16(7):1090-4.

79) O'Farrell PH. High resolution two-dimensional electrophoresis of proteins (1975). J Biol Chem. 250(10):4007-21.

80) Shevchenko A, Wilm M, Vorm O, Mann M. Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels (1996). Anal Chem. 68(5):850-8.

81) Lopez MF, Berggren K, Chernokalskaya E, Lazarev A, Robinson M, Patton WF (2000). A comparison of silver stain and SYPRO Ruby Protein Gel Stain with respect to protein detection

in two-dimensional gels and identification by peptide mass profiling. Electrophoresis. 21(17):3673-83.

82) Washburn MP, Wolters D, Yates JR 3rd. Large-scale analysis of the yeast proteome by multidimensional protein identification technology (2001). Nat Biotechnol. 19(3):242-7.

83) Ethier M, Hou W, Duewel HS, Figeys D (2006). The proteomic reactor: a microfluidic device for processing minute amounts of protein prior to mass spectrometry analysis. J Proteome Res. 5(10):2754-9.

84) Aebersold R, Mann M. Mass spectrometry-based proteomics (2003). Nature. 422(6928):198-207.

85) Domon B, Aebersold R. Mass spectrometry and protein analysis (2006). Science. 14;312(5771):212-7.

86) Lane CS. Mass spectrometry-based proteomics in the life sciences (2005). Cell Mol Life Sci. 62(7-8):848-69.

87) Gaskell SJ. Electrospray: principles and practice (1997). J Mass Spectrom. 32(7):677-688.

88) http://www.matrixscience.com/help/fragmentation_help.html

89) Peng J. Evaluation of proteomic strategies for analyzing ubiquitinated proteins (2008). BMB Reports. 41(3): 177-183.

90) Warren MR, Parker CE, Mocanu V, Klapper D, Borchers CH (2005). Electrospray ionization tandem mass spectrometry of model peptides reveals diagnostic fragment ions for protein ubiquitination. Rapid Commun Mass Spectrom. 19(4):429-37.

91) Huttlin EL, Hegeman AD, Harms AC, Sussman MR (2007). Prediction of error associated with false-positive rate determination for peptide identification in large-scale proteomics experiments using a combined reverse and forward peptide sequence database strategy. J Proteome Res. 6(1):392-8.

92) Higdon R, Hogan JM, Van Belle G, Kolker E (2005). Randomized sequence databases for tandem mass spectrometry peptide and protein identification. OMICS. 9(4):364-79.

# Chapter 2: Proteomic analysis of ubiquitinated proteins from human MCF-7 breast cancer cells by immunoaffinity purification and mass spectrometry

## 2.1 Manuscript status and statement of author contributions

A version of this chapter has been published in the Journal of Proteome Research: *Vasilescu J, Smith JC, Ethier M, Figeys D. Proteomic analysis of ubiquitinated proteins by immunoaffinity purification and mass spectrometry. Journal of Proteome Research. Vol. 4: 2192-2200, 2005.* JV conceived, designed and performed the experiments. JV also wrote the manuscript. JCS and ME assisted with the data analysis. DF edited the manuscript.

## 2.2 Summary

Post-translational modification of proteins via the covalent attachment of Ubiquitin (Ub) plays an important role in the regulation of protein stability and function in eukaryotic cells. In the present study, we describe a novel method for identifying ubiquitinated proteins from a complex biological sample, such as a whole cell lysate, using a combination of immunoaffinity purification and liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis. We have demonstrated the applicability of this approach by identifying 70 ubiquitinated proteins from the human MCF-7 breast cancer cell line after treatment with the proteasome inhibitor MG132. This method will aid the study of protein ubiquitination and may be used as a tool for the discovery of novel biomarkers that are associated with disease progression.

## 2.3 Introduction

Ubiquitin (Ub) is a small 8.5 kDa protein present in all eukaryotes that participates in a host of critical cellular functions by mediating the selective degradation of regulatory proteins. The progression of the cell cycle,[1] the heat shock response,[2] and antigen presentation,[3] are among the many processes regulated by Ub-dependent proteolysis. Not surprisingly, aberrations in this pathway are implicated in the pathogenesis of many diseases including cancer.[4] Ubiquitination also regulates certain processes by mechanisms that, although poorly understood, do not appear to involve proteolysis. These include ribosomal function,[5] DNA repair,[6] and the activation of protein kinases such as IκB kinase.[7]

Post-translational attachment of Ub to a target protein typically occurs via covalent attachment of its C-terminal glycine to a lysine residue on the target protein by a multienzymatic system consisting of activating (E1), conjugating (E2), and ligating (E3) enzymes.[8] Since Ub itself carries seven lysine residues, this initial modification is often followed by the glycine-lysine ligation of additional Ub molecules to the first forming di-, tri-, tetra-, or poly-Ub chains. Recent evidence suggests that for several proteins, initial covalent attachment of Ub may occur via the amino group of the N-terminal residue.[9]

Tetra- and poly-Ub chains serve as consensus signals for protein degradation by the 26S proteasome, which is a multisubunit protease with activity specific for ubiquitinated proteins. Within the lumen of the barrel-shapped proteasome, the target protein is cleaved into peptides of 3-20 residues that are further hydrolyzed to amino acids by cytosolic peptidases.[10] The liberated poly-Ub chains are not degraded by the proteasome, but are reused for modification of other proteins (see Figure 2.1).

Several types of low-molecular weight proteasome inhibitors that can readily enter cells are commercially available. The most widely used are peptide aldehydes, such as Cbz-leu-leu-leucinal (or MG132) and Cbz-leu-leu-norvalinal (or MG115). These agents are substrate analogues and act as potent transition state inhibitors.[11,12] For example, the $K_i$ (or binding affinity) of MG132 *in vitro* is a few nanomolar and its $IC_{50}$ (concentration required for 50% inhibition) is in the micromolar range for inhibition of proteolysis in cultured cells.[13]

Enzymatic digestion of ubiquitinated proteins with trypsin results in a signature peptide containing a di-glycine remnant derived from the C-terminus of Ub that is covalently attached to a lysine residue on the target protein.[14,15] This signature peptide has a mass addition of 114.1026 Da as well as a missed cleavage site because trypsin is unable to cleave after the modified lysine residue (see Figure 2.1). Such a uniquely tagged peptide provides a platform for the identification of ubiquitinated proteins by mass spectrometry because the di-glycine remnant is detected as a variable modification on lysine residues after MS/MS analysis and database searching.

Previous studies in yeast have shown that an affinity purification step prior to MS/MS analysis offers significant advantages for studying ubiquitinated proteins. Peng *et al.* identified 72 Ub-protein conjugates from cells expressing 6×His-Ub while Hitchcock *et al.* identified 211 membrane associated ubiquitinated proteins from cells expressing 6×His- Myc-Ub.[15,16] In both cases, large quantities of yeast cells were cultured and epitope-tagged Ub conjugates were isolated. More recently, Kirkpatrick *et al.* adapted this strategy to HEK293 embryonic kidney cells expressing 6×His-Ub-GFP. Results from a series of experiments enabled the identification of 22 Ub-associated proteins, nearly half of which were identified by a single peptide match.[17]

In the present study, we describe a novel method for identifying ubiquitinated proteins using a combination of immunoaffinity purification and LC-MS/MS analysis. Application of this method enabled the confident identification of 70 ubiquitinated proteins from the human MCF-7 breast cancer cell line after overnight treatment with the proteasome inhibitor MG132. Proteins identified include Ub-ligating enzymes, subunits of the 26S proteasome complex, heat shock proteins, transport proteins, DNA repair proteins, transcription and translation elongation factors, and several proteins involved in cell cycle regulation, apoptosis, and signal transduction. Our findings represent the first proteomic analysis of native ubiquitinated proteins from a genetically unmodified mammalian cell line, and we expect our method to be adapted for the study of other biologically relevant samples soon.

## 2.4 Materials and methods

Antibodies and Reagents. The mouse monoclonal anti-Ub antibody (clone FK2) was obtained from BIOMOL Inc. (Plymouth Meeting, PA). This antibody has been shown by western blotting to recognize K29-, K48-, and K63-linked poly-ubiquitinated and mono-ubiquitinated proteins but not free ubiquitin. Purified mouse $IgG_{1\kappa}$ immunoglobulin was obtained from Sigma-Aldrich (St. Louis, MO). Goat antimouse HRP-conjugated secondary antibody was obtained from DAKO (Glostrup, Denmark). MG132 was purchased from Sigma-Aldrich. One mM stock solutions of MG132 were prepared by reconstituting it in dimethyl sulfoxide. All other reagents, unless specified, were obtained from Fisher Scientific (Burlington, ON).

Cell Culture. Human MCF-7 breast epithelial cancer cells were cultured at 37 °C in 5% $CO_2$ in DMEM medium (GIBCOBRL, Burlington, ON) supplemented with 10% fetal calf serum. After reaching 80% confluence, cells were treated overnight with 5-50 μM MG132 in DMEM medium. Cells were harvested, washed twice with phosphate buffered saline (PBS), and resuspended in modified RIPA lysis buffer (50 mM Tris-HCl pH 7.4, 150 mM NaCl, 1% NP-40, 0.25% Na deoxycholate, 1 mM EDTA) containing a cocktail of protease inhibitors (Roche Diagnostics, Laval, QC). Cell lysates were centrifuged for 10 min at 14 000 rpm to pellet cell debris. Supernatants were collected and protein concentration was assessed using a Bradford protein assay kit (Bio-Rad, Hercules, CA).

Immunoaffinity Purification. Anti-Ub beads were prepared by coupling 2 mg of anti-Ub Ab to 1 mL of protein G-agarose beads (Roche) with dimethyl pimelimidate according to manufacturer's instructions (Pierce, Rockford, IL). 0.8 × 4 cm Poly-Prep chromatography columns (Bio-Rad) were then filled with 0.5 mL of anti-Ub beads. Prior to each purification, cell lysates were pre-cleared with columns containing only protein G agarose beads for 10 min at 4 °C. Pre-cleared lysates containing 12.5 mg of total protein were then loaded onto immunoaffinity purification columns and allowed to incubate for 2 h at 4 °C. Columns were washed 10× with 5 mL of modified RIPA buffer. Bound material was eluted with 2% SDS for 20 min at 37 °C. Eluates from three successive purifications were pooled together and concentrated with an Amicon Ultra-4 10 000 MWCO centrifugal filter (Millipore, Nepean, ON). Five times sample buffer (2% SDS, 10% glycerol, 40 mM Tris pH 6.8, 715 mM β-mercaptoethanol when diluted to 1×) was added to the concentrated samples and proteins were separated on a 10% SDS-PAGE gel and visualized with Colloidal Coomassie (Invitrogen).

LC-MS/MS. Protein bands were excised from SDS-PAGE gels and subjected to in-gel tryptic digestion as previously described.[18] Digestions were carried out for 8 h at 37 °C with sequencing grade trypsin (Promega, Madison, WI). Peptides from each gel band were extracted, brought to a final volume of 10 μL with 5% formic acid, and analyzed by LC-MS/MS. An Agilent 1100 series HPLC system (Agilent Technologies, Palo Alto, CA) was used to load peptides at 2 μL/min onto a 75 μm × 50 mm precolumn packed with 5 μm YMC ODS-A C18 beads (Waters, Milford, MA). Following a desalting step, the flow was split and peptides were eluted through a second 75 μm × 50 mm column packed with the same beads at approximately 200 nL/min using a 5-80% gradient of acetonitrile with 0.1% formic acid for 1 h. The LC effluent was electrosprayed into the sampling orifice of a QSTAR Pulsar quadrupole-TOF mass spectrometer (ABI/MDS Sciex, Concord, ON). MS/MS data was then analyzed and matched to human protein sequences in the MSDB database using the Mascot search engine. Peptide and MS/MS mass tolerances were set at ± 100 ppm and 0.2 Da, respectively.

Western Blotting and Immunoprecipitation. Proteins were separated on NuPage 4-12% Bis-Tris pre-cast gels (Invitrogen) and transferred onto nitrocellulose. Membranes were blocked with 2% BSA (BioShop, Burlington, ON) overnight at 4 °C, incubated with 1° Ab for 1 h at RT, washed 4× with 0.1% Tween-20 in PBS for 5 min, and incubated with 2° Ab for 1 h at RT. After another round of washing, bands were visualized with ECL reagent (Amersham Biosciences, Oakville, ON) and BioMax XAR imaging film (Kodak, New Haven, CT). For immunoprecipitations, pre-cleared cell lysates were incubated with 10 μL of anti-Ub beads (described above) for 2 h at 4 °C. After washing 3× times with modified RIPA buffer, bound material was eluted by heating the beads in 2× sample buffer for 5 min at 95 °C.

## 2.5 Results and discussion

Although many substrates of the Ub-proteasome pathway have been individually characterized by biochemical and genetic approaches,[19,20] only a few large-scale studies of ubiquitinated proteins have been reported primarily due to the difficulties encountered in isolating these low-abundance proteins using conventional purification methods. Several groups have reported that when Ub is tagged at its N-terminus with the Myc epitope or poly-Histidine, the isolation of Ub conjugates is greatly facilitated.[21-23] Recently, Gygi and coworkers have adapted this strategy for the purification of 6×His- and 6×His-Myc tagged Ub-conjugates from yeast cells for mass spectrometry analysis.[15,16]

Epitope tagging strategies, however, are limited to cells that have been genetically modified to express tagged Ub and it is known that such molecules do not completely function as their wild-type counterparts.[21] In addition, many false positives are likely to be obtained because expression levels of these molecules may differ significantly compared to normal physiological conditions. Furthermore, when Nickel (Ni)-affinity chromatography is used to purify His-tagged Ub-conjugates, numerous proteins containing multiple His residues in their primary structure, or that have an alignment of His within their secondary or tertiary structures, are co-purified nonselectively.

Identification of ubiquitinated proteins, whether in cell lysates, tissues, or biological fluids, may enable the discovery of relevant protein-based disease and drug biomarkers. This is because degenerative diseases that destroy specific cells, or diseases in which there is a certain level of necrosis, are known to release ubiquitinated proteins into the blood.[24] In addition, drugs that induce apoptosis may also cause the accumulation and release of ubiquitinated proteins into

biological fluids.[25] With this in mind, it was our goal to develop a novel proteomics method for the analysis of native ubiquitinated proteins from a complex biological sample (i.e., a human whole cell lysate) using a combination of immunoaffinity purification, SDS-PAGE separation, and LC-MS/MS analysis. For immunoaffinity purification, a mouse monoclonal anti-Ub antibody (clone FK2) that recognizes both mono- and poly-ubiquitinated species, but not free Ub, would be employed. This antibody has been successfully used to investigate the dynamics of Ub-conjugation and for the development of immunoassays permitting the quantification of intracellular and serum Ub chains.[26-28] A schematic representation of the method is outlined in Figure 2.2.

MCF-7 cells, which serve as a model system for breast cancer cell transformation and proliferation,[29,30] were treated overnight with the proteasome inhibitor MG132 prior to immunoaffinity purification. Biochemical studies have shown that many regulatory proteins are degraded by the Ub-proteasome pathway and because of their rapid degradation, such proteins are often very hard to isolate.[13] Therefore, their stabilization with proteasome inhibitors is a valuable means to enhance cell content and to facilitate their isolation from mammalian cells. As shown in the Western blot in Figure 2.3, treatment of MCF-7 cells with various concentrations of MG132 resulted in increased levels of ubiquitination. A 5 μM concentration was determined to be optimal based on the signal intensity strength and the number of ubiquitinated species that were detected compared to the untreated control. Concentrations above 50 μM were not assessed due to the low survival rate of cells after overnight treatment.

A series of small-scale immunoprecipitations was performed using the anti-Ub antibody to confirm its ability to purify ubiquitinated proteins from MCF-7 cells and to optimize a number

of conditions. This included the ratio of anti-Ub beads for a given amount of starting material, the necessary incubation time to ensure strong antibody-antigen interactions, the number of washing steps to eliminate the bulk of co-purifying background proteins, and the most efficient type of elution buffer. (Anti-Ub beads were prepared by covalently coupling the anti-Ub antibody to Protein G-agarose beads (see Materials and Methods).) The Western blot in Figure 2.3, for example, clearly shows that when 10 μL of anti-Ub beads was incubated with increasing amounts of protein from MCF-7 cell lysates, there was a proportional increase in the signal intensities of the bands detected. This indicated that the anti-Ub beads displayed specificity toward ubiquitinated proteins and enabled the determination that 250 μg of protein per 10 μL of anti-Ub beads was an optimal ratio that could be scaled-up for subsequent purifications. This optimal ratio was deduced as increasing the amount of protein up to 500 μg or above (data not shown) did not result in increases in the intensities of the bands.

To identify ubiquitinated proteins from MCF-7 cells, sufficient amounts of material needed to be purified for LC-MS/MS analysis. Initially, we attempted to pool material from several immunoprecipitations for visualization on a SDS-PAGE gel. Despite a considerable enrichment, as observed by Western blotting, this proved to be inadequate (data not shown). The pooling of multiple samples did little to improve the yield of ubiquitinated proteins and only resulted in an increased amount of nonspecific background. As a result, we decided to employ a large-scale purification approach based on immunoaffinity chromatography to overcome these constraints. In a recent study, such an approach enabled the successful identification of several low-abundance protein complexes by LC-MS/MS.[31]

Poly-Prep chromatography columns were filled with anti-Ub beads and MCF-7 cell lysates were loaded and allowed to incubate at 4 °C. Prior to loading, cell lysates were subjected to a pre-clearing step by passing them through chromatography columns containing only Protein-G agarose beads. After a 2 h incubation period, which allowed sufficient time for ubiquitinated proteins to bind with high-affinity to the anti-Ub beads, a series of stringent washing steps was performed. Bound material was then eluted with a solution containing a low percentage of ionic detergent.

Eluates from three successive purifications were pooled, concentrated, and fractionated by SDS-PAGE (see Figure 2.4). Staining the gel with Colloidal Coomassie revealed that numerous proteins from the two anti-Ub purifications (lanes 2 & 3) were not present in the IgG isotype control (lane 1). Visual inspection of the two anti-Ub purifications also revealed the presence of additional proteins specific to MCF-7 cells that were treated overnight with 5 μM MG132. A corresponding Western blot analysis of 2% of the final eluates is shown in Figure 2.4. As expected, no ubiquitinated proteins were detected when the mouse anti-IgG antibody was used for purification, thus confirming the specificity of the anti-Ub antibody.

For the present study, it is necessary to point out that it was not our intention to perform a comparative or differential analysis of ubiquitinated proteins from MG132-treated and untreated MCF-7 cells. Instead, it was our goal to develop a method for the analysis of ubiquitinated proteins and to validate it using a biologically relevant sample. An important conclusion that can be drawn from a qualitative/visual inspection of the results shown in Figure 2.4 is that application of our immunoaffinity purification protocol will result in the isolation of additional target proteins when there is an increase in the level of ubiquitination in a given sample (e.g.,

after MG132 treatment). Currently, we are developing a gel-free protocol for the differential analysis of multiple samples using LC-MS/MS, and a commercially available mass spectrometry software package. In this way, we will be able to quantitatively monitor the expression levels of ubiquitinated proteins before and after drug treatment.

Thirty bands were excised from the MG132-treated lane shown in Figure 2.4 and were subjected to in-gel tryptic digestion. Peptides were extracted, separated by nanoflow LC, and then introduced into a quadrupole-TOF mass spectrometer MS/MS analysis and database searching resulted in the identification of 70 proteins, each matching with 2 or more unique peptide sequences. Identification of signature peptides (see Figure 2.1) was made possible by selecting an ubiquitinated lysine residue with a monoisotopic mass of 242.13788 Da in the Mascot variable modification list prior to database searching. This mass corresponds to a lysine residue with an attached di-glycine moiety derived from Ub (128.09496 Da + 114.04292 Da).[32] A complete list of the ubiquitinated proteins identified with their corresponding Mascot scores is found in Table 2.1.

Representative MS/MS spectra of three signature peptides are shown Figure 2.5. Doubly charged $[M + 2H]^+$ peptide ions at m/z 698.3 (a), 730.9 (b), and 401.2 (c) were fragmented via collision-induced dissociation. The masses of the fragments were measured and the peptide sequences MQIFVK$_{GG}$TLTGK, LIFAGK$_{GG}$QLEDGR, and EGK$_{GG}$LLK, all of which contain an internal lysine with a di-glycine remnant that is derived from Ub, were obtained. Subsequent database searching matched these peptide sequences to proteins found in bands 5, 12, and 24. In all cases, signature peptides were manually verified to confirm protein ubiquitination.

The two bands present in the IgG control lane in Figure 2.4, with molecular weights of approximately 50 kDa and 40 kDa, were also subjected to LC-MS/MS analysis and database searching. This resulted in the identification of Tubulin and Actin isoforms as co-purifying background proteins (see Table 2.2). As a result, they were removed from consideration as ubiquitinated proteins. Tubulin and Actin have also been reported as commonly observed contaminants with other affinity purifications using agarose and sepharose-based resins.[33,34] Recently, isoforms of each were reported as nonspecific proteins when Ni-NTA-agarose beads were used for the isolation His-tagged GFP-Ub conjugates from HEK293 cells.[17]

Among the proteins listed in Table 2.1 are the E3 and EDD ubiquitin protein ligases, both of which were identified from band 2 with Mascot scores of 769 and 201, respectively (the statistically significant threshold score was 38). These enzymes are attached to Ub within cells and act as readily accessible storage pools of Ub. Also identified were Poly-Ub and a protein containing multiple Ub-like sequences (FLJ32377). Interestingly, both proteins have theoretical molecular weights well below the apparent molecular weight of band 2. Therefore, it is likely that they comigrated with either of the two Ub-ligating enzymes and were still covalently attached due to incomplete tryptic digestion. A similar explanation is likely for all of the bands that were found to contain Poly-Ub and/or Ub.

Numerous proteins belonging to the 26S proteasome complex are listed in Table 2.1, including 6 ATPase and 3 non-ATPase regulatory subunits. Each subunit is known to interact with a number of related regulatory proteins as well as a distinct subset of cytosolic, nuclear, and/or membrane proteins that are known to be ubiquitinated. For example, the 26S proteasome non-ATPase regulatory subunit 2, identified from Band 9 with a Mascot score of 707, is known

to directly bind to ATPase regulatory subunits 1 and 2, non-ATPase regulatory subunit 5, the E3 ubiquitin protein ligase, and the intracellular domain of the Tumor Necrosis Factor (TNF)-type 1 receptor.[35-37]

Eleven members of the heat-shock protein family and three members of chaperonin protein family were identified from bands 11, 13, and 15-19. This was not surprising as the heat shock response is known to be directly regulated through the Ub-proteasome pathway. By tagging and degrading unfolded or damaged proteins, the Ub-proteasome pathway helps to prevent the accumulation of heat shock proteins under normal conditions.[13] However, when this degradative process is blocked by a proteasome inhibitor such as MG132, there is an induction of heat shock proteins and molecular chaperones (e.g., BiP and glucose-regulated stress proteins) which will in turn bind to Ub and Ub-tagged proteins.[38,39]

All other ubiquitinated proteins that were identified are listed in Table 2. This includes thirteen metabolic enzymes, eight DNA repair proteins, five cell cycle and apoptosis-related proteins, five proteins involved in signal transduction, four transcription and translation elongation factors, four transport proteins, four structural proteins, and two hypothetical proteins with currently unknown functions. A pie-chart distribution of these proteins into different functional categories is shown in Figure 2.6. As expected, the diversity of proteins identified in this study confirms that the important role that ubiquitination plays in the regulation of many key cellular functions in MCF-7 cells.

## 2.6 Conclusion

Taken together, our findings represent the first proteomic analysis of native ubiquitinated proteins from a genetically unmodified/untransfected mammalian cell line. Our method was shown to enable the profiling of ubiquitinated proteins from human MCF-7 cells under normal physiological conditions and/or after drug treatment. Further work is underway to improve the sensitivity of the method in order to permit the identification of very low-abundance ubiquitinated proteins which may serve as biomarkers of disease progression or as targets of novel therapeutics that interfere with the Ub-proteasome pathway. We expect our method to soon be adapted for the study of ubiquitinated proteins in other complex biological material, including human serum and plasma samples.

**Figure 2.1 Ubiquitin-proteasome pathway and tryptic digestion of ubiquitinated proteins.** (a) Post-translational modification of proteins by Ub occurs via covalent attachment of its C-terminal glycine to a lysine on the target protein by 3 enzymes (E1, E2, E3). Ubiquitination serves as a consensus signal for protein degradation by the 26S proteasome complex. (b) After tryptic digestion, ubiquitinated proteins contain a di-glycine remnant attached to a lysine residue. A signature peptide with a 114.1026 Da mass increase and a missed cleavage site can be identified by MS/MS.

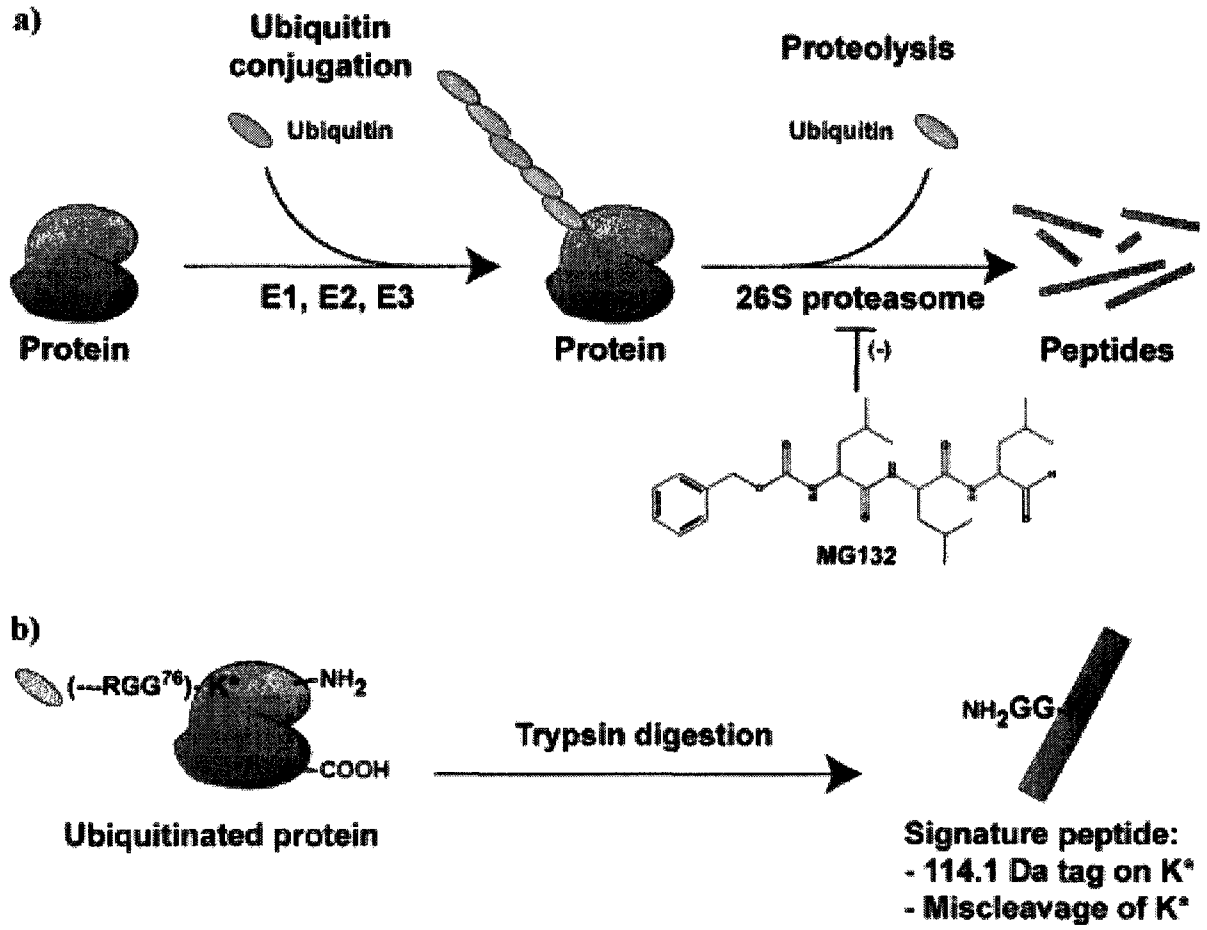


**Figure 2.2 Experimental method.** (1) MCF7 cells were treated with MG132, or were left untreated. (2) Ubiquitinated proteins were isolated by immunoaffinity purification. (3) Proteins were separated by SDS-PAGE and visualized with Colloidal Coomassie. Gel bands were excised and subjected to in-gel tryptic digestion. (4) Peptides were analyzed by LC-MS/MS. (5) MS/MS spectra were matched to human protein sequences after database searching and were manually verified to confirm ubiquitination sites.

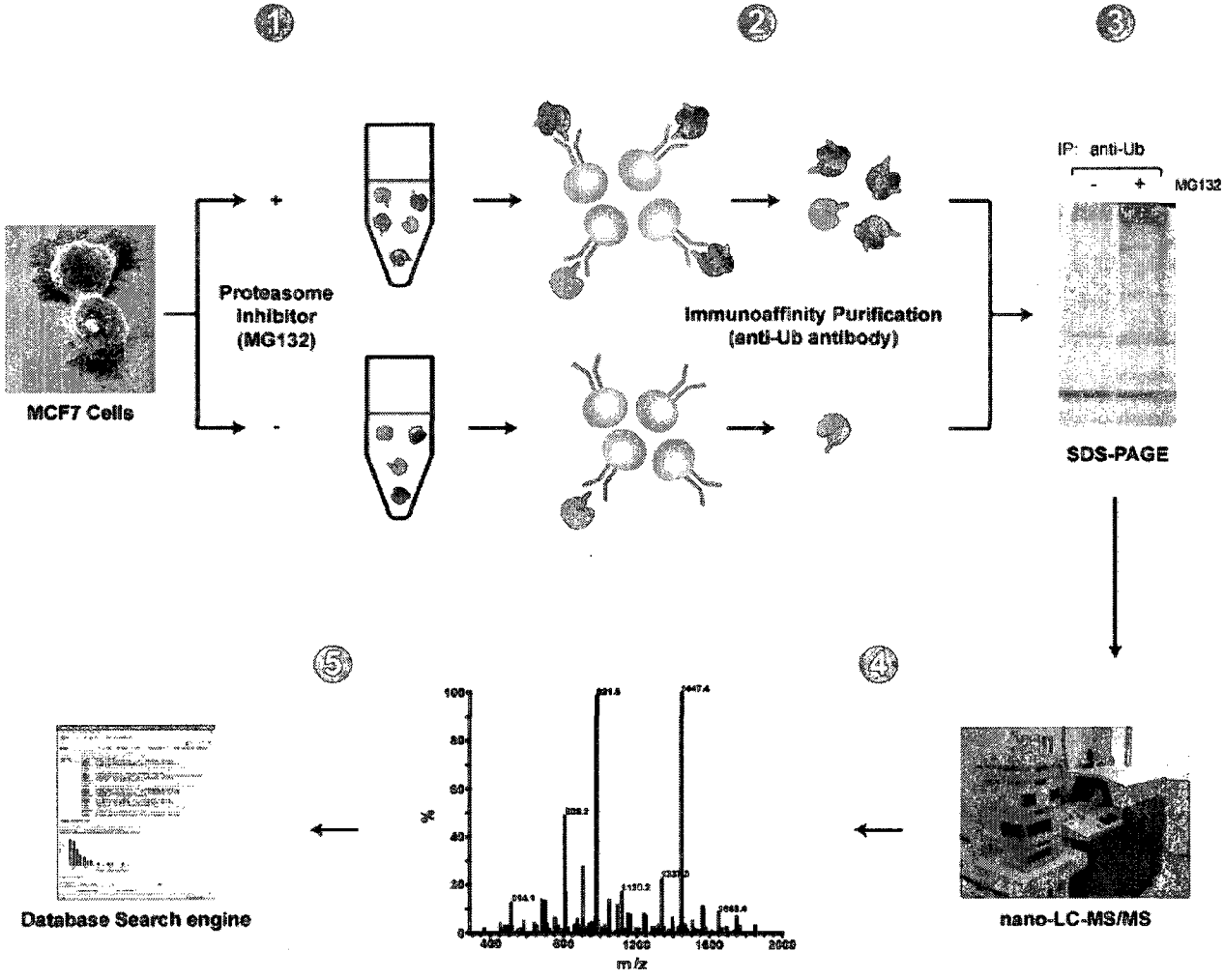


**Figure 2.3 Treatment of MCF-7 cells with the proteasome inhibitor MG132.** (a) MCF7 cells were treated overnight with various concentrations (5-50 μM) of MG132. Cell lysates containing 40 μg of protein were analyzed by Western blotting with an anti-Ub antibody to monitor differential levels of ubiquitination. (b) Small-scale immunoprecipitations with anti-Ub beads (see materials and methods) and MCF7 cell lysates were performed and analyzed by Western blotting.

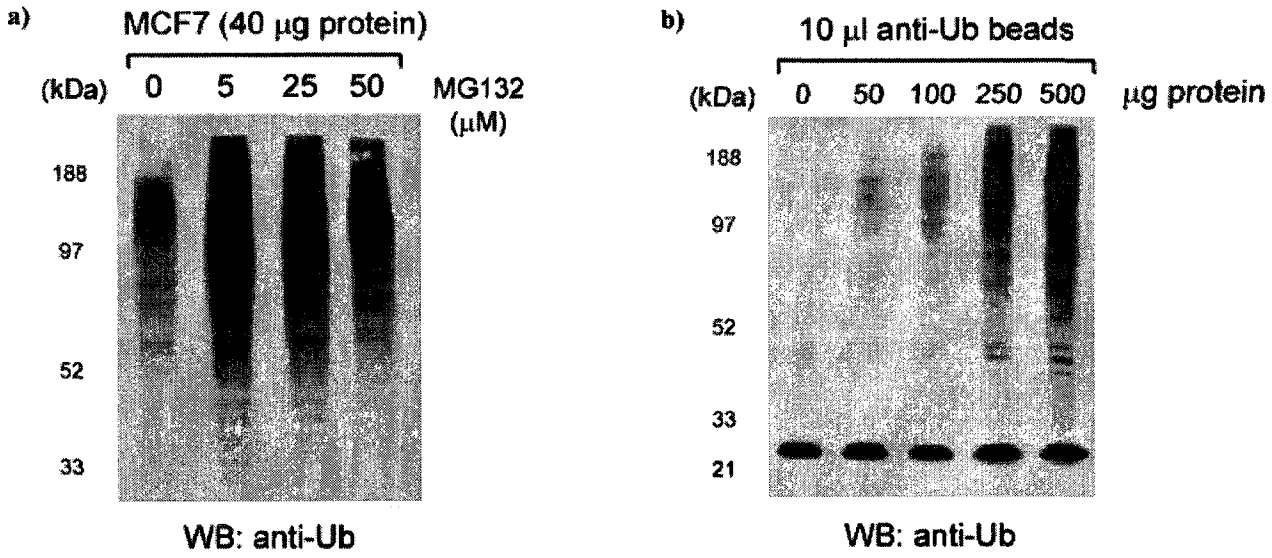


**Figure 2.4 Immunoaffinity purification of ubiquitinated proteins from MCF-7 cells.** (a) Ubiquitinated proteins from untreated MCF7 cells (lane 2) and MG132-treated MCF7 cells (lane 3) were purified by immunoaffinity chromatography, separated by SDS-PAGE, and visualized with Colloidal Coomassie. (b) Aliquots containing 2% of the purified samples were also analyzed by Western blotting.

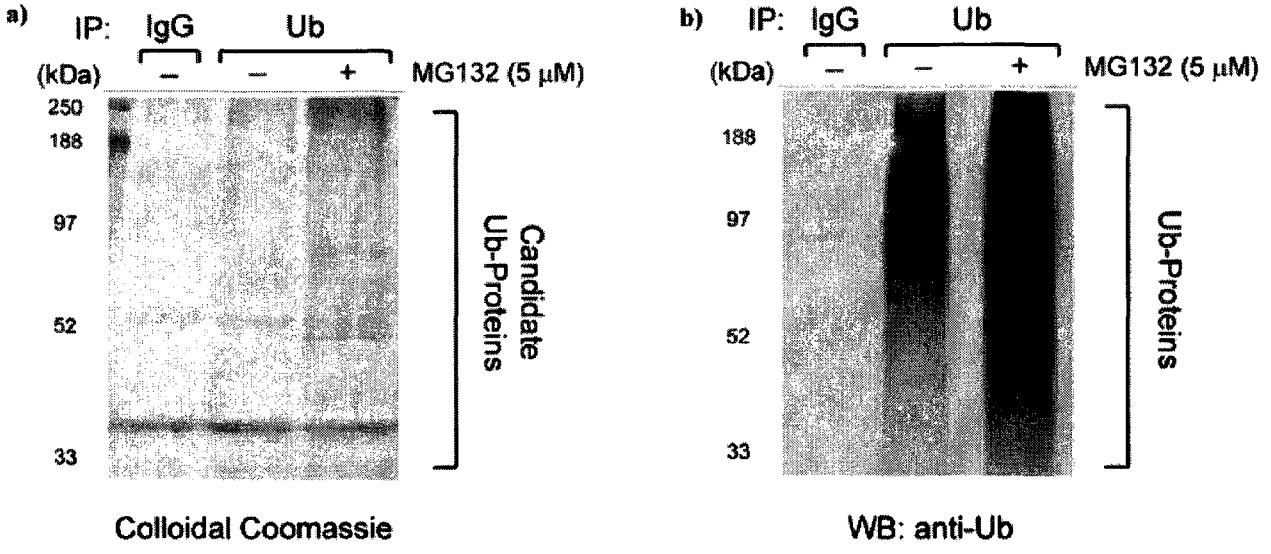


**Figure 2.5 Representative MS/MS spectra of signature peptides.** MS/MS spectra of doubly charged $[M + 2H]^+$ peptide ions at m/z 698.3 (a), 730.9 (b), and 401.2 (c). The peptide sequences MQIFVK$_{GG}$TLTGK, LIFAGK$_{GG}$QLEDGR, and EGK$_{GG}$LLK all contain an internal lysine with a di-glycine remnant that was derived from Ub.

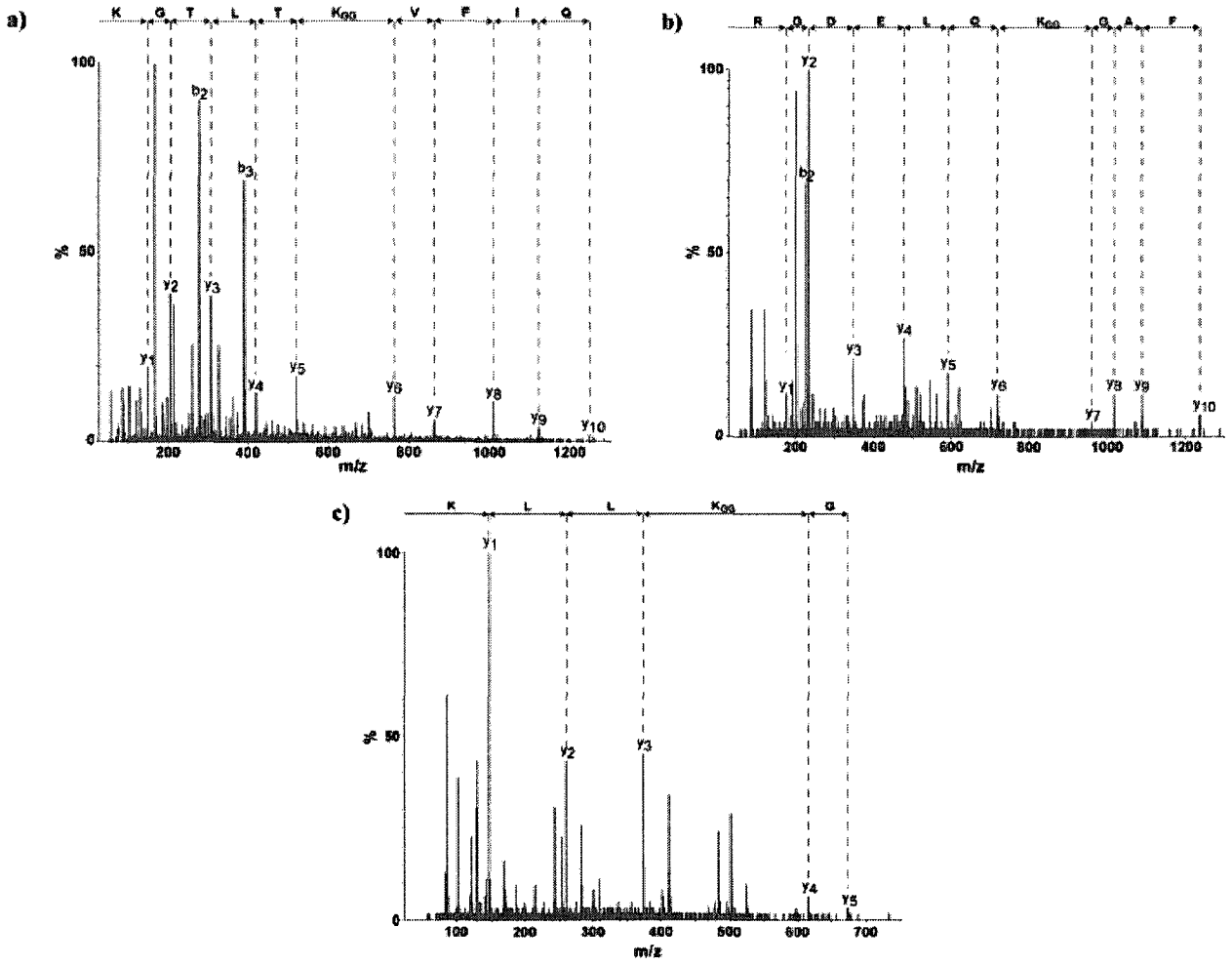
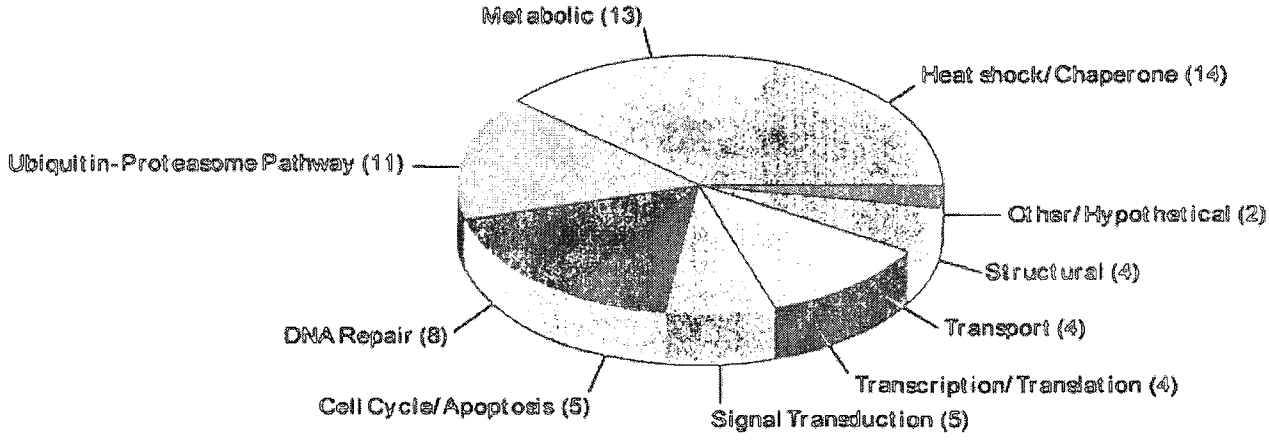


Figure 2.6 Pie-chart distribution of ubiquitinated proteins into functional categories. Ubiquitinated proteins identified from MG132-treated MCF-7 cells were grouped into 10 functional categories.



## Table 2.1 Ubiquitinated proteins from MCF7 cells identified by LC-MS/MS.

| band | protein | accession no. | Mascot score | no. of peptides | $M_w$ (kDa) |
|---|---|---|---|---|---|
| 1 | PolyUbiquitin | Q9UEG1 | 297 | 4 | 68.4 |
| 1 | E3 ubiquitin protein ligase | Q8NG67 | 94 | 4 | 372.5 |
| 2 | Ubiquitin protein ligase EDD | O95071 | 769 | 19 | 312.3 |
| 2 | FLJ32377 *Ubiquitin-like | Q96MH4 | 681 | 12 | 43.6 |
| 2 | PolyUbiquitin | Q9UEG1 | 624 | 11 | 68.4 |
| 2 | FLJ00343 | Q8NF52 | 258 | 6 | 283.9 |
| 2 | Fatty acid synthase | P49327 | 203 | 5 | 275.5 |
| 2 | E3 ubiquitin protein ligase | Q8NG67 | 201 | 5 | 372.5 |
| 2 | DNA-dependent protein kinase catalytic subunit | P78527 | 143 | 3 | 470.0 |
| 3 | Fatty acid synthase | P49327 | 753 | 17 | 275.5 |
| 3 | PolyUbiquitin | Q9UEG1 | 495 | 9 | 68.4 |
| 3 | Glutamine-dependent carbamoyl-phosphate synthase | P27708 | 128 | 2 | 245.1 |
| 4 | Fatty-acid synthase | P49327 | 856 | 23 | 275.5 |
| 4 | Filamin A | P21333 | 520 | 13 | 283.3 |
| 4 | PolyUbiquitin | Q9UEG1 | 376 | 6 | 68.4 |
| 4 | Filamin B | Q9UEV9 | 80 | 3 | 280.1 |
| 5 | Ubiquitin | P62958 | 272 | 5 | 8.5 |
| 5 | Eukaryotic translation initiation factor 4-gamma 1 | Q8N102 | 59 | 2 | 176.1 |
| 6 | Ubiquitin | P62958 | 475 | 7 | 8.5 |
| 6 | PolyUbiquitin | Q9UEG1 | 472 | 7 | 68.4 |
| 6 | Myosin VI | Q9UM54 | 208 | 5 | 149.9 |
| 6 | co-atomer protein complex, subunit alpha | Q8IXZ9 | 144 | 5 | 140.8 |
| 6 | Splicing factor 3B, subunit 1 | O75533 | 93 | 2 | 146.4 |
| 7 | PolyUbiquitin | Q9UEG1 | 367 | 5 | 68.4 |
| 7 | HLA-B associated transcript 3, isoform b | Q9BCN4 | 190 | 3 | 119.0 |
| 8 | KIAA0090 | Q8N766 | 443 | 10 | 112.1 |
| 8 | Ubiquitin | P62958 | 375 | 5 | 8.5 |
| 8 | 26S proteasome non-ATPase regulatory subunit 1 | Q86VU1 | 165 | 4 | 94.1 |
| 9 | 26S proteasome non-ATPase regulatory subunit 2 | Q13200 | 707 | 17 | 100.8 |
| 9 | Transitional endoplasmic reticulum ATPase | P55072 | 627 | 14 | 89.9 |
| 9 | Polyubiquitin | Q9UEG1 | 304 | 5 | 68.4 |
| 10 | 26S proteasome non-ATPase regulatory subunit 2 | Q13200 | 1398 | 26 | 100.8 |
| 10 | Transitional endoplasmic reticulum ATPase | P55072 | 702 | 15 | 89.9 |
| 10 | Ubiquitin | P62958 | 204 | 3 | 8.5 |
| 10 | Glycogen phosphorylase | P11216 | 109 | 2 | 97.3 |
| 11 | Heat shock protein HSP 90-beta | P08238 | 662 | 15 | 83.4 |
| 11 | Heat shock protein HSP 90-alpha | P07900 | 571 | 12 | 84.8 |
| 11 | Ubiquitin | P62958 | 327 | 4 | 8.5 |
| 12 | Ubiquitin | P62958 | 342 | 5 | 8.5 |
| 12 | ATP-dependent helicase DDX1 | Q92499 | 168 | 3 | 83.3 |
| 12 | Werner helicase interacting protein, isoform 2 | Q8WV26 | 138 | 3 | 70.1 |
| 12 | Ewing sarcoma breakpoint region 1, isoform EWS | Q96FE8 | 100 | 2 | 68.6 |
| 12 | DNA replication licensing factor MCM5 | P33992 | 99 | 3 | 83.0 |
| 12 | Signal transducer and activator of transcription 1 | P42224 | 60 | 2 | 83.5 |
| 13 | Ubiquitin | P62958 | 371 | 6 | 8.5 |
| 13 | Werner helicase interacting protein, isoform 2 | Q8WV26 | 354 | 7 | 70.1 |
| 13 | DNA replication licensing factor MCM7 | P33993 | 213 | 6 | 81.8 |
| 13 | Ewing sarcoma breakpoint region 1, isoform EWS | Q96FE8 | 193 | 3 | 68.6 |
| 13 | Vacuolar protein sorting 35 | Q96QK1 | 136 | 3 | 92.4 |
| 13 | Heat shock protein HSP 90-beta | P08238 | 136 | 3 | 83.4 |
| 13 | 6-phosphofructokinase, type C | Q01813 | 92 | 3 | 86.4 |
| 13 | Endoplasmin | P14625 | 85 | 2 | 92.6 |
| 13 | Junction plakoglobin | P14923 | 59 | 2 | 82.2 |
| 14 | Werner helicase interacting protein, isoform 2 | Q8WV26 | 275 | 6 | 70.1 |
| 14 | Ubiquitin | P62958 | 227 | 4 | 8.5 |
| 14 | Ewing sarcoma breakpoint region 1, isoform EWS | Q96FE8 | 136 | 2 | 68.6 |
| 14 | DNA replication licensing factor MCM7 | P33993 | 119 | 4 | 81.8 |
| 14 | 6-phosphofructokinase | Q01813 | 48 | 2 | 86.4 |
| 15 | BiP protein | Q9UK02 | 298 | 8 | 71.0 |
| 15 | Ubiquitin | P62958 | 283 | 5 | 8.5 |
| 15 | Heat shock 70kD protein 9B | Q8N1C8 | 182 | 3 | 74.0 |
| 15 | Heat shock 70 kDa protein 1 | P08107 | 129 | 4 | 70.2 |
| 15 | Heat shock 70 kDa protein 6 | P17066 | 96 | 3 | 71.2 |
| 16 | Heat shock cognate 71 kDa protein | P11142 | 814 | 17 | 71.0 |
| 16 | Stress-70 protein | P38646 | 574 | 12 | 73.9 |
| 16 | Heat shock 70 kDa protein 6 | P17066 | 249 | 5 | 71.2 |
| 16 | Heat shock 70 kDa protein 1 | P08107 | 235 | 5 | 70.2 |
| 16 | Ubiquitin | P62958 | 204 | 3 | 8.5 |
| 17 | Heat shock 70 kDa protein 1 | P08107 | 1003 | 18 | 70.2 |
| 17 | Heat shock 70 kDa protein 1L | Q8NE72 | 364 | 7 | 70.7 |
| 17 | Heat shock 70 kDa protein 6 | P17066 | 338 | 5 | 71.2 |
| 17 | Ubiquitin | P62958 | 189 | 3 | 8.5 |
| 17 | Heat shock cognate 71 kDa protein | P11142 | 151 | 3 | 71.0 |
| 17 | Similar to ribophorin I | Q96HX3 | 120 | 3 | 64.6 |
| 17 | Heat shock-related 70 kDa protein 2 | P54652 | 114 | 3 | 70.2 |
| 18 | Heat shock 70 kDa protein 1 | P08107 | 396 | 8 | 70.2 |
| 18 | Heat shock cognate 71 kDa protein | P11142 | 257 | 5 | 71.0 |
| 18 | Ubiquitin | P62958 | 245 | 5 | 8.5 |

**Table 2.1 continued**

| band | protein | accession no. | Mascot score | no. of peptides | $M_w$ (kDa) |
|---|---|---|---|---|---|
| 18 | Heat shock 70 kDa protein 6 | P17066 | 238 | 5 | 71.2 |
| 18 | Similar to ribophorin I | Q96HX3 | 82 | 2 | 64.6 |
| 18 | Programmed cell death protein 8 | O95831 | 71 | 2 | 67.1 |
| 19 | Ubiquitin | P62958 | 202 | 4 | 8.5 |
| 19 | T-complex protein 1, alpha subunit | P17987 | 100 | 3 | 60.8 |
| 19 | 60 kDa heat shock protein | P10809 | 98 | 3 | 61.1 |
| 20 | Pyruvate kinase, isozymes M1/M2 | P14618 | 351 | 8 | 58.3 |
| 20 | Ubiquitin | P62958 | 227 | 4 | 8.5 |
| 21 | PolyUbiquitin | Q9UEG1 | 272 | 4 | 68.4 |
| 21 | 26S protease regulatory subunit 4 | P62191 | 197 | 4 | 49.3 |
| 21 | P60 protein | Q13446 | 171 | 4 | 48.3 |
| 21 | T-complex protein 1, delta subunit | P50991 | 167 | 3 | 58.3 |
| 21 | Pyruvate kinase, isozymes M1/M2 | P14618 | 142 | 3 | 58.3 |
| 21 | T-complex protein 1, eta subunit | Q99832 | 77 | 2 | 59.8 |
| 22 | 26S protease regulatory subunit 4 | P62191 | 491 | 11 | 49.3 |
| 22 | Pyruvate kinase, isozymes M1/M2 | P14618 | 467 | 9 | 58.3 |
| 22 | P60 protein | Q13446 | 449 | 8 | 48.3 |
| 22 | Phosphotyrosine independent ligand p62B | Q13502 | 272 | 5 | 46.4 |
| 22 | T-complex protein 1, delta subunit | P50991 | 249 | 7 | 58.3 |
| 22 | Ubiquitin | P62958 | 198 | 4 | 8.5 |
| 22 | FBP-interacting repressor | Q9NZA0 | 77 | 2 | 58.3 |
| 23 | 26S protease regulatory subunit 4 | P62191 | 608 | 11 | 49.3 |
| 23 | Ubiquitin | P62958 | 259 | 4 | 8.5 |
| 23 | Glucose-6-phosphate 1-dehydrogenase | P11413 | 142 | 4 | 59.7 |
| 23 | UDP-glucose 6-dehydrogenase | O60701 | 123 | 3 | 55.6 |
| 23 | T-complex protein 1, beta subunit | P78371 | 91 | 2 | 57.7 |
| 23 | UV excision repair protein RAD23 homologue B | P54727 | 73 | 2 | 43.2 |
| 24 | 26S protease regulatory subunit 6A | P17980 | 617 | 12 | 55.2 |
| 24 | 26S protease regulatory subunit 6B | P43686 | 572 | 13 | 47.4 |
| 24 | Eukaryotic translation elongation factor 1 alpha 1 | Q6IPN6 | 445 | 9 | 50.4 |
| 24 | Ubiquitin | P62958 | 272 | 4 | 8.5 |
| 24 | Alpha enolase | P06733 | 251 | 5 | 47.3 |
| 24 | 52 kDa Ro protein | P19474 | 101 | 2 | 55.1 |
| 24 | 26S protease regulatory subunit 7 | P35998 | 73 | 2 | 48.8 |
| 24 | Gamma enolase | P09104 | 68 | 2 | 47.4 |
| 24 | Chromosome 1 open reading frame 26 | Q8NEK9 | 67 | 2 | |
| 25 | 26S protease regulatory subunit 6A | P17980 | 363 | 7 | 55.2 |
| 25 | 26S protease regulatory subunit 6B | P43686 | 227 | 5 | 47.4 |
| 25 | Alpha enolase | P06733 | 89 | 2 | 47.3 |
| 25 | Eukaryotic translation elongation factor 1 alpha 1 | Q6IPN6 | 81 | 2 | 50.4 |
| 26 | 26S protease regulatory subunit 7 | P35998 | 474 | 9 | 48.8 |
| 26 | 26S protease regulatory subunit 8 | P62196 | 473 | 10 | 45.7 |
| 26 | Ubiquitin | P62958 | 225 | 4 | 8.5 |
| 26 | Elongation factor Tu | P49411 | 116 | 2 | 49.8 |
| 26 | Eukaryotic translation elongation factor 1 alpha 1 | Q6IPN6 | 93 | 2 | 50.4 |
| 27 | 26S protease regulatory subunit 8 | P62196 | 366 | 9 | 45.7 |
| 27 | Ubiquitin | P62958 | 232 | 4 | 8.5 |
| 27 | Elongation factor Tu | P49411 | 205 | 5 | 49.8 |
| 27 | Calcium-binding transporter | Q9P129 | 95 | 2 | 46.0 |
| 28 | Ubiquitin | P62958 | 261 | 4 | 8.5 |
| 28 | 26S protease regulatory subunit 8 | P62196 | 190 | 4 | 45.7 |
| 28 | Ubiquinol-cytochrome-c reductase complex core protein 2 | P22695 | 98 | 2 | 48.5 |
| 28 | 26S proteasome non-ATPase regulatory subunit 6 | Q15008 | 97 | 2 | 45.7 |
| 28 | Isocitrate dehydrogenase | P48735 | 84 | 2 | 51.3 |
| 29 | 26S protease regulatory subunit S10B | P62333 | 501 | 11 | 44.4 |
| 29 | Ubiquitin | P62958 | 226 | 4 | 8.5 |
| 29 | Guanine nucleotide-binding protein G-s-alpha-3 | Q14433 | 134 | 3 | 44.6 |
| 30 | Ubiquitin | P62958 | 176 | 3 | 8.5 |
| 30 | Fructose-bisphosphate aldolase A | P04075 | 125 | 3 | 39.8 |

Note: Lane 3 of the gel shown in Figure 2.4 (MG132-treated MCF7 cells) was excised and sectioned into 30 bands of equal size for MS/MS analysis.

## Table 2.2 Co-purified background proteins.

| band | protein | accession no. | $M_w$ (kDa) |
|---|---|---|---|
| 1 | Class IVb beta tubulin | Q8IWP6 | 50.1 |
| 1 | Tubulin beta-2 chain | P07437 | 50.0 |
| 1 | Tubulin alpha-ubiquitous chain | P68363 | 50.8 |
| 2 | Actin, beta | Q96HG5 | 41.3 |
| 2 | Actin-like protein 2 | P61160 | 45.0 |

## 2.7 References

(1) Koepp, D. M.; Harper, J. W.; Elledge, S. J. Cell. 1999, 97, 431- 334.

(2) Fujimoro, M.; Sawada, H.; Yokosawa, H. Eur. J. Biochem. 1997, 249, 427-433.

(3) Rock, K. L.; Goldberg, A. L. Annu. Rev. Immunol. 1999, 17, 739-779.

(4) Glickman, M. H.; Ciechanover, A. Physiol Rev. 2002, 82, 373- 428.

(5) Spence, J.; Gali, R.; Dittmar, G.; Sherman, F.; Karin, M.; Finley, D. Cell. 2000, 102, 67-76.

(6) Hofmann, R. M.; Pickart, C. M. Cell. 1999, 96, 645-653.

(7) Sun, L.; Chen, Z. J. Curr. Opin. Cell Biol. 2004, 16, 119-126.

(8) Schwartz, A. L.; Ciechanover, A. Annu. Rev. Med. 1999, 50, 57- 74.

(9) Ciechanover, A.; Ben-Saadon, R. Trends Cell Biol. 2004, 14, 103-106.

(10) Kisselev, A. F.; Akopian, T. N.; Goldberg, A. L. J. Biol. Chem. 1998,273, 1982-1989.

(11) Rock, K. L. Gramm, C.; Rothstein, L.; Clark, K.; Stein, R.; Dick, L.; Hwang, D.; Goldberg, A. L. Cell. 1994, 78, 761-771.

(12) Lee, D. H.; Goldberg, A. L. J. Biol. Chem. 1996, 271, 27280-27284.

(13) Lee, D. H.; Goldberg, A. L. Trends Cell Biol. 1998, 8, 397-403.

(14) Wang, D.; Cotter, R. J. Anal. Chem. 2005, 77, 1458-1466.

(15) Peng, J.; Schwartz, D.; Elias, J. E.; Thoreen, C. C.; Cheng, D.; Marsischky, G.; Roelofs, J.; Finley, D.; Gygi, S. P. Nat. Biotech. 2003, 21, 921-926.

(16) Hitchcock, A. L.; Auld, K.; Gygi, S. P.; Silver, P. A. Proc. Natl. Acad. Sci. U.S.A. 2003, 100, 12735-12740.

(17) Kirkpatrick, D. S.; Weldon, S. F.; Tsaprailis, G.; Liebler, D. C.; Gandolfi, A. J. Proteomics 2005, 5, 2104-2111.

(18) Wilm, M.; Shevchenko, A.; Houthaeve, T.; Breit, S.; Schweigerer, L.; Fotsis, T.; Mann, M. Nature 1996, 379, 466-469.

(19) Smith, S. E.; Koegl, M.; Jentsch, S. Biol. Chem. 1996, 377, 437- 446.

(20) Hochstrasser, M. Annu. Rev. Genet. 1996, 30, 405-439.

(21) Ellison, M. J.; Hochstrasser, M. J. Biol. Chem. 1991, 266, 21150- 21157.

(22) Beers, E. P.; Callis, J. J. Biol. Chem. 1993, 268, 21645-21649.

(23) Ling, R.; Colon, E.; Dahmus, M. E.; Callis, J. Anal. Biochem. 2000, 282, 54-64.

(24) Majetschak, M.; Krehmeier, U.; Bardenheuer, M.; Denz, C.; Quintel, M.; Voggenreiter, G.; Obertacke, U. Blood. 2003, 101, 1882-1890.

(25) Canu, N.; Barbato, C.; Ciotti, M. T.; Serafino, A.; Dus, L.; Calissano, P. J Neurosci. 2000, 20, 589-99.

(26) Fujimoro, M.; Sawada, H.; Yokosawa, H. Eur. J. Biochem. 1997, 249, 427-433.

(27) Takada, K.; Nasu, H.; Hibi, N.; Tsukada, Y.; Ohkawa, K.; Fujimoro, M.; Sawada, H.; Yokosawa, H. Eur. J. Biochem. 1995, 233, 42-47.

(28) Takada, K.; Nasu, H.; Hibi, N.; Tsukada, Y.; Shibasaki, T.; Fukise, K.; Fujimoro, M.; Sawada, H.; Yokosawa, H.; Ohkawa, K. Clin. Chem. 1997, 43, 1188-1195.

(29) Soule, H. D.; Vazquez, J.; Albert, S.; Brenan, S. J. Natl. Cancer Inst. 1973, 51, 1409-1413.

(30) Keen, J. C.; Davidson, N. E. Cancer 2003, 97, 825-833.

(31) Vasilescu, J.; Guo, X.; Kast, J. Proteomics 2004, 4, 3845-3854.

(32) UNIMOD: protein modifications for mass spectrometry. http://www.unimod.org.

(33) Shevchenko, A.; Shaft, D.; Roguev, A.; Pijnappel, W. W.; Steward, A. F.; Shevchenko, A. Mol. Cell. Proteomics 2002, 1, 204-212.

(34) Kocks, C.; Maehr, R.; Overkleeft, H. S.; Wang, E. W.; Iyer, L. K.; Lennon-Dummenil, A. M.; Ploegh, H. L.; Kessler, B. M. Mol. Cell. Proteomics 2003, 2, 1188-1197.

(35) Gorbea, C.; Taillandier, D.; Rechsteiner, M. J. Biol. Chem. 2000, 275, 875-882.

(36) You, J.; Wang, M.; Aoki, T.; Tamura, T. A.; Pickart, C. M. J. Biol. Chem. 2003, 278, 23369-23375.

(37) Dunbar, J. D.; Song, H. Y.; Guo, D.; Wu, L. W.; Donner, D. B. J. Immunol. 1997, 158, 4252-4259.

(38) Zhou, M.; Wu, X.; Ginsberg, H. N. J. Biol. Chem. 1996, 271, 24769- 24775.

(39) Bush, K. T.; Goldberg, A. L.; Nigam, SK. J. Biol. Chem. 1996, 271, 24769-24775.

# Chapter 3: The proteomic reactor facilitates the analysis of affinity-purified proteins by mass Spectrometry: application for identifying ubiquitinated proteins in human cells

## 3.1 Manuscript status and statement of author contributions

A version of this chapter has been published in the Journal of Proteome Research: *Vasilescu J, Zweitzig D, Denis NJ, Smith JC, Ethier M, Haines DS, Figeys D. The proteomic reactor facilitates the analysis of affinity-purified proteins by mass spectrometry: application for identifying ubiquitinated proteins in human cells. Journal of Proteome Research. Vol. 6: 298-305, 2007.* JV conceived, designed and performed most of the experiments. JV also wrote the manuscript. DZ performed the cell culture and siRNA experiments. NJD processed samples on the proteomic reactor. JCS and ME assisted with the data analysis. DSH and DF edited the manuscript.

## 3.2 Summary

Mass spectrometry (MS) coupled to affinity purification is a powerful approach for identifying protein-protein interactions and for mapping post-translational modifications. Prior to MS analysis, affinity purified proteins are typically separated by gel electrophoresis, visualized with a protein stain, excised, and subjected to in-gel digestion. An inherent limitation of this series of steps is the loss of protein sample that occurs during gel processing. Although methods employing in-solution digestion have been reported, they generally suffer from poor reaction

kinetics. In the present study, we demonstrate an application of a microfluidic processing device, termed the Proteomic Reactor, for enzymatic digestion of affinity-purified proteins for liquid chromatography tandem mass spectrometry (LC-MS/MS) analysis. Use of the Proteomic Reactor enabled the identification of numerous ubiquitinated proteins in a human cell line expressing reduced amounts of the ubiquitin-dependent chaperone, valosin-containing protein (VCP). The Proteomic Reactor is a novel technology that facilitates the analysis of affinity-purified proteins and has the potential to aid future biological studies.

## 3.3 Introduction

When mass spectrometry (MS) is coupled to affinity purification techniques, such as immunopurification, tandem affinity purification (TAP), and metal-affinity chromatography, components of targeted protein complexes may be identified.[1,2] In addition, key post-translational modifications that regulate the interaction of proteins within a complex may be mapped to specific amino acid residues. MS coupled to affinity purification has also been used in large-scale proteomic studies to map glycosylation, phosphorylation, and ubiquitination sites in yeast and human cells.[3-6]

Prior to MS analysis, affinity-purified proteins are typically separated by 1-D or 2-D gel electrophoresis, visualized with a MS-compatible stain (e.g., Coomassie, silver nitrate, or sypro ruby), excised, and subjected to in-gel digestion.[7] Limitations of this approach include the loss of sample that occurs during gel processing and the possibility of introducing contaminants. Although gel-free protein separation methods are available, including size-exclusion

chromatography and free-flow electrophoresis,[8,9] in-solution digestions need to be performed for every fractionated sample. Furthermore, when affinity-purified proteins are present in low (submicromolar) concentrations, in-solution digestion efficiencies are hindered because most proteolytic enzymes have $K_m$ values in the order of 5-50 μM.[10] To compensate for this, enzyme concentrations may be increased, but this results in an increased rate of enzyme autoproteolysis, which can suppress MS signals for the protein(s) of interest.

Alternatives to in-solution digestion have been reported, but they were evaluated with standard proteins and have yet to be tested with biologically relevant samples. For example, Doucette *et al.* developed a method for immobilizing proteins on polymeric microbeads and digesting them with trypsin. Peptides were analyzed by MS after depositing the beads directly on a MALDI target.[11] This procedure was demonstrated using cytochrome c, lysozyme, and albumin. Craft *et al.* later modified this approach by preparing microcolumns packed with hydrophobic beads and loaded standard proteins on these columns. Bound proteins were digested with trypsin, and peptides were eluted and analyzed by MS.[12] Recently, a complex automated version of this method was reported for liquid chromatography-tandem mass spectrometry (LC-MS/MS) that employs numerous pumps and switching valves.[13]

In the following study, we demonstrate an application of a microfluidic processing device, termed the Proteomic Reactor,[14] for rapid preconcentration, derivatization, and enzymatic digestion of affinity-purified proteins for LC-MS/MS analysis. Briefly, the Proteomic Reactor consists of a short length of fused silica capillary tubing packed with strong cation exchange (SCX) resin. Proteins samples are acidified, mixed with trypsin, and loaded on the Proteomic Reactor. At low pH, most of the proteins are positively charged and bind to the SCX resin. After

disulfide bond reduction, the pH is increased to permit simultaneous alkylation and enzymatic digestion. Peptides are then eluted in a buffer that is compatible with LC-MS/MS analysis.[14]

We integrated the Proteomic Reactor with a method that we have previously developed[15] and applied this novel approach to identify ubiquitinated proteins in human cells harboring reduced levels of the valosin-containing protein (VCP). VCP is a highly abundant member of the AAA ATPase family.[16,17] VCP participates in numerous cellular processes, including transcription, DNA repair, mitosis, membrane fusion, and ER-associated degradation of misfolded proteins, and these functions are mediated via ubiquitin adaptor proteins that target VCP to ubiquitinated substrates.[18] Although the biochemical functions of VCP containing complexes are poorly understood, they are thought to function in the separation of ubiquitinated substrates from unmodified binding partners, the retrotranslocation of ubiquitinated proteins from membranes, and/or the delivery of ubiquitinated proteins to the proteasome.[19,20]

Use of the Proteomic Reactor enabled 27 ubiquitination sites to be precisely mapped on 21 proteins and enabled the identification of 58 candidate ubiquitinated proteins from control and VCP depleted H1299 lung adenocarcinoma cells. Differences in ubiquitination were also observed when two populations of H1299 cells (i.e. -/+ VCP depletion) were compared. Several proteins designated as candidates were confirmed to be ubiquitinated by subsequent immunoprecipitation and Western blotting experiments. Our results demonstrate that the Proteomic Reactor facilitates the analysis of ubiquitinated proteins and may be used for the study of other post-translational modifications. In addition, the Proteomic Reactor has the potential to aid future biological studies requiring MS analysis of affinity-purified proteins or protein complexes.

## 3.4 Materials and methods

Antibodies and Plasmids. Mouse monoclonal anti-Ub antibody (clone FK2) was obtained from Biomol. Mouse $IgG_{1k}$ immunoglobulin was obtained from Sigma-Aldrich. Mouse monoclonal anti-VCP, anti-Sequestosome 1, anti-Actin, and anti-HSP90 antibodies were obtained from Abcam, BD Biosciences, Chemicon, and Upstate, respectively. Goat antimouse HRP-conjugated secondary antibody was obtained from DakoCytomation. Anti-VCP siRNA constructs were generated as previously described[21] against VCP sequences 5'-AGATGATGTGGACCTGGAA-3' (VCP3) and 5'-AAGTAGGGTATGATGACAT-3' (VCP6).

Cell Culture and siRNA Treatment. H1299 lung adenocarcinoma cells were cultured at 37°C in 5% $CO_2$ in DMEM medium supplemented with 10% FBS and 1% penicillinstreptomycin (Cellgro, Mediatech). After reaching 50% confluence, cells were transfected with 5 μg of anti-VCP siRNA constructs or empty vector (pSuper), along with 0.5 μg of a plasmid conferring puromycin resistance (pBabe-Puro), using the FuGENE 6 transfection reagent (Roche). A total of 48 h after transfection, cells were split into media containing 3 μg/mL of puromycin (BD Bioscience). Five days later, cells were harvested and lysed in modified RIPA buffer (50 mM Tris-HCl pH 7.4, 150 mM NaCl, 1% NP-40, 0.25% Na deoxycholate, and 1 mM EDTA) containing leupeptin (10 μg/mL), pepstatin (10 μg/mL), aprotinin (2 μg/mL), PMSF (100 μg/mL), MG115 (10 μM), and NEM (1 μM). Whole cell lysates were centrifuged for 10 min at 14 000 rpm to pellet cell debris. Protein concentrations were determined using a Bradford protein assay kit (Bio-Rad).

Immunopurification of Ubiquitinated Proteins and Acetone Precipitation. Polyprep chromatography columns (Bio-Rad) were filled with 250 μL of anti-Ub beads, prepared as previously described.[15] Pre-cleared H1299 WCLs were then loaded onto immunopurification columns and allowed to incubate for 4 h at 4 °C. Columns were washed with modified RIPA buffer, and bound proteins were eluted with 2% SDS. Trichloroacetic acid was added to immunopurified samples up to a final concentration of 25%, and samples were put on ice for 30 min with periodic vortexing. Samples were then centrifuged for 30 min at 14 000 rpm. Supernatants were removed, and protein pellets were washed with chilled (-20 °C) acetone and centrifuged for another 20 min at 14 000 rpm. Acetone was removed, and protein pellets were reconstituted in 1% NP-40, 150 mM NaCl, and 10 mM potassium phosphate (KPB buffer). Formic acid (FA) up to 10% was added to help dissolve the pellets. Western blotting and immunoprecipitations have been described previously.[15]

Proteomic Reactor Assembly and Protein Processing. The Proteomic Reactor was assembled by connecting 8 cm of fused silica capillary tubing (200 μm i.d.) to an in-line microfilter (Upchurch Scientific). A pressure bomb was then used to pack 4 cm of strong cation exchange (SCX) beads (The Nest Group) into the opposite end of the capillary tubing by applying 150 psi of $N_2$ gas. SCX resin was then washed with KPB buffer. Immunopurified samples were acidified, and 2 μg of sequencing-grade trypsin (Promega) was added (0.5 μg/μL in KPB buffer, pH 3.0). Protein samples were centrifuged at 14 000 rpm for 5 min to pellet insoluble material. The supernatant was loaded on the SCX beads using the pressure bomb and $N_2$ gas. Bound proteins were then washed with 20% acetonitrile (ACN), followed by milli-Q $H_2O$. Proteins were subjected to disulfide bond reduction with 100 mM dithiothreitol (in 10 mM

ammonium bicarbonate) for 30 min. The SCX resin was then dried and washed with KPB buffer to quench the reaction. Proteins were subjected to simultaneous alkylation and trypsin digestion with 10 mM iodoacetamide (in 100 mM Tris-HCl, pH 8.0) for 2 h. Peptides were then eluted with 200 mM ammonium bicarbonate.

LC-MS/MS and Data Analysis. Peptide samples were acidified with FA and loaded on a 200 μm × 50 mm fused silica precolumn packed with 5 cm of YMC ODS-A C18 beads (Waters Co.) using an 1100 series HPLC system (Agilent Technologies). Following a desalting step, the flow was split, and peptides were eluted through a second 75 μm × 50 mm column packed with the same beads at approximately 200 nL/min using a 5-80% gradient of ACN with 0.1% FA (J.T. Baker) for 1 h. The LC effluent was electrosprayed into a QSTAR Pulsar quadrupole-TOF mass spectrometer (ABI/MDS Sciex). MS and MS/MS spectra were acquired in a data-dependent mode, with one MS scan followed by four MS/MS scans. MS/MS spectra were matched to human sequences in the NCBI database using the Mascot (Matrix Science) search engine with carbamidomethyl (C) as a fixed modification and oxidation (M) and ubiquitination (K) as variable modifications. Peptide and MS/MS mass tolerances were set at ±100 ppm and 0.2 Da, respectively. Ion cutoff scores were set to 15. Reverse database searches using the same criteria were also performed to determine false positive rates. MS/MS spectra of all signature peptides were manually verified to confirm internal lysine residues with GG or LRGG moieties.

## 3.5 Results and discussion

Strategy for Processing Proteins on the Proteomic Reactor. The overall strategy for processing proteins on the Proteomic Reactor is a three-step procedure (Figure 3.1). In the first step, protein samples (~100 μL) are acidified, mixed with trypsin, and loaded on the Proteomic Reactor using a pressure bomb and nitrogen gas. At pH 3, most of the proteins are positively charged and bind to the SCX resin. Nonionic detergents, salts, and other buffer components are removed by washing the beads with a dilute organic solvent and water. Since the interstitial volume in the Proteomic Reactor is approximately 50 nL, a 2000-fold pre-concentration effect is achieved when 100 μL of protein sample is loaded.[14] Therefore, chemical derivatizations and enzymatic reaction rates are significantly enhanced as compared to in-gel and in-solution digestions.

In the second step, bound proteins are subjected to disulfide bond reduction with dithiothreitol. A buffer solution containing iodoacetamide is then added to increase the pH and permit simultaneous cysteine alkylation and trypsin digestion. Only sufficient volumes of dithiothreitol and iodoacetamide solutions are added to hydrate the SCX beads. In the third and final step of the procedure, peptides are eluted using an ammonium bicarbonate buffer, which is readily compatible with LC-MS/MS analysis. The total time for processing protein samples on the Proteomic Reactor is approximately 2.5 h, which is shorter than standard in-gel or in-solution digestion protocols.

Immunopurification of Ubiquitinated Proteins from H1299 Cells -/+ VCP Depletion. To demonstrate a biological application of the Proteomic Reactor, we integrated this technology with a protocol that we have previously developed for purifying and identifying ubiquitinated

proteins in human cells.[15] We believed that by eliminating SDS-PAGE separation and in-gel digestions after immunopurification, a significant reduction in sample processing times could be achieved (Figure 3.1).

Two populations of human H1299 lung adenocarcinoma cells were used in this study. The first consisted of cells transfected with anti-VCP siRNA, while the second consisted of control cells transfected with empty vector (pSuper). As depicted in Figure 3.2, knock-down of VCP by two different anti-VCP siRNA constructs (VCP3 and VCP6) resulted in an accumulation of ubiquitinated proteins as assessed by Western blotting with an anti-ubiquitin (Ub) antibody. Cells transfected with anti-VCP3 siRNA were used for large-scale immunopurification experiments because slightly higher levels of knockdown and ubiquitin accumulation were observed with this construct.

Poly-prep chromatography columns were filled with anti-Ub beads, and H1299 whole cell lysates (WCLs) were loaded and allowed to incubate at 4 °C. Prior to loading, WCLs were pre-cleared by passing them through columns containing only Protein G-agarose beads. After a 4 h incubation period, which was a sufficient time to permit ubiquitinated proteins to bind the anti-Ub beads with high-affinity, a series of washing steps was performed. Bound proteins were then eluted with a solution containing a low percentage of ionic detergent (2% SDS). A control experiment using anti-IgG beads was also performed in parallel.

Aliquots of VCP3 and pSuper WCLs before and after immunopurification were analyzed by Western blotting (Figure 3.2). Prior to immunopurification, numerous ubiquitinated proteins were observed in both cell populations (lanes 1 and 4). The stronger band intensities in lane 4 confirmed the previous observation of increased ubiquitination as a result of anti-VCP3 siRNA

transfection. After immunopurification, WCLs were depleted of ubiquitinated proteins (lanes 2 and 5), indicating that they were captured and retained by the anti-Ub beads. Aliquots containing 0.2% of the immunopurified samples were also analyzed to confirm the efficiency of elution (lanes 3 and 6). As expected, no ubiquitinated proteins were detected when anti-IgG beads were used (lane 7), thus confirming the specificity of the anti-Ub beads.

Loading Affinity-Purified Proteins on the Proteomic Reactor. When dissolved in 2% SDS, purified protein samples did not bind to the SCX beads in the Proteomic Reactor. This was determined by analyzing reactor flow-throughs after SDS-PAGE and silver staining. Therefore, we explored the use of acetone precipitation and spin columns with a low molecular weight cutoff (MWCO) to permit a buffer exchange step. We tested both procedures using bovine serum albumin (BSA) dissolved in 2% SDS. Acetone precipitation resulted in a BSA pellet that was dissolved in a reactor-compatible solution, while spin columns concentrated BSA in a small enough volume to be diluted in the same solution. Only acetone-precipitated samples displayed efficient binding to SCX beads. Therefore, this procedure was selected as the method of choice for removing SDS from immunopurified samples.

Interestingly, we found that ubiquitinated protein pellets did not dissolve as easily as the BSA protein pellet. To aid the solubilization process, formic acid up to 10% was added. This step was fully compatible with the Proteomic Reactor as protein samples require an acidification step prior to loading.[14] After samples were loaded on the Proteomic Reactor, bound proteins were subjected to reduction, alkylation, and trypsin digestion. The resulting peptides were then eluted for LC-MS/MS analysis.

Mass Spectrometric Analysis of Immunopurified Proteins. Tryptic peptides were separated by nanoflow LC and introduced into a quadrupole time-of-flight (q-TOF) mass spectrometer.[22,23] Mapping of ubiquitination sites was performed by identifying internal lysine (K) residues with monoisotopic masses of 242.13788 or 511.32305 Da. These masses correspond to lysine residues (128.09496 Da) with attached GG (+114.04292 Da) or LRGG (+383.22809 Da) moieties derived from the carboxy terminus of Ub.[15,24] Signature peptides containing modified lysine residues were produced by trypsin digestion.

Representative MS/MS spectra of signature peptides are shown in Figure 3.3. Doubly charged $[M + 2H]^+$ peptide ions at m/z 543.27, 475.72, 436.72, and 415.24 were fragmented via collision-induced dissociation. The masses of these fragments were measured, and the peptide sequences $LK_{GG}DLVQQK$, $FGSLEK_{GG}R$, $ALAEAK_{GG}R$, and $ALLK_{GG}DR$, all of which contain an internal lysine with a diglycine remnant ($K_{GG}$), were obtained. Database searching matched these peptides to a Kinesin-like motor protein, KIF-14, MTG8-related protein, and Neu tumor-associated kinase.

A total of 16 signature peptides matching to 14 proteins were identified in the VCP3 sample, while 12 signature peptides matching to 11 proteins were identified in the pSuper sample (see Table 3.1). Reverse database searching using the same criteria for protein identifications were performed to determine false positive rates. A single signature peptide was identified with the reverse database that had a higher score in the forward database for the VCP3 MS/MS data, which represents a false positive rate of approximately 6% (1/16), while three signature peptides were identified with the reverse database that had higher scores in the forward database for the pSuper MS/MS data, which represents a false positive rate of 25% (3/12).

A total of 58 candidate ubiquitinated proteins were also identified from the VCP3 and pSuper samples. These proteins were selectively purified using the anti-Ub antibody and were not identified after immunopurification with the IgG antibody. Candidate ubiquitinated proteins that were identified include the following: Sequestosome-1, which is a protein that contains a ubiquitin-associated (UBA) domain and is involved in proteasomal degradation of ubiquitinated proteins;[25] numerous E2 and E3 ubiquitin ligases; subunits of the 26S proteasome complex; Actin, which is known to be ubiquitinated on lysine-118;[26] and Rac1, which is a regulator of actin dynamics that was recently shown to be ubiquitinated.[27]

These proteins were designated as candidates because they were isolated by immunopurification with the anti-Ub antibody, but no ubiquitination sites could be precisely mapped by MS/MS. The inability to map ubiquitination sites on these proteins may be attributed to sensitivity issues, but it is also possible that for at least a subset of these proteins, they, or a large percentage of their total pool, were not ubiquitinated but instead interacted strongly with various ubiquitinated substrates (especially in the case of VCP depletion) and were copurified.

The distribution of ubiquitinated and candidate ubiquitinated proteins identified from the VCP3 and pSuper samples is shown in Figure 3.4. Five ubiquitination sites were mapped on four proteins that were common to both samples (4/21 or 19%), while 12 ubiquitination sites were exclusively mapped on 10 proteins that were identified from the VCP3 sample. A greater degree of overlap was observed for the candidate ubiquitinated proteins, as 23 were common to samples (23/58 or 40%), while 21 were exclusively identified from the VCP3 sample. Pie-chart distributions of all proteins into functional categories are also shown in Figure 3.4.

Confirmation of Candidate Ubiquitinated Proteins. Three candidate ubiquitinated proteins were selected for immunoprecipitation and Western blotting experiments to confirm their ubiquitination status in H1299 cells -/+ VCP depletion. Sequestosome-1 (also known as p62 or p60) and HSP90 were selected because their higher abundance in VCP depleted cells versus the control cells (based on the number of identified peptides) could be due to increases in their ubiquitination or to interactions with ubiquitinated proteins. Actin, on the other hand, was selected because was no apparent difference in the abundance of this protein in control versus VCP depleted cells was observed.

Protein extracts from H1299 cells transfected with empty vector (pSuper) or anti-VCP3 siRNA (VCP3) were subjected to immunoprecipitation with anti-Sequestosome-1, anti-HSP90, and anti-Actin beads and were analyzed by Western blotting with the anti-Ub antibody (Figure 3.5). For Sequestosome-1, numerous bands above 62 kDa were detected in both samples, confirming the presence of higher molecular weight polyubiquitinated isoforms (lanes 1 and 2). Stronger band intensities in lane 2 were also observed, confirming an increase in the amount of ubiquitinated Sequestosome-1 as a result of VCP knockdown (lane 2).

For Actin (Figure 3.5), a single band at approximately 50 kDa was detected in both samples, which confirmed the presence of a mono-ubiquitinated isoform of actin (lanes 1 and 2). Band intensities were similar for both samples, suggesting that VCP knockdown had no effect on the ubiquitination status of actin. For HSP90 (Figure 3.5), bands above 90 kDa were detected in both samples, which confirmed the presence of higher molecular weight poly-ubiquitinated isoforms (lanes 1 and 2). Stronger band intensities in lane 2 were also observed, confirming an increase in the amount of ubiquitinated HSP90 as a result of VCP knockdown (lane 2).

Preliminary Data Analysis. Although the goal of this study was to demonstrate an application of the Proteomic Reactor for identifying affinity-purified proteins, it is worth discussing briefly the potential significance of the data. Many proteins that were identified in VCP depleted cells are consistent with an accumulation of ubiquitinated proteins when VCP expression/function is suppressed. These proteins include various subunits of the 26S proteasome complex, known ubiquitin binding partners, and members of the heat shock family of proteins. Interestingly, studies have shown that several other proteins, including Sequestosome-1, may be substrates of the ubiquitination machinery,[28,29] and our results indicate that ubiquitinated forms of these proteins accumulate in cells containing low VCP levels. It also deserves to be noted that our approach has mapped many novel ubiquitination sites on proteins that were purified from cells containing and/or expressing reduced amounts of VCP. Some of these proteins have been documented to be regulated by the ubiquitin-proteasome pathway,[30,31] while others have not. Thus, it will be interesting to determine the effect of ubiquitination on these proteins and if mutations of lysine residues harboring this modification impair or suppress their function.

## 3.6 Conclusion

Taken together, our results demonstrate that the Proteomic Reactor is a microfluidic processing device that permits the analysis of affinity-purified proteins by LC-MS/MS. A specific application was demonstrated that enabled the identification of ubiquitinated and candidate ubiquitinated proteins in human H1299 lung adenocarcinoma cells. The Proteomic Reactor may be adapted for the study of other posttranslational modifications and may be employed for future studies requiring an analysis of affinity-purified proteins.

**Figure 3.1 Experimental method and the Proteomic Reactor.** (a) 1: H1299 cells were transfected with empty vector (pSuper) or anti-VCP siRNA. 2: Ubiquitinated proteins were isolated by immunopurification. 3: Immunopurified samples were processed on the Proteomic Reactor. 4: Peptides were analyzed by LC-MS/MS. (b) 1: Immunopurified samples were acidified (pH < 3), mixed with trypsin, and loaded on the Proteomic Reactor. 2: Proteins were reduced, and the pH was raised to permit alkylation and enzymatic digestion (pH = 8). 3: Peptides were eluted in a buffer compatible with LC-MS/MS analysis.

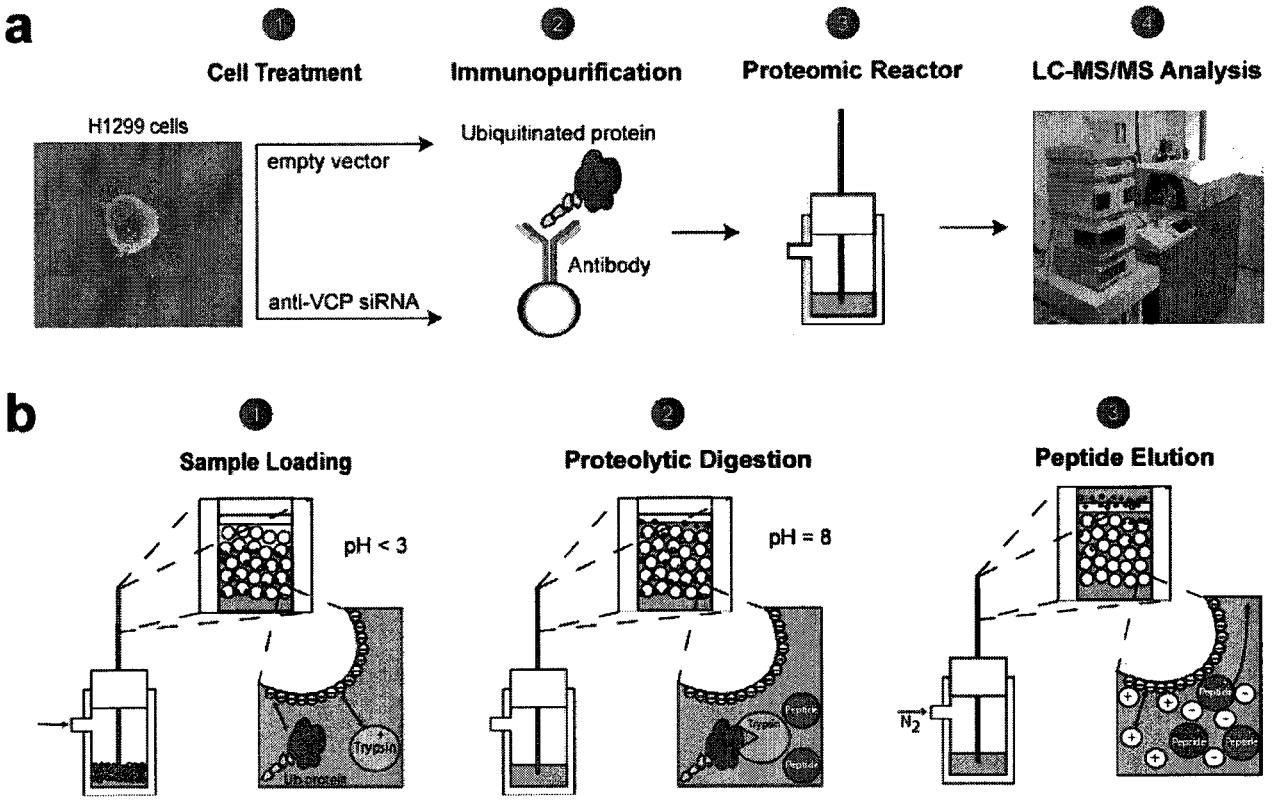


**Figure 3.2 Anti-VCP siRNA constructs and immunopurification of ubiquitinated proteins.** (a) Protein extracts (20 μg) from H1299 cells transfected with empty vector (pSuper) or anti-VCP siRNA constructs (VCP3 and VCP6) were analyzed by Western blotting with anti-VCP, anti-Ub, and anti-Actin (as a loading control) antibodies. (b) Protein extracts (100 μg) from H1299 cells were analyzed by Western blotting with the anti-Ub antibody before and after immunopurification (WCL and WCL post-IP). Aliquots containing 0.2% of the immunopurified samples were also analyzed (Eluate). A negative control was performed using an anti-IgG antibody for immunopurification.

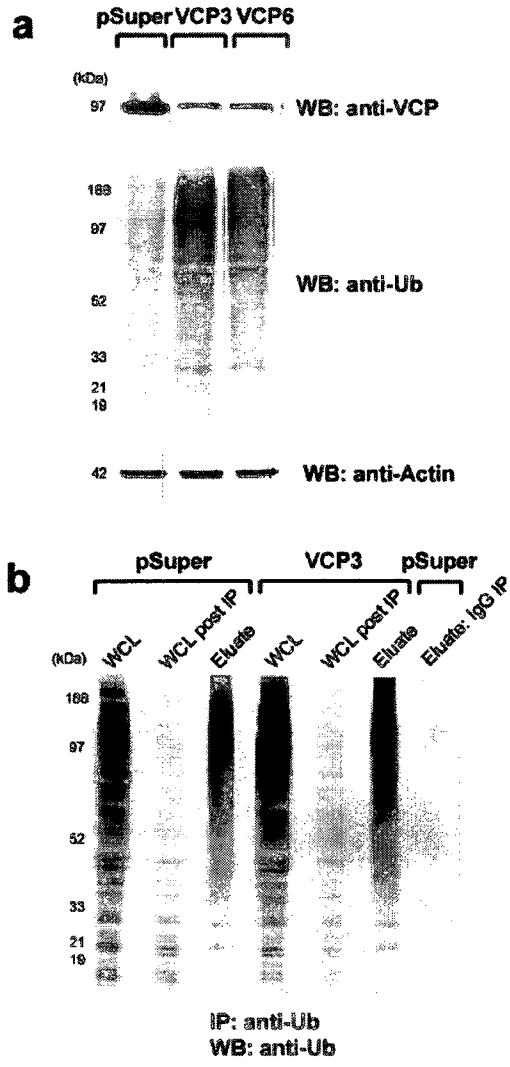


**Figure 3.3 Representative MS/MS spectra of signature peptides.** MS/MS spectra of peptide ions at m/z (a) 543.27 (2+), (b) 475.72 (2+), (c) 436.72 (2+), and (d) 415.24 (2+) matching to Kinesin-like motor protein, KIF14, MTG8-related protein, and Neu tumor-associated kinase, respectively. The peptide sequences $LK_{GG}DLVQQK$ (ion score 21), $FGSLEK_{GG}R$ (ion score 17), $ALAEAK_{GG}R$ (ion score 28), and $ALLK_{GG}DR$ (ion score 15) all contain an internal lysine (K) residue with a diglycine (GG) remnant that was derived from Ub.

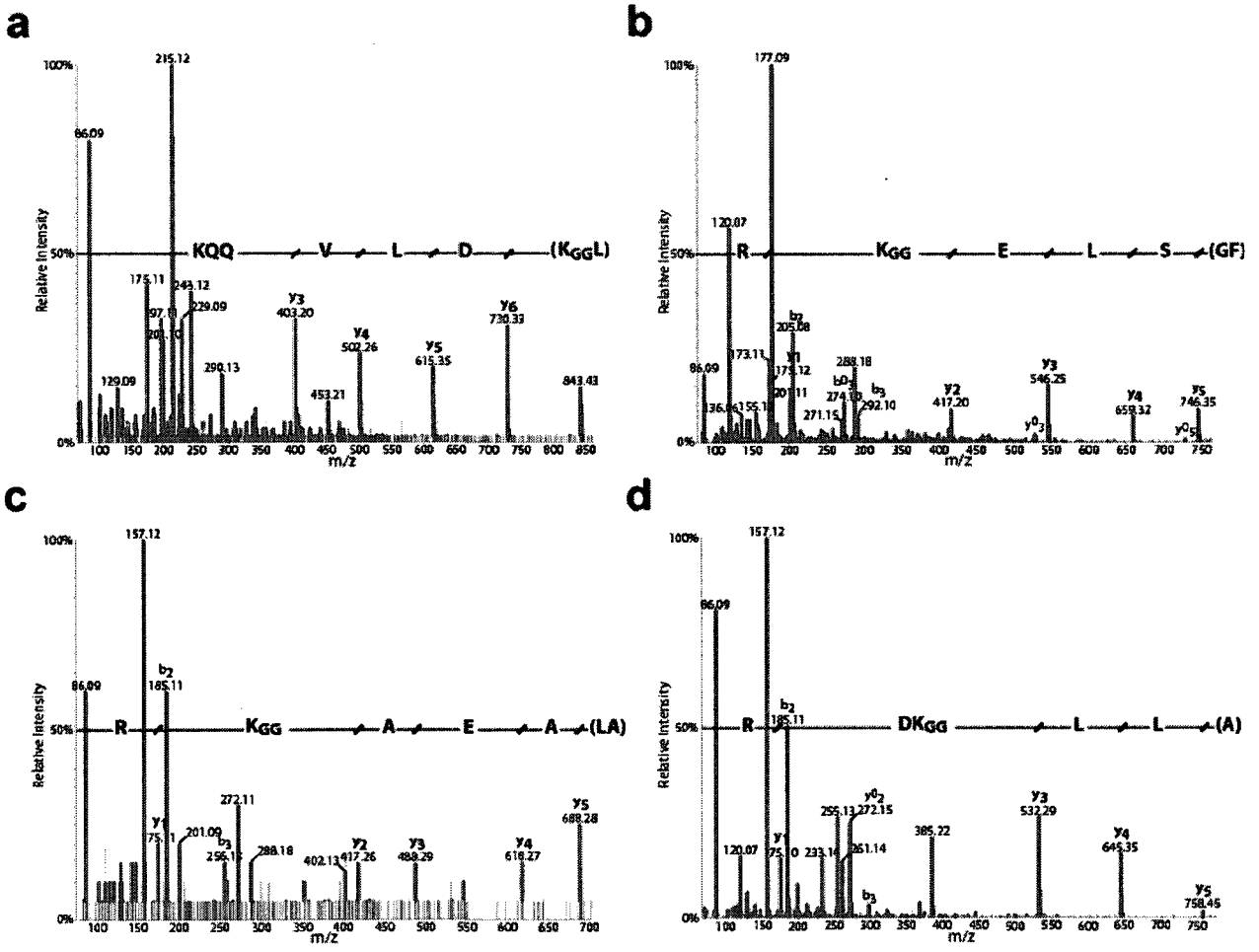


**Figure 3.4 Ubiquitinated proteins and candidate ubiquitinated proteins.** (a and c) Venn diagrams of ubiquitinated proteins and candidate ubiquitinated proteins identified in two populations of H1299 cells. (b and d) Pie-chart distributions of ubiquitinated proteins and candidate ubiquitinated proteins into functional categories.

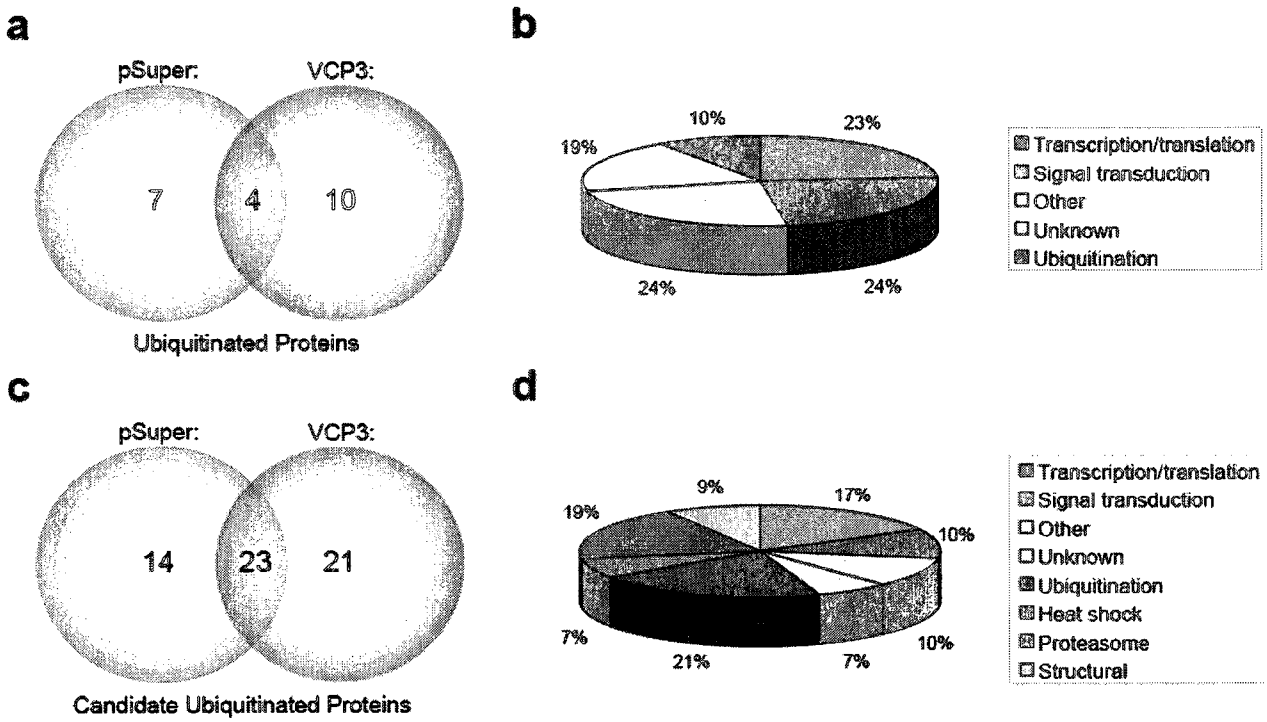
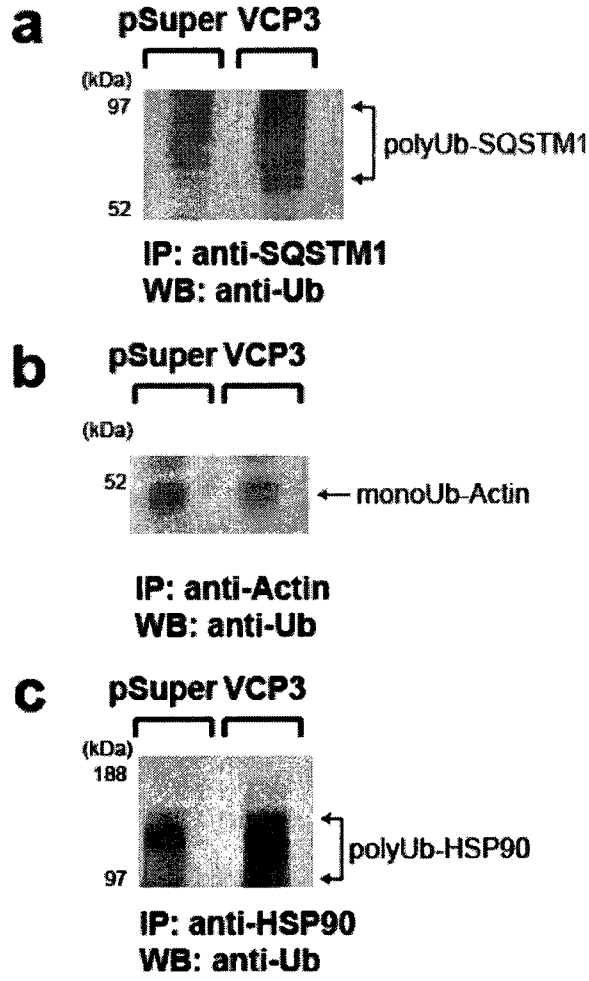


**Figure 3.5 Confirmation of candidate ubiquitinated proteins.** Protein extracts (500 μg) from H1299 cells transfected with empty vector (pSuper) or anti-VCP3 siRNA (VCP3) were subjected to immunoprecipitation with (a) anti-Sequestosome 1, (b) anti-Actin, or (c) anti-HSP90 antibodies and were analyzed by Western blotting with the anti-Ub antibody.



## Table 3.1 Ubiquitinated Proteins and Signature Peptides Identified in Human H1299 Lung Adenocarcinoma Cells.

| protein name[a] | accession no. (gi) | function[b] | sample[c] | signature peptide | ion score |
|---|---|---|---|---|---|
| Ubiquitin | 229532 | ubiquitination | pSuper, VCP3 | LIFAGK$_{GG}$QLEDGR | 58, 51 |
| | | | pSuper, VCP3 | LIFAGK$_{LRGG}$QLED GR | 19, 19 |
| RP11-334J6.1 | 57162363 | unknown | pSuper, VCP3 | K$_{GG}$AEKEAK | 22, 20 |
| kinesin-like motor protein | 58036802 | transport (other) | pSuper, VCP3 | LK$_{GG}$DLVQQK | 21, 18 |
| FLJ32745 | 29791660 | unknown | pSuper, VCP3 | K$_{GG}$IDIQAKR | 21, 17 |
| fibroblast growth factor receptor 4 variant | 62088842 | signal transduction | pSuper | IGTVTPSK$_{GG}$VSR | 20 |
| Nudix-type motif 6 isoform a variant | 62089102 | hydrolase (other) | pSuper | LDAAAFQK$_{GG}$GLQ GK | 20 |
| TRAF3-interacting JNK-activating modulator | 4678741 | signal transduction | pSuper | K$_{GG}$VLLEMEDQKN SYEQK | 20 |
| Homeobox protein Hox-B7 | 123255 | transcription | pSuper | RMK$_{GG}$WK$_{GG}$KENK | 18 |
| kinesin-like protein KIF14 | 40788943 | motor protein (other) | pSuper | FGSLEK$_{GG}$R | 17 |
| GUF1 | 54035062 | signal transduction | pSuper | NNK$_{LRGG}$QVLDK | 17 |
| zinc finger protein HRX | 4007383 | transcription | pSuper | DSK$_{GG}$SIEK$_{LRGG}$K $_{GG}$R | 15 |
| Polyubiquitin | 2627129 | ubiquitination | VCP3 | LIFAGK$_{GG}$QLEDGR | 51 |
| | | | VCP3 | LIFAGK$_{LRGG}$QLED GR | 19 |
| MTG8-related protein | 2981446 | transcription | VCP3 | ALAEAK$_{GG}$R | 28 |
| ATPase family AAA domain-containing protein 1 | 18033709 | ATPase (other) | VCP3 | WMVDAIDPTRK$_{GG}$ QK | 21 |
| threonyl-tRNA synthetase | 1464742 | translation | VCP3 | QAFERLEVK$_{GG}$K | 19 |
| ASH1L | 7243221 | transcription | VCP3 | KIVASSAMGLVNK$_{GG}$DIGK$_{GG}$K | 18 |
| LOC143684 | 46195787 | unknown | VCP3 | LESK$_{GG}$PSNGDRSM | 18 |
| ATR-interacting protein | 17227176 | ribonuclease (other) | VCP3 | LLDGMSK$_{GG}$NPSG K | 17 |
| v-fos FBJ osteosarcoma oncogene homologue | 4885241 | unknown | VCP3 | AGVVK$_{GG}$TMTGGRAQSIGRR | 16 |
| Neu-tumor associated kinase | 7657214 | signal transduction | VCP3 | ALLK$_{GG}$DR | 15 |
| 5-hydroxytryptamine 1A receptor | 35524 | signal transduction | VCP3 | K$_{GG}$MALAR | 15 |

[a,b]Commonly used protein names and functional information obtained from SwissProt (http://ca.expasy.org) and the Human Protein Reference Database (http://www.hprd.org) [c] pSuper refers to H1299 cells transfected with an empty vector; VCP3 refers to H1299 cells transfected with anti-VCP siRNA.

## 3.7 References

(1) Vasilescu, J.; Figeys, D. Mapping protein-protein interactions by mass spectrometry. Curr. Opin. Biotechnol. 2006, 17, 394-399.

(2) Vasilescu, J.; Guo, X.; Kast, J. Identification of protein-protein interactions using in vivo cross-linking and mass spectrometry. Proteomics 2004, 4, 3845-3854.

(3) Lewandrowski, U.; Moebius, J.; Walter, U.; Sickmann, A. Elucidation of N-glycosylation sites on human platelet proteins: a glycoproteomic approach. Mol. Cell. Proteomics 2006, 5, 226-233.

(4) Ficarro, S. B.; McCleland, M. L.; Stukenberg, P. T.; Burke, D. J.; Ross, M. M.; Shabanowitz, J.; Hunt, D. F.; White, F. M. Phosphoproteome analysis by mass spectrometry and its application to Saccharomyces cerevisiae. Nat. Biotechnol. 2002, 20, 301-305.

(5) Rush, J.; Moritz, A.; Lee, K. A.; Guo, A.; Goss, V. L.; Spek, E. J.; Zhang, H.; Zha, X. M.; Polakiewicz, R. D.; Comb, M. J. Immunoaffinity profiling of tyrosine phosphorylation in cancer cells. Nat. Biotechnol. 2005, 23, 94-101.

(6) Peng, J.; Schwartz, D.; Elias, J. E.; Thoreen, C. C.; Cheng, D.; Marsischky, G.; Roelofs, J.; Finley, D.; Gygi, S. P. A proteomics approach to understanding protein ubiquitination. Nat. Biotechnol. 2003, 21, 921-926.

(7) Bauer, A.; Kuster, B. Affinity purification-mass spectrometry; powerful tools for the characterization of protein complexes. Eur. J. Biochem. 2003, 270, 570-578.

(8) Lecchi, P.; Gupte, A. R.; Perez, R. E.; Stockert, L. V.; Abramson, F. P. Size-exclusion chromatography in multi-dimensional separation schemes for proteome analysis. J. Biochem. Biophys. Methods 2003, 56, 141-152.

(9) Poggel, M.; Melin, T. Free-flow zone electrophoresis: a novel approach and scale-up for preparative protein separation. Electrophoresis 2001, 22, 1008-1015.

(10) Quadroni, M.; James, P. Proteomics and automation. Electrophoresis 1999, 20, 664-677.

(11) Doucette, A.; Craft, D.; Li, L. Protein concentration and enzyme digestion on microbeads for MALDI-TOF peptides mass mapping of proteins from dilute solutions. Anal. Chem. 2000, 72, 3355-3362.

(12) Craft, D.; Doucette, A.; Li, L. Microcolumn capture and digestion of proteins combined with mass spectrometry for protein identification. J. Proteome Res. 2002, 1, 537-547.

(13) Craft, D.; Li, L. Integrated sample processing system involving on-column protein adsorption, sample washing, and enzyme digestion for protein identification by LC-ESI MS/MS. Anal. Chem. 2005, 77, 2649-2655.

(14) Ethier, M.; Hou, W.; Duewel, H.; Figeys, D. The Proteomic Reactor: a microfluidic device for processing minute amounts of protein prior to mass spectrometry analysis. J. Proteome Res. 2006, 5, 2754-2759.

(15) Vasilescu, J.; Smith, J. C.; Ethier, M.; Figeys, D. Proteomic analysis of ubiquitinated proteins from human MCF-7 breast cancer cells by immunoaffinity purification and mass spectrometry. J. Proteome Res. 2005, 4, 2192-2200.

(16) Ye, Y. Diverse functions with a common regulator: ubiquitin takes command of an AAA ATPase. J. Struct. Biol. 2006, 156, 29-40.

(17) Wang, Q.; Song, C.; Li, C. H. Molecular perspectives on p97- VCP: progress in understanding its structure and diverse biological functions. J Struct Biol. 2004, 146, 44-57.

(18) Elsasser, S.; Finley, D. Delivery of ubiquitinated substrates to protein-unfolding machines. Nat. Cell Biol. 2005, 7, 742-749.

(19) Kobayashi, T.; Tanaka, K.; Inoue, K.; Kakizuka, A. Functional ATPase activity of p97/valosin-containing protein is required for the quality control of endoplasmic reticulum in neuronally differentiated mammalian PC12 cells. J. Biol. Chem. 2002, 277, 47358-47365.

(20) Wojcik, C.; Yano, M.; DeMartino, G. N. RNA interference of valosin-containing protein (VCP/p97) reveals multiple cellular roles linked to ubiquitin/proteasome-dependent proteolysis. J. Cell Sci. 2004, 117, 281-292.

(21) Brummelkamp, T. R.; Bernards, R.; Agami, R. A system for stable expression of short interfering RNAs in mammalian cells. Science 2002, 296, 550-553.

(22) Pinto, D. M.; Ning, Y.; Figeys, D. An enhanced microfluidic chip coupled to an electrospray Qstar mass spectrometer for protein identification. Electrophoresis 2000, 21, 181-190.

(23) Steen, H.; Kuster, B.; Mann, M. Quadrupole time-of-flight versus triple-quadrupole mass spectrometry for the determination of phosphopeptides by precursor ion scanning. J. Mass Spectrom. 2001, 36, 782-790.

(24) Warren, M. R.; Parker, C. E.; Mocanu, V.; Klapper, D.; Borchers, C. H. Electrospray ionization tandem mass spectrometry of model peptides reveals diagnostic fragment ions for protein ubiquitination. Rapid Commun. Mass Spectrom. 2005, 19, 429-437.

(25) Seibenhener, M. L.; Babu, J. R.; Geetha, T.; Wong, H. C.; Krishna, N. R.; Wooten, M. W. Sequestosome 1/p62 is a polyubiquitin chain binding protein involved in ubiquitin proteasome degradation. Mol. Cell. Biol. 2004, 24, 8055-8068.

(26) Burgess, S.; Walker, M.; Knight, P. J.; Sparrow, J.; Schmitz, S.; Offer, G.; Bullard, B.; Leonard, K.; Holt, J.; Trinick, J. Structural studies of arthrin: monoubiquitinated actin. Mol. Biol. 2004, 341, 1161-1173.

(27) Lynch, E. A.; Stall, J.; Schmidt, G.; Chavrier, P.; D'Souza-Schorey, C. Proteasome-mediated degradation of Rac1-GTP during epithelial cell scattering. Mol. Biol. Cell. 2006, 17, 2236-2242.

(28) Nakaso, K.; Yoshimoto, Y.; Nakano, T.; Takeshima, T.; Fukuhara, Y.; Yasui, K.; Araga, S.; Yanagawa, T.; Ishii, T.; Nakashima, K. Transcriptional activation of p62/A170/ZIP during the formation of the aggregates: possible mechanisms and the role in Lewy body formation in Parkinson's disease. Brain Res. 2004, 1012, 42-51.

(29) Blank, M.; Mandel, M.; Keisari, Y.; Meruelo, D.; Lavie, G. Enhanced ubiquitinylation of heat shock protein 90 as a potential mechanism for mitotic cell death in cancer cells induced with hypericin. Cancer Res. 2003, 63, 8241-8247.

(30) Eide, E. J.; Woolf, M. F.; Kang, H.; Woolf, P.; Hurst, W.; Camacho, F.; Vielhaber, E. L.; Giovanni, A.; Virshup, D. M. Control of mammalian circadian rhythm by CKIepsilon-regulated proteasome-mediated PER2 degradation. Mol. Cell. Biol. 2005, 25, 2795-2807.

(31) Hermida-Matsumoto, M. L.; Chock, P. B.; Curran, T.; Yang, D. C. Ubiquitinylation of transcription factors c-Jun and c-Fos using reconstituted ubiquitinylating enzymes. J Biol. Chem. 1996, 271, 4930-4936.

# Chapter 4: Systematic determination of ion score cutoffs based on calculated false positive rates: application for identifying ubiquitinated proteins by tandem mass spectrometry

## 4.1 Manuscript status and statement of author contributions

A version of this chapter has been published in the Journal of Mass Spectrometry: *Vasilescu J, Smith JC, Zweitzig D, Denis NJ, Haines DS, Figeys D. Systematic determination of ion score cutoffs based on calculated false positive rates: Application for identifying ubiquitinated proteins by tandem mass spectrometry. Journal of Mass Spectrometry. Vol. 43: 296-304, 2008.* JV and JCS conceived, designed and performed the experiments. JV wrote the manuscript. DZ performed the cell culture and siRNA experiments. NJD processed samples on the proteomic reactor. DSH and DF edited the manuscript.

## 4.2 Summary

We report a simple approach for determining ion score cutoffs that permit the confident identification of ubiquitinated proteins by tandem mass spectrometry (MS/MS). Initial experiments involving the analysis of gel bands containing multi-Ubiquitin chains with quadrupole time-of-flight and quadrupole ion trap mass spectrometers revealed that standard ion score cutoffs used for database searching were not sufficiently stringent. We also found that false positive and false negative rates (FPR and FNR) varied significantly depending on the cutoff scores used and that appropriate cutoffs could only be determined following a systematic

evaluation of false positive rates. When standard cutoff scores were used for the analysis of complex mixtures of ubiquitinated proteins, unacceptably high FPR were observed. Finally, we found that FPR for ubiquitinated proteins are affected by the size of the protein database that is searched. These observations may be applicable for the study of other post-translational modifications.

## 4.3 Introduction

The majority of proteomic studies that have been conducted over the last decade have relied on tandem mass spectrometry (MS/MS) to identify proteins and their post-translational modifications in a wide variety of samples including tissue extracts, biological fluids, cell lysates and subcellular fractions. Consequently, our understanding of signaling pathways and the interactions that regulate key cellular processes have improved significantly. With an increasing number of laboratories relying on proteomic experiments to guide their research efforts and an escalating amount of data in the literature, straightforward methods to estimate the accuracy of MS/MS results are necessary.[1–9] This is especially true for mapping post-translational modifications, including phosphorylation and ubiquitination, since the standard criteria used for identifying modified proteins may not suitable.[10–16]

Prior to MS/MS analysis, protein samples are typically subjected to an enzymatic digestion step to generate shorter length peptides. The peptides are then introduced into a mass spectrometer by electrospray ionization (ESI) or matrix-assisted laser desorption ionization (MALDI) so their mass to charge ratios (m/z) and fragmentation patterns may be determined.

The resulting MS/MS data are searched against a protein database using a software algorithm, such as Mascot or Sequest.[17,18] These algorithms function by comparing the MS/MS fragmentation patterns of the unknown peptides to the predicted fragmentation patterns of all peptides of the same mass in a protein database and reporting the best matches.

An important consequence of this database searching strategy is that incorrect matches will occur since the best match is not always a perfect match. To differentiate between correct and incorrect matches, algorithms assign ion scores which reflect the likelihood that a peptide match occurred by chance. An ion score cutoff is then employed so that peptides that match with scores below this value are rejected, while those with higher scores are accepted. This means that for any given cutoff, some correct matches will be rejected (false negatives) while a number of incorrect matches (false positives) will be accepted. One method for estimating false positive rates (FPR) involves searching a MS/MS dataset against a database that also contains non-sense sequences.[19,20] These non-sense sequences may be generated by randomly scrambling or reversing the protein sequences in the original (forward only) database. While correct matches will be made to the forward protein sequences, matches that occur randomly will be made with equal frequency to the forward and scrambled/reversed protein sequences.

Numerous factors can influence FPR, including the mass accuracy and resolution of the mass spectrometer, the size of the protein database that is searched, and the nature of the sample.[4,6,21] Therefore, if a single ion score cutoff is used for every MS/MS experiment, a variety of FPR may be observed. For example, if a single yeast protein in a gel band is digested and analysed, most of the resulting peptides should be matched with scores well-above a standard cutoff and the FPR should be negligible. However, if a protein extract from whole cell lysates

(WCLs) is analysed using the same cutoff, hundreds or thousands of peptide matches could be accepted at, or just above the cutoff. In this case, many will represent incorrect matches and a significant FPR will be observed.

In the current study, we report a simple approach for determining appropriate ion score cutoffs for the identification of ubiquitinated proteins that ensures a low FPR. For our initial experiments, we analysed multi-Ubiquitin (Ub) chains with quadrupole time-of-flight (QqTOF; Qstar)[22,23] and quadrupole ion trap (QIT; LTQ)[24,25] mass spectrometers in combination with the Mascot database search engine. We found that standard ion score cutoffs (i.e. 15 for the Qstar and 26 for the LTQ) were not sufficiently stringent as significant FPR were observed. We also found that FPR and false negative rates (FNR) varied significantly depending on the cutoff score used and that appropriate cutoff scores could only be determined following a systematic evaluation of FPR. When a range of cutoff scores was used for the analysis of complex mixtures of ubiquitinated proteins obtained from human cell lysates by immunopurification with an anti-Ub antibody, unacceptably high FPR were observed. As a final set of experiments, we searched the MS/MS data obtained for the analysis of immunopurified ubiquitinated proteins against five protein databases of increasing size to determine if FPR for ubiquitinated proteins are affected by the size of the protein database that is searched.

## 4.4 Materials and methods

Reagents and ubiquitinated proteins. Mouse monoclonal anti-Ub antibody (clone FK2) and K48- & K63-linked multi-Ub chain standards were purchased from Biomol Inc. 5 μg of

multi-Ub chains were separated by SDS-PAGE and stained with Coomassie blue. H1299 lung adenocarcinoma cells were cultured and harvested as previously described.[26] WCLs were prepared and protein content was determined using a Bradford protein assay kit (Bio-Rad).

Immunopurification of ubiquitinated proteins. Pre-cleared WCLs were loaded onto columns containing 250 μL of protein G-agarose beads coupled to 500 μg of anti-Ub antibody and allowed to incubate for 4 h at 4 °C.[26,27] Columns were washed with RIPA buffer (50 mM Tris-HCl pH 7.4, 150 mM NaCl, 1% NP-40, 0.25% Na deoxycholate, 1 mM EDTA) and bound proteins were eluted with 2% SDS. Trichloroacetic acid was added to a final concentration of 25% and samples were put on ice for 30 min. Samples were then centrifuged for 30 min at 14 000 rpm. Supernatants were removed and protein pellets were washed with chilled (–20 °C) acetone and centrifuged for another 20 min at 14 000 rpm. Acetone was removed and protein pellets were reconstituted in 1% NP-40, 150 mM NaCl, and 10 mM potassium phosphate (KPB buffer). Formic acid up to 10% was added to help dissolve the pellets.

Protein digestion and sample preparation. Gel bands containing multi-Ub chains were excised and subjected to tryptic digestion using a standard in-gel digestion protocol.[28] Immunopurified samples were digested with the Proteomic Reactor,[26] which was assembled by connecting 8 cm of fused silica capillary tubing (200 μM i.d.) to an in-line microfilter (Upchurch Scientific). A pressure bomb was then used to pack 4 cm of strong cation exchange (SCX) beads (The NestGroup) into the opposite end of the capillary tubing by applying 150 PSI of $N_2$ gas. Immunopurified samples were acidified and 2 μg of sequencing-grade trypsin (Promega) was added (0.5 μg/μl in KPB buffer, pH 3.0) to give a final protein: trypsin ratio of approximately 5:1. The supernatant was loaded on the SCX beads using the pressure bomb and $N_2$ gas. Bound

proteins were then washed with 20% acetonitrile (ACN), followed by milli-Q $H_2O$. Disulfide bond reduction was achieved with 100 mM dithiothreitol (in 10 mM ammonium bicarbonate) for 30 min. The SCX resin was then dried and washed with KPB buffer to quench the reaction. Proteins were subjected to simultaneous alkylation and tryptic digestion with 10 mM iodoacetamide (in 100 mM Tris-HCl, pH 8.0) for 2 h at room temperature. Tryptic peptides were eluted with 200 mM ammonium bicarbonate.

LC-MS/MS analysis. Peptides were loaded onto a microtrap column (Dionex, Sunnyvale,CA) for desalting/concentration using an Agilent 1100 HPLC (Agilent, Santa Clara, CA). Peptides were then separated on a C18 Pepmap100 analytical column (Dionex, Sunnyvale, CA) using a 250 nl/min flow with the following gradient of acetonitrile: 0 min: 5%, 2 min: 10%, 50 min: 30%, 57.5 min: 50%, 62 min: 80%, 70 min: 80%, 72.5 min: 5%, 85 min: 5%. Peptides were analysed with a Qstar Elite (Applied Biosystems, Foster City, CA) and a LTQ mass spectrometer (Thermo Fisher Scientific, Waltham, MA) both operated in an information-dependent acquisition mode with a 1 survey scan (MS) followed by 5 product ion scans (MS/MS). MS times were 250 ms (Qstar) or 50 ms (LTQ); MS/MS experiments were 150 ms (LTQ) or dynamic based on the quality of the spectrum (Qstar). Additional immunopurified samples were analysed in duplicate with the Qstar Elite.

Data analysis and bioinformatics. MS/MS data was searched against a concatenated IPI protein database[29] containing forward and reversed human protein sequences using the Mascot search engine (Matrixscience Ltd, Boston, MA). Carbamidomethylation of cysteine (C) residues was selected as a fixed modification while oxidation of methionine (M) residues and ubiquitination of lysine (K) residues were selected as variable modifications. Peptide mass

tolerances were set to 50 ppm (Qstar) or 2 Da (LTQ) and fragment ion mass tolerances were set to 0.15 Da (Qstar) or 0.8 Da (LTQ). The Mascot output was formatted as a peptide summary with standard scoring and a minimum ion score cutoff of 15. The requirement that each protein contain at least one bold red peptide was also selected, which effectively removes duplicate homologous proteins from the results (for more information see www.matrixscience.com/help/interpretation_help.html). The HTML output was copied into a Microsoft Excel spreadsheet and parsed using a modified version of a recently reported tool.[30] The parser removed all non-signature peptides and examined the MS/MS data to ensure that there were at least 3 y- or b-ions in a row (excluding $b_1$-$^{+}$ ions), or 3 y-$^{++}$ or b-$^{++}$ ions in a row (excluding $b_1$-$^{++}$ ions) if the charge state of the precursor was > $2^{+}$. The parser also discarded MS/MS data if the RMS error was greater than 200 ppm (Qstar) or 1000 ppm (LTQ). Signature peptides were then manually validated to ensure that all ubiquitination sites were mapped to internal lysine residues.

## 4.5 Results and discussion

Ubiquitin (Ub) and signature peptides. Ub is a highly conserved 76-amino acid protein that is present in all eukaryotes and differs in only four residues among yeast, plants and mammals.[31,32] In its simplest form, a single Ub molecule is covalently attached via its C-terminal glycine (G76) to a lysine on the target protein through an isopeptide bond. This process, known as mono-ubiquitination, is often followed by the glycine–lysine ligation of additional Ub molecules to the first, forming di-, tri-, tetra or multi-Ub chains. Since Ub itself contains seven

lysines (K6, K11, K27, K29, K33, K48 and K63) a number of different chain linkages are possible (Figure 4.1).

Enzymatic digestion of ubiquitinated proteins with trypsin typically results in the production of signature peptides with a di-glycine (GG) tag that is derived from the carboxy-terminus of Ub. These signature peptides have a monoisotopic mass addition of 114.04292 Da on lysine residues and are characterized by a missed cleavage site because trypsin is unable to cleave after these modified lysine residues. This provides a platform for the identification of ubiquitinated proteins because the GG-tag can be detected as a variable modification after MS/MS analysis and database searching.[26,27] Recently, we reported a more comprehensive approach for identifying ubiquitinated proteins that also involves searching for signature peptides with a LRGG-tag (+383.22809 Da), which is the result of a missed cleavage site at arginine-74 of Ub.[33]

MS/MS analysis of multi-Ub chains. Although several proteomic studies have reported the identification of ubiquitinated proteins in yeast and mammalian cell lysates,[34-36] no systematic method for estimating the accuracy of the results was described. This represents a concern for several large-scale studies that have reported the identification of hundreds or even thousands of post-translational modifications by MS/MS since a number of experimental details, including individual ion scores, have only recently become a requirement for publication.[37,38]

Historically, the proteomics field has relied on standard ion score cutoffs to achieve acceptable FPR for protein identifications. Gygi and co-workers, for example, demonstrated that Mascot ion score cutoffs of 15 and 16 (for $3^+$ and $2^+$ ions, respectively) using a Qstar mass spectrometer or 26 and 30 (for $3^+$ and $2^+$ ions, respectively) using a LTQ mass spectrometer

achieved a 1% FPR for the identification of tryptic peptides from a yeast cell lysate using a gel-based approach.[22] Although these cutoff scores are generally useful for achieving low FPR for protein identifications, they may not be suitable for mapping post-translational modifications, such as phosphorylation and ubiquitination.

To determine if standard ion score cutoffs are suitable for mapping ubiquitination sites, we analysed commercially available multi-Ub chains since they produce a simple mixture of signature peptides following tryptic digestion. Multi-Ub chain isoforms (Stnd 1 and 2) were mixed (Mix) in a 50:50 ratio, separated by SDS-PAGE, and stained with Coomassie Blue. Three gel bands were excised and subjected to in-gel trypsin digestion as shown in Figure 4.2. The digested samples were split into equal fractions and analysed with a QqTOF (Qstar) and a QIT (LTQ) mass spectrometer. The Mascot search engine was then used to search the MS/MS data against a concatenated database containing both forward and reversed human protein sequences. Ion score cutoffs between 15 and 40 were employed. The Mascot output was then copied into Microsoft Excel and filtered using a parser program as described in the Materials and methods section. A complete list of database searching parameters can be found in Table 4.1.

All signature peptide matches were manually validated to ensure that ubiquitination sites were mapped to internal lysine residues. Representative MS/MS spectra of signature peptides obtained with the Qstar and the LTQ are shown in Figure 4.2. The spectrum on the left is a triply charged $[M + 3H]^+$ peptide ion at m/z 487.6 that matched to the sequence LIFAG$K_{GG}$QLEDGR with an ion score of 65, while the spectrum on the right is a triply charged $[M + 3H]^+$ ion at m/z 749.1 that matched to the sequence TLSDYNIQ$K_{GG}$ESTLHLVLR with an ion score of 111.

The top scoring signature peptides matching to Ub were tabulated to determine which lysine residues were involved in Ub-chain branching (Table 4.2). A total of six Ub-chain branching sites (out of a possible of seven) were identified with the Qstar and the LTQ, as shown in Figure 4.2. Each of the five sites identified with the Qstar were observed for all three analyses, while the five sites identified with the LTQ were observed with less consistency. However, the LTQ enabled the identification of a unique branching site that was missed by the Qstar (K27), which is likely due to experimental variation. Overall, our results indicate that the sensitivity of the Qstar is slightly higher than that of the LTQ.

False positive and false negative rates for the analysis of multi-Ub chains. In theory, the probability of non-peptide spectral noise matching to forward or reversed protein sequences in a concatenated database should be equivalent.[19,39] As an example, if five peptides are matched to reversed protein sequences, one can assume that there are also five false positives hidden amongst the matches to the forward protein sequences.[39] Therefore, the total number of false positives in a MS/MS dataset is equal to the number of matches to reversed protein sequences multiplied by two. For this study, we assume that the number of signature peptide matches to reversed protein sequences corresponds to the number of false positives that would have occurred if the original (forward only) database were searched. Therefore, FPR for signature peptides were calculated as follows: FPR = no. of reversed signature peptide matches/no. of forward signature peptide matches.

FPR for signature peptides and standard deviations were calculated for the multi-Ub chain analyses and plotted as a function of the ion score cutoff (Figure 4.3). The average FPR (n = 3) for the Qstar at an ion score cutoff of 15 was 16.54%, which is considerably higher than the

1% FPR for tryptic peptides observed by Gygi and co-workers at the same cutoff.[22] The average FPR for signature peptides improved to 9.02% at a cutoff of 20 and eventually fell to 3.78% at a cutoff of 27. Improved FPR were also observed for the LTQ as the cutoff scores were increased. The average FPR for the LTQ at a cutoff score of 25 was 27.47% and improved to 8.64% at a cutoff score of 35 and eventually fell to 4.60% at a cutoff score of 40. Overall, FPR for signature peptides were lower for the Qstar when compared to the LTQ.

The number of false negatives in a MS/MS dataset can be estimated by calculating the number of correct peptide matches that are rejected at any given cutoff score. Therefore, FNR for signature peptides were calculated as follows: FNR = no. of Ub signature peptides rejected/total no. of Ub signature peptides. FNR for signature peptides and standard deviations were calculated for the multi-Ub chain analyses and plotted as a function of the ion score cutoff (Figure 4.3). The average FNR (n = 3) for the Qstar was negligible for cutoff scores between 15 and 25, indicating that no correct signature peptide matches were rejected in this range. For cutoff scores between 27 and 40, the FNR increased from 5.69 to 24.13%. A nearly identical trend for FNR was observed with the LTQ between cutoff scores of 27 and 40, indicating that roughly the same proportion of correct signature peptide matches are rejected when cutoff scores in this range are used.

Taken together, Figure 4.3 demonstrates that standard ion score cutoffs used with the Qstar (i.e. 15 for $3^+$ ions and 16 for $2^+$ ions) and the LTQ (i.e. 26 for $3^+$ ions and 30 for $2^+$ ions) mass spectrometers are not sufficiently stringent for mapping ubiquitination sites as significant FPR are observed. We believe that the reason why higher cutoff scores are necessary for signature peptides is because database searching using variable modifications on lysine residues

(GG- and LRGG-tags) significantly increases the size of the protein database (i.e. the peptide pool) that is searched, thus increasing the probability of obtaining incorrect signature peptide matches.

MS/MS analysis of complex mixtures of ubiquitinated proteins. To further investigate the effect of ion score cutoff on FPR, we analysed two samples containing a mixture of ubiquitinated proteins (samples 1 and 2). Ubiquitinated proteins were obtained from two populations of human H1299 lung adenocarcinoma cells (± anti-VCP siRNA treatment)[26] by immunopurification with an anti-Ub antibody. Briefly, poly-prep chromatography columns were filled with anti-Ub beads and H1299 WCLs were loaded and allowed to incubate at 4°C. After a 4 h incubation period, which was sufficient time to permit ubiquitinated proteins to bind the anti-Ub beads with high-affinity, a series of washing steps was performed. Bound proteins were eluted with a solution containing a low percentage of ionic detergent and subjected to trypsin digestion step using a microfluidic protein processing device known as the Proteomic Reactor.[40]

The two digested samples were split into equal fractions and analysed with the Qstar and LTQ mass spectrometers. The Mascot search engine was used to search the MS/MS data using the same criteria as the gel band analyses with ion score cutoffs between 15 and 40. The Mascot output was filtered and FPR were calculated as previously described. The number of signature peptides matching to forward and reversed protein sequences for sample 1 is shown in Figure 4.4. At a cutoff score of 15, a total of 35 signature peptides were identified with the Qstar, seven of which were false positives; FPR = 7/28 = 25%. At the same cutoff score (15), 235 signature peptides were identified with the LTQ. Although this represents a seven-fold increase in the number of signature peptides identified, the FPR was extremely high at roughly 93%; FPR =

113/122 = 92.6%. At a cutoff score of 23, 11 signature peptides were identified with the Qstar, only one of which was a false positive; FPR = 1/10 = 10%. At the same cutoff score (23), 49 signature peptides were identified with the LTQ but once again, the FPR was beyond any acceptable level at 104.1%; FPR = 25/24 = 104.1%. The remaining FPR for sample 1 were calculated and running averages were plotted as a function of the ion score cutoff (Figure 4.4).

A similar decreasing trend for signature peptide FPR was observed with the Qstar for sample 2. At a cutoff score of 15, a total of 25 signature peptides were identified with the Qstar, six of which were false positives; FPR = 6/19 = 31.5% (Figure 4.4). While at a higher cutoff score of 23, 11 signature peptides were identified, with only one false positive; FPR = 1/10 = 10%. The results for sample 2 with the LTQ were slightly improved compared to sample 1. At a cutoff score of 15, 182 signature peptides were identified, with a corresponding FPR of 111.6%. For cutoff scores between 23 and 30, FPR decreased from 50 to 33.3%. Only at a cutoff score of 40, did the FPR fall to below 25% for the LTQ, which was problematic since only three signature peptide matches to forward protein sequences were observed. The remaining FPR for sample 2 were calculated and running averages were plotted as a function of the ion score cutoff.

When FPR for samples 1 and 2 were compared to those obtained for the multi-Ub chain analyses, a significant increase was observed for both the Qstar and the LTQ. We believe that this is a result of the complex nature of samples 1 and 2 which increased the total amount of MS/MS data obtained and decreased its overall quality due to matrix effects, such as ionization competition, greater numbers of co-eluting peptides, and isobaric interferences. The differences in FPR observed with the Qstar and the LTQ for the same samples, however, may be attributed

to the fact that the Qstar provides higher mass accuracy and resolution and its ability to select only multiply charged ions for fragmentation, thus reducing the amount of spectral noise.

False positive rates and protein database size. One of the factors that can influence the FPR for a MS/MS dataset is the size of the protein database that is searched. This is because larger databases contain a greater number of sequences, which in turn, increases the probability that a search engine will randomly match non-peptide spectral noise to peptide sequences. To test whether protein database size influences FPR for signature peptides, we searched the MS/MS data for samples 1 and 2 obtained with the Qstar against five databases of increasing size; 1) Homo sapiens (human) containing 153 777 sequences; 2) modified Homo sapiens containing 248 060 sequences; 3) Chordata (vertebrates and relatives) containing 690 662 sequences; 4) Eukaryota (eukaryotes) containing 1 604 804 sequences; and 5) All entries (all species) containing 3 893 302 sequences. Chymotrypsin, rather than trypsin, was specified as the proteolytic enzyme so all signature peptide matches would represent known false positives. An ion score cutoff of 25 was employed for all database searches.

The number of signature peptide and non-signature peptide false positives for sample 1 is shown in Figure 4.5. A single signature peptide false positive was observed when the Homo sapiens and modified Homo sapiens databases were searched, 8 when the Chordata database was searched, 26 when the Eukaryota database was searched, and 50 when the All entries database was searched. Therefore, as the database size increased, there was a corresponding increase in the number of signature peptide false positives. The $R^2$ value for this relationship was 0.9838, indicating a linear correlation. A linear correlation was also observed between database size and the number of non-signature peptide false positives as the $R^2$ value was 0.9811.

The number of signature peptide and non-signature peptide false positives for sample 2 is shown in Figure 4.5. Two signature peptide false positives were observed when the Homo sapiens and modified Homo sapiens databases were searched, 10 when the Chordata database was searched, 38 when the Eukaryota database was searched, and 55 when the All entries database was searched. The calculated $R^2$ value for this relationship was 0.9232, which is slightly lower than the $R^2$ value for sample 1 but was still sufficiently high to indicate a linear correlation. A linear correlation was also observed between database size and the number of non-signature peptide false positives as the calculated $R^2$ value was 0.9801. Overall, these results clearly demonstrate that protein database size affects the FPR for both signature peptides and non-signature peptides.

## 4.6 Conclusion

MS/MS analysis of ubiquitinated proteins requires the use of higher ion score cutoffs compared to conventional protein identifications. This is likely the result of an increase in protein database size when variable modifications on lysine residues (GG- and LRGG-tags) are considered. When complex samples containing a mixture of ubiquitinated proteins are analysed, even more stringent criteria may be necessary due to significant matrix effects. Appropriate ion score cutoffs should be determined on a sample by sample basis by performing a systematic evaluation of FPR for a range of cutoff scores. To further ensure a low false positive rate for the identification of ubiquitinated proteins, mass spectrometers capable of high mass accuracy and resolution should be employed.

**Figure 4.1 Ubiquitin and signature peptides.** Ub is 76-amino acid protein that is covalently attached via its C-terminal glycine (G76) to a lysine (K) on a target protein. When ubiquitinated proteins are subjected to trypsin digestion, signature peptides with a GG– or LRGG–tag are produced. Signature peptides have monoisotopic mass increases of +114.04292 or +383.22809 Da and missed cleavage sites that can be identified by tandem mass spectrometry (MS/MS) analysis.

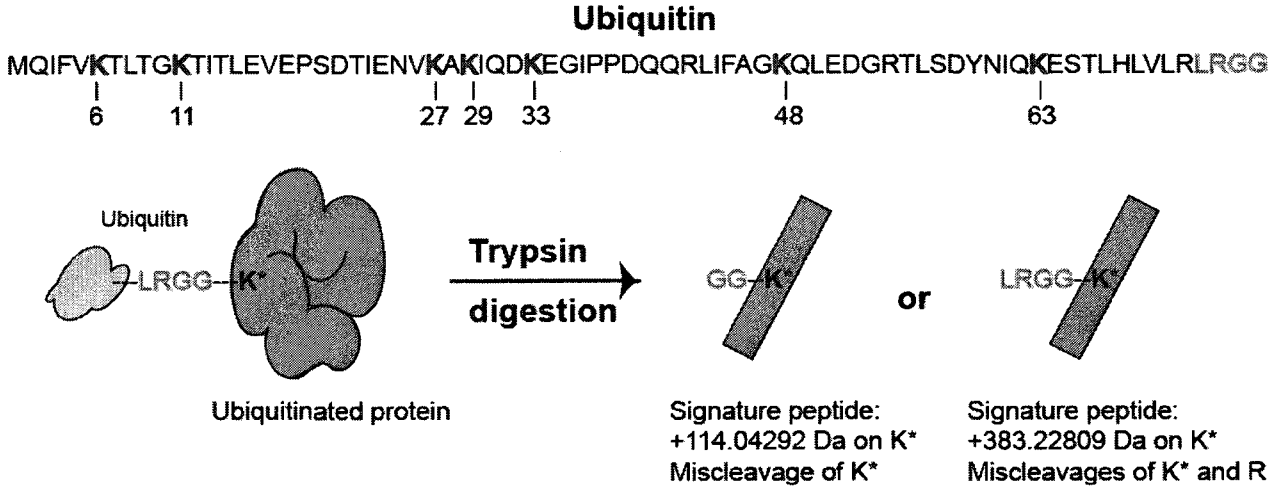


**Figure 4.2 MS/MS analysis of multi-Ub chains.** (a) Multi-Ub chains (Stnd 1 and 2) were mixed (Mix) in a 50:50 ratio, separated by SDS-PAGE, and stained with Coomassie blue. Three bands were excised and subjected to in-gel trypsin digestion. (b) The digested samples were split into equal fractions and analysed with QqTOF (Qstar) and QIT (LTQ) mass spectrometers. Representative MS/MS spectra of signature peptides are shown. (c) The top scoring signature peptides after database searching with the Mascot search engine were tabulated and the number of times that each lysine (K) residue was involved in Ub-chain branching is shown.

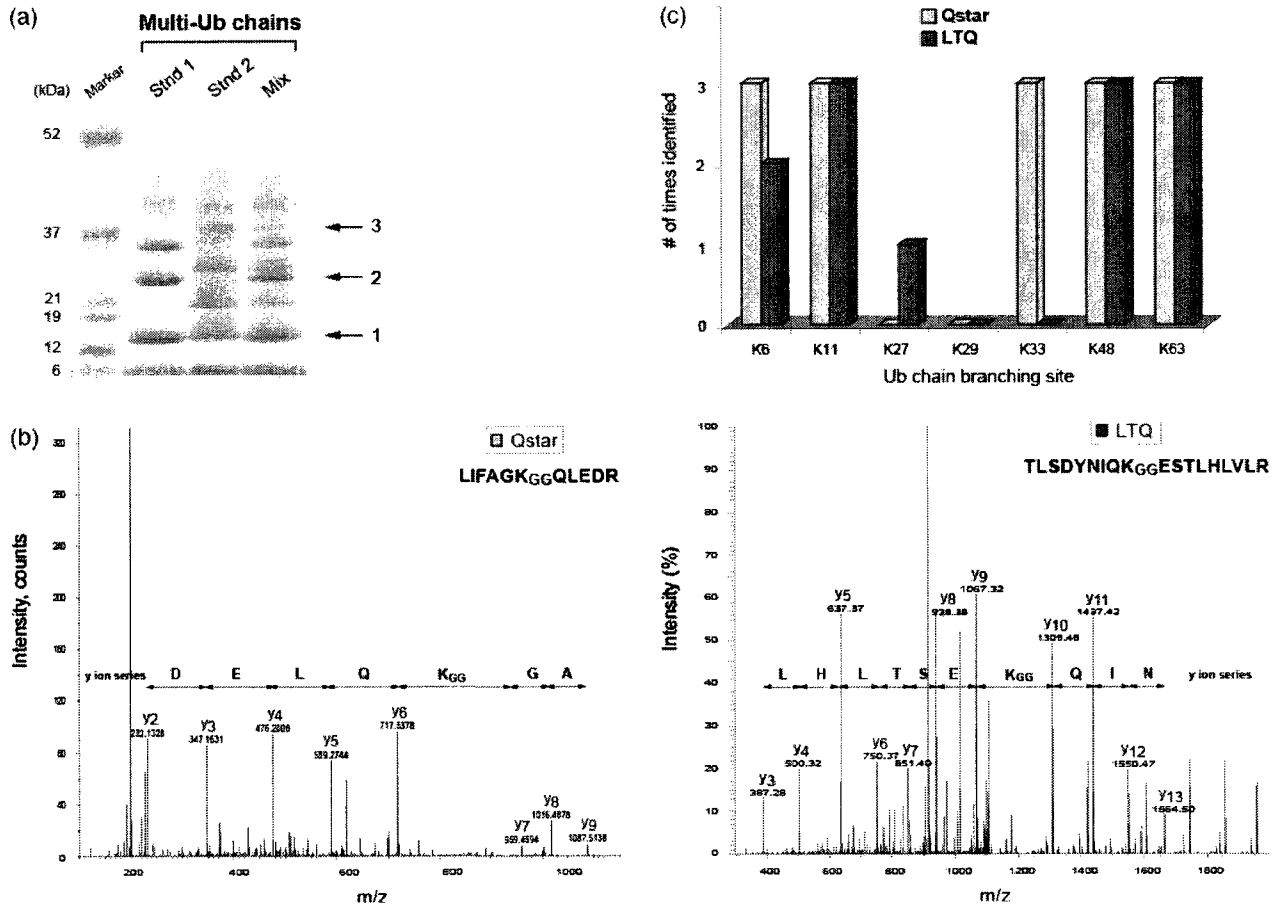


**Figure 4.3 False positive and false negative rates for multi-Ub chains.** MS/MS data from the multi-Ub chain analyses were searched against a concatenated database containing forward and reversed human protein sequences with the Mascot search engine using ion score cutoffs between 15 and 40. (a) False positive rates for signature peptides were calculated and averages (n = 3) were plotted as a function of the ion score cutoff. (b) False negative rates for Ubiquitin signature peptides were calculated and averages (n = 3) were plotted as a function of the ion score cutoff.

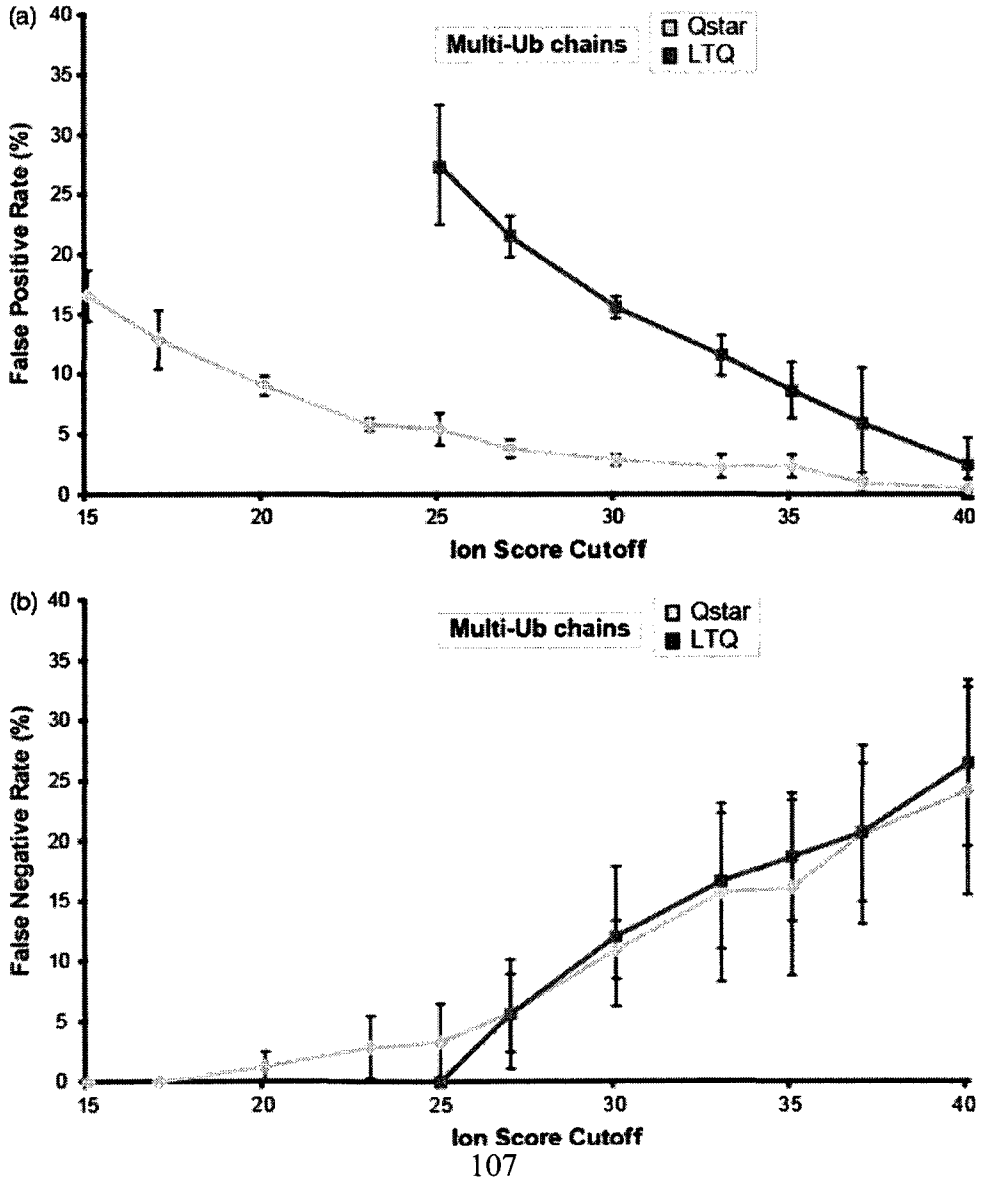


**Figure 4.4 False positive rates for complex mixtures of ubiquitinated proteins.** Immunopurified ubiquitinated proteins (samples 1 and 2) were subjected to trypsin digestion with the Proteomic Reactor. The digested samples were split into equal fractions and analysed with the Qstar and LTQ mass spectrometers. MS/MS data was searched against a concatenated database containing forward and reversed human protein sequences with the Mascot search engine using ion score cutoffs between 15 and 40. (a) and (c) The number of signature peptides that matched to forward and reversed protein sequences for each cutoff score is shown. (b) and (d) False positive rates were calculated for each sample (n = 1) and plotted as a function of the ion score cutoff.

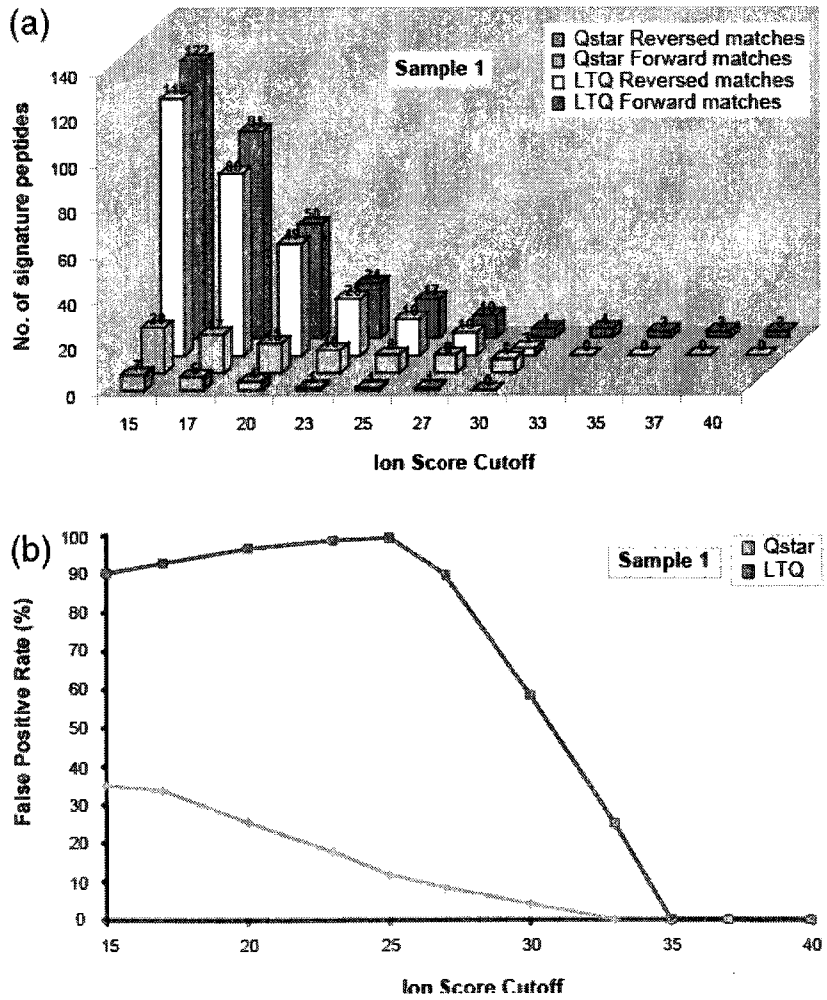


**Figure 4.4 continued**

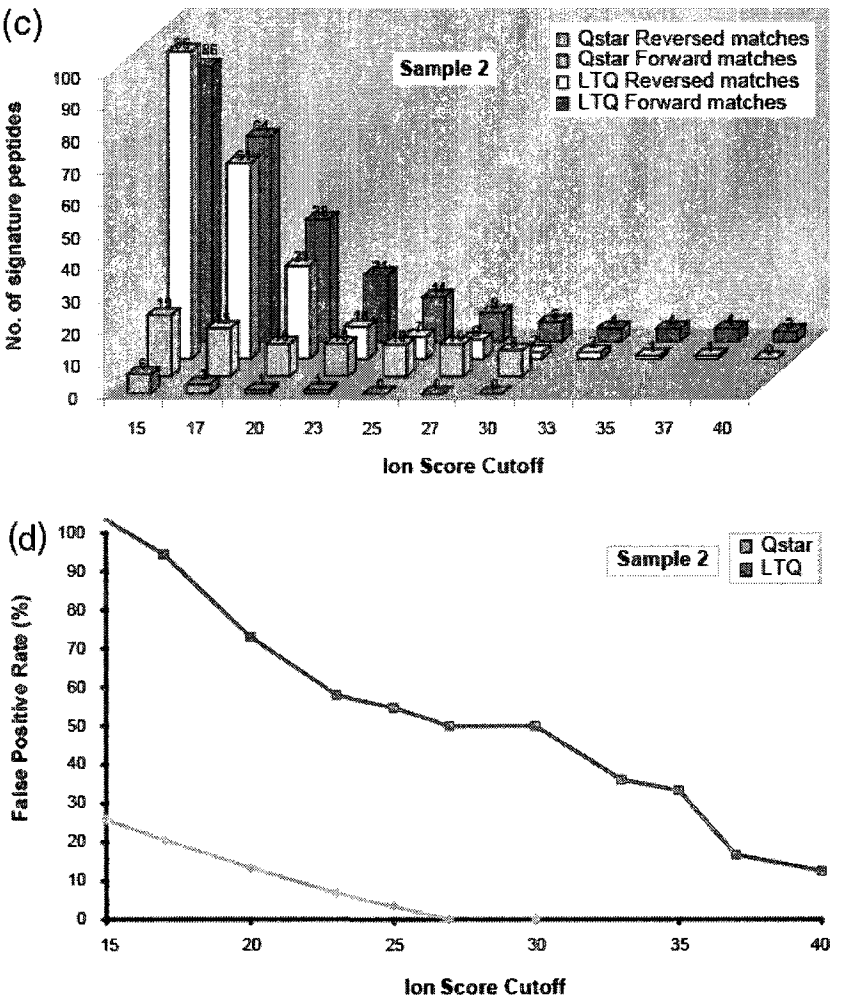
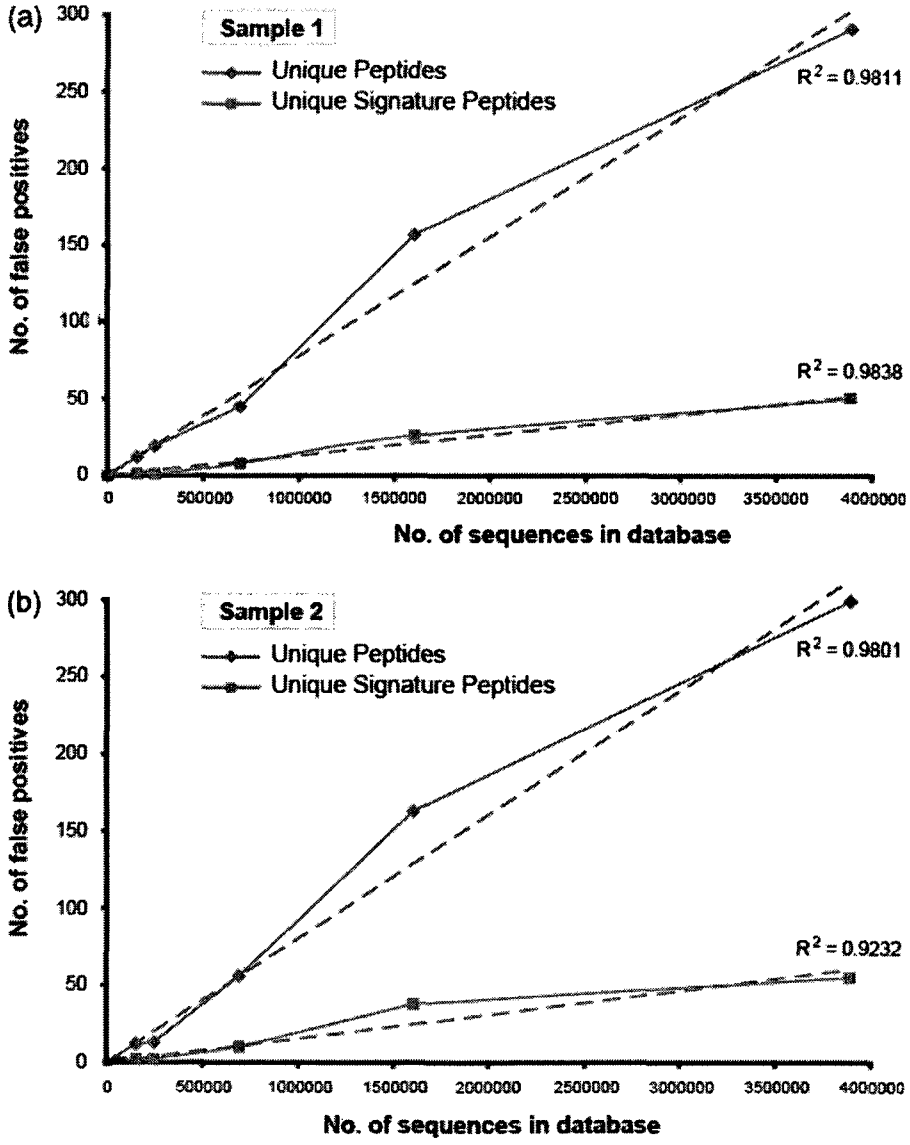


**Figure 4.5 False positive rates and protein database size.** MS/MS data obtained with the Qstar for the analysis of immunopurified ubiquitinated proteins (samples 1 and 2) was searched against five protein databases with the Mascot search engine; (1) Homo sapiens; (2) modified Homo sapiens; (3) Chordata; (4) Eukaryota; and (5) All species. Chymotrypsin, rather than trypsin, was specified as the proteolytic enzyme. (a) and (b) The number of signature peptide and non-signature peptide false positives for the five databases are shown. Trend lines and correlation coefficients ($R^2$) are also shown.



## Table 4.1 Mascot database searching parameters.

| Instrument | Database | Missed cleavages | Fixed modifications | Variable modifications[a] | Peptide tolerance | MS/MS tolerance |
|---|---|---|---|---|---|---|
| Qstar | Human IPI | 2 | Carbamidomethyl (C) | Oxidation (M) | ±50 ppm | ±0.15 Da |
| – | – | – | – | Ubiquitination (K) | – | – |
| LTQ | Human IPI | 2 | Carbamidomethyl (C) | Oxidation (M) | ±2.0 Da | ±0.8 Da |
| – | – | – | – | Ubiquitination (K) | – | – |

[a] Only internal lysines (K) with a GG- or LRGG- tag were considered for ubiquitination.

## Table 4.2 Signature peptides identified from multi-ubiquitin (Ub) chains.

| Band no. | Signature peptide | Instrument | | Ion score[a] | | Ub site |
|---|---|---|---|---|---|---|
| | | Qstar | LTQ | Qstar | LTQ | |
| 1 | MQIFVK$_{GG}$TLTGK | x | – | 49 | – | K6 |
| | MQIFVKTLTGK$_{LRGG}$TITLEVEPSDTIENVK | – | x | – | 32 | K11 |
| | TLTGK$_{GG}$TITLEVEPSDTIENVK | x | x | 111 | 81 | K11 |
| | TITLEVEPSDTIENVK$_{GG}$AK | – | x | – | 28 | K27 |
| | IQDK$_{GG}$EGIPPDQQR | x | – | 53 | – | K33 |
| | LIFAGK$_{GG}$QLEDGR | x | x | 71 | 87 | K48 |
| | LIFAGK$_{GG}$QLEDGRTLSDYNIQK | x | x | 35 | 26 | K48 |
| | LIFAGK$_{LRGG}$QLEDGR | x | – | 17 | – | K48 |
| | TLSDYNIQK$_{GG}$ESTL | x | x | 78 | 49 | K63 |
| | TLSDYNIQK$_{GG}$ESTLHLVLR | x | x | 113 | 110 | K63 |
| | TLSDYNIQK$_{LRGG}$ESTLHLVLR | x | x | 25 | 82 | K63 |
| 2 | MQIFVK$_{GG}$TLTGK | x | x | 48 | 80 | K6 |
| | TLTGK$_{GG}$TITLEVEPSDTIENVK | x | x | 90 | 100 | K11 |
| | IQDK$_{GG}$EGIPPDQQR | x | – | 53 | – | K33 |
| | LIFAGK$_{GG}$QLEDGR | x | x | 85 | 89 | K48 |
| | LIFAGK$_{GG}$QLEDGRTLSDYNIQK | x | x | 37 | 89 | K48 |
| | LIFAGK$_{LRGG}$QLEDGR | – | x | – | 43 | K48 |
| | LIFAGK$_{LRGG}$QLEDGRTLSDYNIQK | x | x | 20 | 38 | K48 |
| | TLSDYNIQK$_{GG}$ESTLHLVLR | x | x | 73 | *111* | K63 |
| | TLSDYNIQK$_{GG}$ESTL | x | – | 43 | – | K63 |
| | TLSDYNIQK$_{LRGG}$ESTLHLVLR | – | x | – | 63 | K63 |
| 3 | MQIFVK$_{GG}$TLTGK | x | x | 56 | 76 | K6 |
| | TLTGK$_{GG}$TITLEVEPSDTIENVK | x | x | 92 | 80 | K11 |
| | TITLEVEPSDTIENVK$_{GG}$AK | – | x | – | 41 | K27 |
| | IQDK$_{GG}$EGIPPDQQR | x | – | 45 | – | K33 |
| | LIFAGK$_{GG}$QLEDGR | x | x | 65 | 89 | K48 |
| | LIFAGK$_{LRGG}$QLEDGR | – | – x | – | 50 | K48 |
| | LIFAGK$_{GG}$QLEDGRTLSDYNIQK | x | – | 28 | – | K48 |
| | TLSDYNIQK$_{GG}$ESTL | x | x | 62 | 62 | K63 |
| | TLSDYNIQK$_{GG}$ESTLHLVLR | x | x | 126 | 114 | K63 |
| | TLSDYNIQK$_{LRGG}$ESTLHLVLR | x | x | 30 | 76 | K63 |

[a] Only the top ion score is shown for each signature peptide.

## 4.7 References

(1) Cannon WR, Jarman KH, Webb-Robertson BJ, Baxter DJ, Oehmen CS, Jarman KD, Heredia-Langner A, Auberry KJ, Anderson GA. Comparison of probability and likelihood models for peptide identification from tandem mass spectrometry data. Journal of Proteome Research 2005; 4: 1687.

(2) Eriksson J, Fenyo D. The statistical significance of protein identification results as a function of the number of protein sequences searched. Journal of Proteome Research 2004; 3: 979.

(3) Haas W, Faherty BK, Gerber SA, Elias JE, Beausoleil SA, Bakalarski CE, Li X, Villen J, Gygi SP. Optimization and use of peptide mass measurement accuracy in shotgun proteomics. Molecular and Cellular Proteomics 2006; 5: 1326.

(4) Huttlin EL, Hegeman AD, Harms AC, Sussman MR. Prediction of error associated with false-positive rate determination for peptide identification in large-scale proteomics experiments using a combined reverse and forward peptide sequence database strategy. Journal of Proteome Research 2007; 6: 392.

(5) Keller A, Nesvizhskii AI, Kolker E, Aebersold R. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. Analytical Chemistry 2002; 74: 5383.

(6) Qian WJ, Liu T, Monroe ME, Strittmatter EF, Jacobs JM, Kangas LJ, Petritis K, Camp DG, Smith RD. Probability-based evaluation of peptide and protein identifications from tandem mass spectrometry and SEQUEST analysis: the human proteome. Journal of Proteome Research 2005; 4: 53.

(7) Shadforth I, Dunkley T, Lilley K, Crowther D, Bessant C. Confident protein identification using the average peptide score method coupled with search-specific, ab initio thresholds. Rapid Communications in Mass Spectrometry 2005; 19: 3363.

(8) Stead DA, Preece A, Brown AJ. Universal metrics for quality assessment of protein identifications by mass spectrometry. Molecular and Cellular Proteomics 2006; 5: 1205.

(9) Weatherly DB, Astwood JA, Minning TA, Cavola C, Tarleton RL, Orlando R. A Heuristic method for assigning a falsediscovery rate for protein identifications from Mascot database search results. Molecular and Cellular Proteomics 2005; 4: 762.

(10) Beausoleil SA, Jedrychowski M, Schwartz D, Elias JE, Villen J, Li J, Cohn MA, Cantley LC, Gygi SP. Large-scale characterization of HeLa cell nuclear phosphoproteins. Proceedings of the National Academy of Sciences of the United States of America 2004; 101: 12130.

(11) Beausoleil SA, Villen J, Gerber SA, Rush J, Gygi SP. A probability-based approach for high-throughput protein phosphorylation analysis and site localization. Nature Biotechnology 2006; 24: 1285.

(12) DeGnore JP, Qin J. Fragmentation of phosphopeptides in an ion trap mass spectrometer. Journal of the American Society for Mass Spectrometry 1998; 9: 1175.

(13) Ficarro SB, McCleland ML, Stukenberg PT, Burke DJ, Ross MM, Shabanowitz J, Hunt DF, White FM. Phosphoproteome analysis by mass spectrometry and its application to Saccharomyces cerevisiae. Nature Biotechnology 2002; 20: 301.

(14) Karty JA, Reilly JP. Deamidation as a consequence of beta elimination of phosphopeptides. Analytical Chemistry 2005; 77: 4673.

(15) Tao WA, Wollscheid B, O'Brien R, Eng JK, Li XJ, Bodenmiller B, Watts JD, Hood L, Aebersold R. Quantitative phosphoproteome analysis using a dendrimer conjugation chemistry and tandem mass spectrometry. Nature Methods 2005; 2: 591.

(16) Yang F, Stenoien DL, Strittmatter EF, Wang J, Ding L, Lipton MS, Monroe ME, Nicora CD, Gristenko MA, Tang K, Fang R, Adkins JN, Camp DG, Chen DJ, Smith RD. Phosphoproteome profiling of human skin fibroblast cells in response to low- and high-dose irradiation. Journal of Proteome Research 2006; 5: 1252.

(17) Eng JK, McCormack AL, Yates JR. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. Journal of the American Society for Mass Spectrometry 1994; 5: 976.

(18) Perkins DN, Pappin DJ, Creasy DM, Cottrell JS. Probability based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis 1999; 20: 3551.

(19) Higdon R, Hogan JM, Van Belle G, Kolker E. Randomized sequence databases for tandem mass spectrometry peptide and protein identification. Omics 2005; 9: 364.

(20) Li X, Gerber SA, Rudner AD, Beausoleil SA, Haas W, Villen J, Elias JE, Gygi SP. Large-scale phosphorylation analysis of alpha factor-arrested Saccharomyces cerevisiae. Journal of Proteome Research 2007; 6: 1190.

(21) MacCoss MJ, Wu CC, Yates JR. Probability-based validation of protein identifications using a modified SEQUEST algorithm. Analytical Chemistry 2002; 74: 5593.

(22) Elias JE, Haas W, Faherty BK, Gygi SP. Comparative evaluation of mass spectrometry platforms used in large-scale proteomics investigations. Nature Methods 2005; 2: 667.

(23) Steen H, Kuster B, Mann M. Quadrupole time-of-flight versus triple-quadrupole mass spectrometry for the determination of phosphopeptides by precursor ion scanning. Journal of Mass Spectrometry 2001; 36: 782.

(24) Mayya V, Rezaul K, Cong YS, Han D. Systematic comparison of a two-dimensional ion trap and a three-dimensional ion trap mass spectrometer in proteomics. Molecular and Cellular Proteomics 2005; 4: 214.

(25) Schwartz JC, Senko MW, Syka JE. A two-dimensional quadrupole ion trap mass spectrometer. Journal of the American Society for Mass Spectrometry 2002; 13: 659.

(26) Vasilescu J, Zweitzig DR, Denis NJ, Smith JC, Ethier M, Haines DS, Figeys D. The proteomic reactor facilitates the analysis of affinity-purified proteins by mass spectrometry: application for identifying ubiquitinated proteins in human cells. Journal of Proteome Research 2007; 6: 298.

(27) Vasilescu J, Smith JC, Ethier M, Figeys D. Proteomic analysis of ubiquitinated proteins from human MCF-7 breast cancer cells by immunoaffinity purification and mass spectrometry. Journal of Proteome Research 2005; 4: 2192.

(28) Wilm M, Shevchenko A, Houthaeve T, Breit S, Schweigerer L, Fotsis T, Mann M. Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry. Nature 1996; 379: 466.

(29) Kersey PJ, Duarte J, Williams A, Karavidopoulou Y, Birney E, Apweiler R. The International Protein Index: an integrated database for proteomics experiments. Proteomics 2004; 4: 1985.

(30) Smith JC, Duchesne MA, Tozzi P, Ethier M, Figeys D. A differential phosphoproteomic analysis of retinoic Acid-treated p19 cells. Journal of Proteome Research 2007; 6: 3174.

(31) Catic A, Ploegh HL. Ubiquitin--conserved protein or selfish gene? Trends in Biochemical Sciences 2005; 30: 600.

(32) Glickman MH, Ciechanover A. The ubiquitin-proteasome proteolytic pathway: destruction for the sake of construction. Physiological Reviews 2002; 82: 373.

(33) Denis NJ, Vasilescu J, Lambert JP, Smith JC, Figeys D. Tryptic digestion of ubiquitin standards reveals an improved strategy for identifying ubiquitinated proteins by mass spectrometry. Proteomics 2007; 7: 868.

(34) Kirkpatrick DS, Weldon SF, Tsaprailis G, Liebler DC, Gandolfi AJ. Proteomic identification of ubiquitinated proteins from human cells expressing His-tagged ubiquitin. Proteomics 2005; 5: 2104.

(35) Matsumoto M, Hatakeyama S, Oyamada K, Oda Y, Nishimura T, Nakayama KI. Large-scale analysis of the human ubiquitin related proteome. Proteomics 2005; 5: 4145.

(36) Peng J, Schwartz D, Elias JE, Thoreen CC, Cheng D, Marsischky G, Roelofs J, Finley D, Gygi SP. A proteomics approach to understanding protein ubiquitination. Nature Biotechnology 2003; 21: 921.

(37) Bradshaw RA, Burlingame AL, Carr S, Aebersold R. Reporting protein identification data: the next generation of guidelines. Molecular and Cellular Proteomics 2006; 5: 787.

(38) Carr S, Aebersold R, Baldwin M, Burlingame A, Clauser K, Nesvizhskii A. The need for guidelines in publication of peptide and protein identification data: Working Group on Publication Guidelines for Peptide and Protein Identification Data. Molecular and Cellular Proteomics 2004; 3: 531.

(39) Elias JE, Gygi SP. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. Nature Methods 2007; 4: 207.

(40) Ethier M, HouW, Duewel HS, Figeys D. The proteomic reactor: a microfluidic device for processing minute amounts of protein prior to mass spectrometry analysis. Journal of Proteome Research 2006; 5: 2754.

# Chapter 5: Identification of lysines within membrane-anchored Mga2p120 that are targets of Rsp5p ubiquitination and mediate mobilization of tethered Mga2p90

## 5.1 Manuscript status and statement of author contributions

A version of this chapter has been published in the Journal of Molecular Biology: *Bhattacharya S, Shcherbik N, Vasilescu J, Smith JC, Figeys D, Haines DS. Identification of lysine residues within the carboxy-terminus of Mga2p120 that are targets of Rsp5p-dependent ubiquitination and mediate mobilization of interaction Mga2p90. Journal of Molecular Biology. Vol. 385: 718-725, 2009.* SB and NS conceived, designed and performed most of the experiments. JV performed the mass spectrometry and data analysis to map protein ubiquitination sites. JCS assisted with the data analysis. JV and DF edited the manuscript. DSH wrote the manuscript.

## 5.2 Summary

Mga2p90 is an endoplasmic reticulum (ER)-localized transcription factor that is released from the ER membrane by a unique ubiquitin (Ub)-dependent mechanism. Mga2p90 mobilization requires polyubiquitination of its associating membrane-bound Mga2p120 anchor and subsequent Mga2p120–Mga2p90 complex disassembly that is mediated by ATPase Cdc48p and its heteromeric Ub-binding adaptor Npl4p-Ufd1p. Although previous studies have identified the Ub ligase (i.e., Rsp5p) and ligase binding site on Mga2p120 that play a role in this process,

the amino acids of Mga2p120 that are targets of ubiquitination and promote Mga2p90 mobilization are unknown. We have identified, using mass spectrometry analysis of *in vitro* ubiquitinated Mga2p120–Mga2p90 complex, that lysine residues 983 and 985 contained within the carboxy-terminal domain of Mga2p120 are Rsp5p-directed Ub-conjugation sites. Mutation of these residues as well as proximally located lysine 980 results in suppression of Mga2p120 ubiquitination *in vitro* and *in vivo*, inefficient liberation of Mga2p90 by $Cdc48p^{Npl4p/Ufd1p}$ *in vitro*, and ER retention of Mga2p in cells. Moreover, mga2Δ/spt23ts harboring Rsp5p binding and conjugation mga2 mutants express low OLE1 (an Mga2p90 target gene) transcripts and display reduced growth. We conclude that residues 980, 983, and 985 are targets of Rsp5pinduced polyubiquitination and mediate $Cdc48p^{Npl4p/Ufd1p}$-dependent Mga2p90–Mga2p120 separation and Mga2p90 mobilization.

## 5.3 Introduction

Ubiquitination is a post translation modification process where a 76-amino-acid polypeptide called ubiquitin (Ub) is covalently attached to a lysine residue of substrate.[1,2] This process requires the coordinated activities of a Ub-activating enzyme (E1), a Ub conjugating enzyme (E2), and a Ub ligase (E3). While an attachment of a single Ub moiety can occur, generally, this reaction occurs repeatedly, leading to the attachment of Ub to internal lysines of Ub and/or proximal lysine residues on the substrate. The best understood consequence of substrate ubiquitination is targeting the protein for destruction by the proteasome.[3] It is becoming increasingly apparent, however, that a Ub signal can regulate protein function by non-degradative mechanisms, including modulating protein–protein interactions.[4]

Ubiquitination plays unique roles in regulating the expression and localization of the Mga2p transcriptional regulator. Mga2p is one of two (Spt23p being the other) homologous endoplasmic reticulum (ER)-localized transcription factors in *Saccharomyces cerevisiae*.[5] Full-length Mga2p is approximately 120 kDa in size and harbors a carboxy-terminal localized transmembrane domain. It is retained at the ER after synthesis and undergoes homodimerization, which is mediated by the centrally localized IPT (Ig-like/plexin/transcription factor) domain.[6,7] After dimerization, ubiquitination of one of the 120-kDa monomers recruits the proteasome, resulting in endoproteolytic cleavage and bidirectional degradation of one of the monomers.[8,9] While this process destroys the carboxy-terminal domain of the monomer, the proteasome is unable to degrade the central and amino-terminal domains of Mga2p. A 90-kDa Mga2p fragment (termed Mga2p90) remains tethered to the membrane via an interaction with its 120-kDa unprocessed anchor (termed Mga2p120).[6,7]

In addition to facilitating Mga2p90 generation, ubiquitination plays an unusual role in promoting release of the processed product from the ER. This Ub-directed event does not involve the proteasome, but is dependent on the Cdc48p and its Ub-binding adaptor composed of Npl4 and Ufd1p. Cdc48p (also called p97 or VCP) is a highly conserved member of the AAA (ATPase associated with diverse cellular activities) ATPase family of molecular chaperones. This protein, in collaboration with specific adaptor proteins that bind to ubiquitinated substrates, participates in numerous Ub-selective processes such as ubiquitination, deubiquitination, and Ub-dependent protein complex disassembly.[10] It is believed that Cdc48p applies mechanical force onto substrates upon ATP hydrolysis, leading to p97-dependent changes in their structure and subsequent biochemical events.[11] Ubiquitination of Mga2p120 by the Rsp5p E3 ligase recruits $Cdc48p^{Npl4p/Ufd1p}$ to an Mga2p120–Mga2p90 complex. This interaction results in

$Cdc48p^{Npl4p/Ufd1p}$-mediated Mga2p120–Mga2p90 complex disassembly and release of Mga2p90 from the membrane.[12] Upon release, Mga2p90 is capable of migrating to the nucleus where it induces transcription of OLE1 and other genes involved in lipid metabolism.[5,13]

## 5.4 Materials and methods

Plasmids. Plasmids pYes-FLAGMGA2 and pGEX-6p-1-RSP5 have been described previously. Mutants were generated by PCR-based site-directed mutagenesis using the low error rate polymerase PFU. mga2-k980r mutants were generated using the following primers: 5'-CTTGGGGCGAGATGATAAATTGAAAACC-3' and 5'GGTTTTCAATTTATCATCTC GCCCCAAG-3', mga2-k983r mutants were generated using the following primers: 5'-AGATGATAAATTGCGAACCACAAATCAAGACAGTATTGTGGAGC-3' and 5'-GCTCCA CAATACTGTCTTGATTTGTGGTTCGCAATTTATCATCT-3', finally mga2-k985r mutants were generated using the following primers: 5'- AGATGATCGATTGAAAACCACAAAT CAAGACAGTATTGTGGAGC-3' and 5'- GCTCCACAATACTGTCTTGATTTGTGGTTTT CAATCGATCATCT-3'. pYes-FLAGmga2k983/985r (referred to as mga2-2k in the text) was generated by using pYes-FLAGMGA2 as the template and primers: 5'-GGGTCGAGATGAT CGATTGCGAACCACAAATCAAGACAGTATTGTGGAGC-3' and 5'- GCTCCACAATACT GTCTTGATTTGTGGTTCGCAATCGATCATCTCGACCC-3'. pYes-FLAG mga2k980/983/ 985r (referred to as mga2-3k) was generated the same way except pYes-FLAGmga2k983/985r was used as a template and primers 5'-GTTGTCTTGGGGTCGAGATGATCGATTGC-3' and 5'-GCAATCGATCATCTCGACCCCAAGACAAC-3' were used for the PCR reactions.

Yeast strains and viability assays. InvSc1 was purchased from Invitrogen while the mga2Δspt23ts strain was kindly provided by David Garfinkel (National Cancer Institute). mga2Δspt23ts transformed with expressing constructs were grown overnight at 25°C. The OD was measured the next day and equal number of cells from each culture was subjected to serial dilutions (1:10). Cells were placed in 96 well plates and were replica-plated onto plates lacking or containing 1 mM oleic acid. Plates were incubated at 34°C for 3 days.

Isolation of microsomes. Wild type InVsc cells were transformed with plasmids encoding Mga2p. Transformed cells were picked and grown in synthetic dropout media for 24 hours. Mga2p expression was induced with synthetic dropout media containing galactose. Following induction, cells were collected and lysed with acid washed glass beads (Sigma) by vortexing at 4°C in buffer containing, 200 mM D-Sorbitol, 75 mM NaCl, 20 mM NaH2P04 (pH 7.5) and 1mM MgCl2 with protease inhibitors cocktail comprised of Aprotinin, Pepstatin A, Leupeptin, and Phenylmethylsulfonyl Fluoride. Following lysis, the cell debris was separated by centrifugation. The supernatant was collected and subjected to ultracentrifugation at 50,000 rpm for 1 hour in a Sorvall OTD 70B with a 42.2 TI rotor. The microsomes were either used for mobilization assays or the proteins in the microsomal fraction were solubilized by incubating the membranes in SLIP buffer containing, 150 mM NaCl, 50 mM HEPES (pH 7.5), 10% glycerol, 0.1% Triton X-100, 10 μM proteasome inhibitor MG115 and protease inhibitor cocktail. The solubilized proteins were subjected to immunoprecipitation with anti-FLAG affinity gel M2 (Sigma) to isolate Mga2p120-Mga2p90 complexes.

*In vitro* ubiquitination assays. Following immunoprecipitation of Mga2p120-Mga2p90 complexes, the beads were washed several times with RIPA buffer containing 0.5 M NaCl and then once with 0.25M Tris-HCl (pH 7.4). Ubiquitination assays were carried out in 10 mM Tris-HCl (pH 7.5), 20 mM NaCl, 125 μM DTT, 5 mM MgCl2, 5 mM ATP, 50 μg/ml ubiquitin (Sigma), 20 ng of purified human E1 and 10 ng of purified yeast E2 Ubc1p. Recombinant Rsp5p (10-50 ng) or buffer control was added to initiate the reactions. The reactions were carried out for 40 min at room temperature and then the beads were again washed with RIPA buffer containing 0.5M NaCl and then with elution buffer containing, 50mM Tris-HCl (pH 8), 200 mM KCl, 25mM MgCl2, 1% Triton X-100, 5% glycerol, 0.1M DTT, 10 μM MG115, and protease inhibitor cocktail. The FLAG-tagged proteins were eluted from the beads by adding 100 ng/μL of FLAG peptide (Sigma) in the aforementioned elution buffer at 4°C for 2 to 4 hours. The beads were then subjected to micro-centrifugation and the supernatants were collected. The beads were subjected to at least two rounds of extraction with the FLAG peptide. The eluted proteins were then denatured by adding SDS-PAGE loading buffer and boiling for 5 minutes. These proteins were resolved on an 8% SDS-PAGE gel and were detected by western blotting using anti-FLAG M5 (Sigma) or anti-Ub P4D1 (Santa Cruz) antibodies or by coomassie blue staining.

LC-MS/MS analysis. High molecular weight, Rsp5p modified Mga2p was excised from SDS-PAGE gels and subjected to in-gel tryptic digestion as previously described. Digestions were carried out for 8 h at 37oC with sequencing grade trypsin (Promega, Madison, WI). Peptides from each gel band were extracted, brought to a final volume of 10 uL with 5% formic acid, and analyzed by LC-MS/MS. An Agilent 1100 series HPLC system (Agilent Technologies, Palo Alto, CA) was used to load peptides at 2 uL/min onto a precolumn packed with 5 um YMC

ODS-A C18 beads (Waters, Milford, MA). Following a desalting step, the flow was split and peptides were eluted through a second column packed with the same beads at approximately 200 nL/min using a 5-80% gradient of acetonitrile with 0.1% formic acid for 1 h. The LC effluent was electrosprayed into the sampling orifice of a QSTAR Pulsar quadrupole-TOF mass spectrometer (ABI/MDS Sciex, Concord, ON). MS/MS data was then analyzed and matched to yeast protein sequences in the NCBI database using the Mascot search engine. Peptide and MS/MS mass tolerances were set at ±100 ppm and 0.2 Da, respectively. Carbamidomethyl was set as a fixed modification and oxidation and ubiquitination were set as variable modifications. MS/MS spectra of all ubiquitinated peptides were manually verified to confirm internal lysine residues with GG or LRGG moieties.

Microsome-based mobilization assay. Microsomal fractions were prepared as described above and placed in reactions containing 25mM Tris-HCl (pH 7.5), 75 mM NaCl, 5 mM $MgCl_2$, 3 mM ATP, 1 mM creatine phosphate, 0.05 units/μL creatine kinase, 20 μM DTT, 0.5M MG115, protease inhibitor cocktail and 300 ng of recombinant Cdc48p or $Cdc48p^{Npl4p\text{-}Ufd1p}$ complex. Reactions were incubated at 27°C for one hour and were separated into membrane and soluble fractions by ultracentrifugation 50,000 g for 30 minutes at 4°C. Proteins from soluble and membrane fractions were resolved by SDS-PAGE and western blotting was carried out with anti-FLAG antibody.

Immunofluorescence. Cells expressing epitope-tagged Mga2p were fixed with 3.7% formaldehyde and washed with 0.1M potassium phosphate buffer, pH 7.5. These cells were treated with 50 units of lyticase (Sigma) for one hour at 30°C to generate spheroplasts, washed with PBS and placed on polylysine-coated glass slides. Cells were permeabilized with 0.1%

Triton X-100/PBS for 5 minutes. Anti-FLAG antibodies or Anti-Kar2p antibodies were added and incubations were carried out at room temperature for one hr. FITC conjugated anti-rabbit IgG or alexa fluor 647 conjugated anti-mouse antibodies were added and incubated for an additional 30 min. Mounting media was placed on cells and visualized by microscopy.

## 5.5 Results and discussion

Previous studies have identified the Ub ligase and ligase-binding site on Mga2p120 that play a role in Mga2p90 release.[12,14] However, the amino acids of Mga2p120 that are targets of ubiquitination and mediate mobilization have yet to be identified. We employed mass spectrometry analysis with *in vitro* ubiquitinated product to identify lysine residues of Mga2p120 that are ubiquitinated by Rsp5p. To generate the substrate, microsomes were first prepared from yeast harboring an amino-terminal FLAG-tagged MGA2 expression construct. Mga2p120–Mga2p90 complexes were immunopurified from microsomes using an agarose-conjugated anti-FLAG antibody. Bead-bound complexes were then placed in ubiquitination reactions containing E1, E2, and Rsp5p. After termination of these reactions, complexes were washed, resuspended in SDS-PAGE loading buffer, boiled, and resolved by SDS-PAGE. After the gel was stained with Coomassie blue (Figure 5.1), a region containing high molecular weight proteins was excised and subjected to in-gel tryptic digestion and the resulting peptides were analyzed by liquid chromatography–tandem mass spectrometry (LC–MS/MS). The analysis successfully identified a peptide matching Mga2p120 with two Ub-conjugation sites at lysine residues 983 and 985 (Figure 5.1). This ubiquitinated peptide was assigned a Mascot score of 53, which indicated a high-confidence identification.[15]

Based on this information, site-directed mutagenesis was performed to generate an mga2 expression construct encoding FLAG-tagged protein harboring arginine substitutions at residues 983 and 985 (termed Mga2p-2K). Mga2p120–Mga2p90 complexes were purified from cells expressing this mutant and subjected to *in vitro* ubiquitination by Rsp5p. In initial experiments, as well as the one presented in Figure 5.2, mutation of these two amino acids leads to suppression of Rsp5p-induced Mga2p120 ubiquitination. This is evidenced by a reduction in high molecular weight proteins measured by Western blotting using anti-FLAG or anti-Ub antibodies. However, high molecular weight Mga2p120 products were evident in reactions harboring this mutant when Western blots were probed with the more sensitive anti-Ub antibody, suggesting the presence of additional Rsp5p-induced Ub-conjugation sites.

Considering that Mag2p harbors another lysine residue at 980 that is in close proximity to 983 and 985 and a peptide fragment containing this amino acid was not covered by the LC–MS/MS analysis, we generated an mga2 construct encoding anMga2p120 mutant containing arginine substitutions at 980, 983, and 985 (termed Mga2p-3K). Complexes were again purified from cells harboring this triple-lysine mutant as well as those from yeast harboring mga2-2k or mga2Δlpky. mga2Δlpky encodes a mutant protein lacking the Rsp5p binding site14 on Mga2p120 and serves as an important control for this and subsequent experiments. Mga2p120–Mga2p90 complexes were then placed in ubiquitination reactions with varying amounts of Rsp5p. As shown in Figure 5.2, the Mga2p-3K and Mga2pΔLPKY mutants are more resistant to Rsp5p-mediated ubiquitination when compared to Mga2p-2K. It also deserves to be mentioned that mutation of any one of these amino acids individually has no detectable affect on Rsp5p ubiquitination (data not shown). These results suggest that lysine residues 980, 983, and 985 can

serve as potential acceptor sites for Rsp5pinduced Mga2p120 ubiquitination *in vitro* and that mutation of at least two of these sites are needed to suppress Rsp5p-induced ubiquitination.

To determine if mutation of these sites affects Mga2p120 ubiquitination in cells, we measured the amount of Ub-modified protein directly after immunopurification of wild-type and mutant proteins with anti-FLAG antibody. Since the amount of ubiquitinated Mga2p120 represents only a small fraction of total Mga2p120 after immunopurification, a greater amount of cells was used for immunopurification. Also, we analyzed increased amounts of immunopurified protein when compared to the inputs for the Rsp5p-dependent ubiquitination reactions presented in Figure 5.1. Consistent with the *in vitro* ubiquitination data, the Mga2p120 mutant harboring mutations in 983 and 985 displayed reduced ubiquitination (Figure 5.2). This figure also shows that elimination of all three sites results in a further reduction in the amount of ubiquitinated Mga2p120 and the amount of Ub-modified Mga2p120 immunoprecipitated from cells harboring the mga2-3k mutant was similar to that immunopurified from cells harboring the Rsp5p-binding-deficientmga2Δlpky mutant. These results suggest that lysine residues 980, 983, and 985 can also serve as targets of the ubiquitination machinery *in vivo*.

One trivial explanation for the observed results is that the engineered mutations abrogate an Mga2p120–Rsp5p interaction. To investigate this possibility, we performed *in vitro* binding assays with recombinant glutathione S-transferase (GST)-tagged Rsp5p and immunopurified FLAG-tagged Mga2p120–Mga2p90 complexes. As shown in the right panel of Figure 5.2, Rsp5p interacts with Mga2p120–Mga2p90 complexes immunopurified from cells harboringMGA2 or mga2-3k, but not mga2Δlpky. Consistent with this *in vitro* binding data, we

also found that Rsp5p-bound Mga2p120–Mga2p complex in cells containing mga2-3k as assessed by co-immunoprecipitation analysis (Figure 5.2).

Rsp5p promotes mobilization of Mga2p90 from the ER membrane by promoting polyubiquitination of its interacting membrane-bound Mga2p120 anchor. This Rsp5p-dependent activity is dependent on $Cdc48p^{Npl4p/Ufd1p}$, which recognizes the polyubiquitinated signal on Mga2p120 and facilitates segregation of Mga2p90 from Mga2p120.[12] Thus, we next wanted to determine how mutation of the identified carboxy-terminal-localized Mga2p120 conjugation sites affects $Cdc48p^{Npl4p/Ufd1p}$-dependent Mga2p90 mobilization *in vitro*. For these and subsequent experiments, we used only the Mga2p-3K Rsp5pUb-conjugation-defective mutant. Microsomes were prepared from cells expressing amino-terminal FLAG-tagged Mga2p products.

Microsomes were placed into reactions containing ATP, an ATP regeneration system, and recombinant Cdc48p or preformed recombinant $Cdc48p^{Npl4p/Ufd1p}$ complex (left panel of Figure 5.3). The ATP-depleting enzyme apyrase was also included in a set of reactions containing $Cdc48p^{Npl4p/Ufd1p}$ to verify that mobilization is ATP dependent. Reactions were subjected to ultracentrifugation, and the amounts of Mga2p90 and Mga2p120 in the soluble and membrane fractions were measured by Western blotting using anti-FLAG antibody. As shown in the right panel of Figure 5.3, the greatest amount of soluble Mga2p90 was detected in reactions containing microsomes prepared from cells harboringMGA2 and recombinant $Cdc48p^{Npl4p/Ufd1p}$ complex. Only background Mga2p90 release was detected in the other reactions, including those containing $Cdc48p^{Npl4p/Ufd1p}$ and microsomes derived from yeast expressing the Rsp5p binding deficient mutant Mga2pΔLPKY or the Rsp5p conjugation mutant Mga2p-3K (Figure 5.3).

In addition to this *in vitro* approach, we also determined the localization of Mga2p in cells containing MGA2, mga2Δlpky, and mga2-3k by indirect immunofluorescence using the anti-FLAG antibody, which recognizes both Mga2p120 and Mga2p90. As shown in Figure 5.3, Mga2p displays a mix of perinuclear and nuclear staining in cells harboring MGA2. In contrast, cells expressing the Rsp5p binding or lysine conjugation mutants show anti-FLAG positive "half-moon structures," which is a marker of Mga2p90 ER sequestration that has been reported previously.[7,8] The ER sequestration is also confirmed by co-localization with the ER-localized protein Kar2p (Figure 5.3). These experiments suggest that these lysine residues on membrane-bound Mga2p120 mediate the release of the tethered Mga2p90 product.

We next evaluated the biological consequences of mutating these residues on Mga2p90 transcriptional regulation function *in vivo*. We first generated low- copy MGA2, mga2Δlpky, and mga2-3k expression constructs under the control of the native MGA2 promoter. We also generated expression constructs that harbored single mutations at each of the lysine residues. These constructs, alone with an empty vector control, were transformed into mga2Δspt23ts cells. These cells are not viable at elevated temperatures due to lack of both Mga2p and Spt23p, which perform redundant functions inducing the expression of the essential gene, OLE1.5 Their growth defects can be suppressed, however, by the addition of oleic acid to the growth media.5 After transformation with the indicated constructs, cells were grown at 25°C and then placed at the non-permissive temperature of 34°C. Cells were harvested, RNA was prepared, and RNA levels of the Mga2p90 target gene OLE1 were measured by Northern blotting. As shown in Figure 5.4, cells harboring mga2Δlpky or mga2-3k express lower OLE1 RNA amounts when compared to those with MGA2. OLE1 levels are, however, higher in mga2Δlpky or mga2-3k cells when

compared to the vector alone control (Figure 5.4). Cells containing the single-lysine mga2 mutants did not display reduced OLE1 levels.

We also assessed the proliferative capacity of mga2Δspt23ts cells transformed with these constructs when grown at non-permissive temperature, both in the absence and presence of oleic acid. As shown in Figure 5.4, cells harboring the mga2Δlpky and mga2-3k constructs do not proliferate to the same extent as those containing MGA2, mga2-k980r, mga2-k983r, or mga2-k985r in the absence of oleic acid, but do grow better than cells containing the empty vector control. There was no noticeable difference in the growth of any of these transformants when cells were plated on media containing oleic acid (Figure 5.4). These results suggest that elimination of the Rsp5p-binding and Ub-conjugation sites on Mga2p120 suppresses, but does not eliminate, its OLE1-inducing function when cells are grown at elevated temperature.

The conjugation of Ub to lysine residues on a substrate by the ubiquitination machinery is generally promiscuous in nature. This lack of Ub conjugation specificity has made it very difficult to identify specific residues on proteins that mediate Ub-controlled signaling events. Taking Rsp5p as an example, we have identified numerous substrates for this ligase.[16,17] However, Ub-conjugation sites for the majority of Rsp5p substrates have yet to be identified. We set out to identify lysine residues of Mga2p120 that are targets for Rsp5p-induced ubiquitination and provide evidence here that three lysine residues, 980, 983, and 985, all located carboxy-terminal to the LPKY motif, can serve as sites for Rsp5p-induced Ub attachment both *in vitro* and *in vivo*. Moreover, mutation of these residues abrogates $Cdc48p^{Npl4p/Ufd1p}$-mediated mobilization of Mga2p90 from microsomes *in vitro* and ER sequestration of Mga2p90 in cells. Considering that these residues are contained within the membrane bound Mga2p120 anchor and

not in tethered Mga2p90, these findings further support the model[12] where $Cdc48p^{Npl4p/Ufd1p}$-mediated segregation of Mga2p120–Mga2p90 complexes is mediated by a Ub signal present on Mga2p120. Although all three sites can serve as conjugation sites, it remains unclear if there is a preference for a specific residue or residues under conditions where all three sites are available.

As stated in the introduction, SPT23 is a homologue of MGA2, and their encoded protein p90 products undergo Rsp5p-dependent release and possess redundant OLE1 transactivation function.[5] We have also attempted to identify Rsp5p-directed Ub-conjugation sites on Spt23p120. Unfortunately, using a similar approach described here for Mga2p120, we have not been able to identify Rsp5p-induced Ub-conjugation sites on Spt23p120. Moreover, we have mutated 11 lysine residues (alone and in combination) in the carboxy-terminal region of Spt23p120 extending from amino acid 900 to the transmembrane domain and have seen no effect on Spt23p120 ubiquitination, processing, or mobilization (S.B., N.S., and D.S.H., unpublished data). These lysine residues include five that are amino-terminal (901, 903, 908, 914, and 917) and six that are carboxy-terminal (947, 948, 950, 957, 996, and 997) to the Rsp5p-binding LPKY site (938–941). It is conceivable that the Rsp5p-induced Ub signal is placed on lysine residues present in a different locale and there are important functional consequences for this differential placement. Alternatively, the location of Ub conjugation on Spt23p120 may not be as stringent when compared to Mga2p120 and this is preventing us from detecting the true ligation sites, which are contained within the carboxy-terminus of Spt23p120.

## 5.6 Conclusion

Although Rsp5p is largely dispensable for Mga2p120 processing, we have previously shown that it mediates release of Mga2p90 from the ER.[12,14] We were thus somewhat surprised by the rescue data presented here, as we were expecting that both the Rsp5p-binding and Ub-conjugation mutants would be unable to rescue the growth defects of mga2Δspt23ts. While mga2Δspt23ts cells transformed with mga2Δlpky and mga2-3k do not grow as well when compared to those containing MGA2, they proliferate better than the vector-alone control. It is possible, based on these results, that multiple ligases promote Mga2p90 release and Rsp5p plays a contributing rather than an obligatory role. Consistent with this idea, we have found that despite displaying reduced ubiquitination, Ub-modified forms of Mga2pΔLPKY and Mga2p-3k can be detected after immunopurification of these mutant proteins from cells (Figure 5.2 and our unpublished overexposed blots). It is also interesting to note that deletion of the Rsp5p binding and conjugation sites on Mga2p also has a slight effect on Mga2p processing, as the ratio of Mga2p90 to Mga2p120 is reduced in cells harboring these mutants. Thus, it is possible that Rsp5p and yet to be identified ligase(s) play roles in providing both the processing and release signals. However, in the absence of knowing ligases that perform such functions, we cannot rule out the possibility that Rsp5p-independent Mga2p processing and release is mediated by signals other than ubiquitination. Notwithstanding, it will be interesting to elucidate Rsp5p-independent processes that contribute to Mga2p120 processing and/or inducing release of processed Mga2p90 from the ER.

**Figure 5.1 Mga2p120 is ubiquitinated on lysine 983 and 985 by Rsp5p.** (a) Microsomes from cells expressing FLAG-tagged Mga2p120–Mga2p90 complex were collected and the proteins from the microsomes were eluted and subjected to immunoprecipitation with anti-FLAG antibodies. *In vitro* ubiquitination assays were performed using the immunoprecipitated complexes and recombinant E1, E2, and E3 (Rsp5p) as described in the supplemental information. Proteins were separated by SDS-PAGE and the gel was stained with Coomassie blue. (b) Model depicting Mga2p120 (unprocessed membrane-bound form) and Mga2p90 (processed membrane-tethered form) with noted IPT domain, two ankyrin (Anks) repeats, and transmembrane domain. Also shown is the Rsp5p interaction motif LPKY, the ubiquitinated lysine residues that were identified (983 and 985, blue) by MS/MS analysis, and the lysine residue in proximity to the identified residues (980, green) that was not covered by the MS/MS analysis.

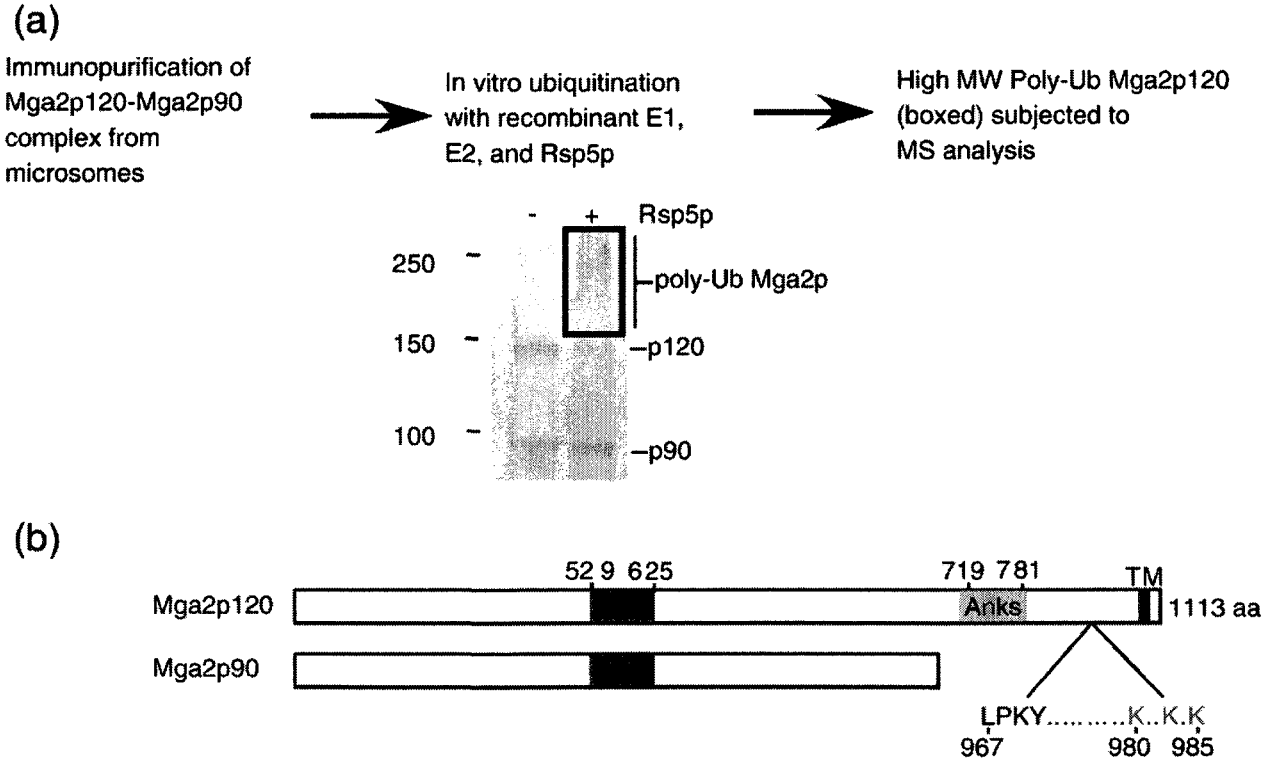


**Figure 5.2 Mutations at lysine 980, 983, and 985 of Mga2p120 abolish Rsp5p-mediated polyubiquitination.** (a) Mga2p120–Mga2p90 complexes were immunopurified from cells harboring galactose-inducible FLAGMGA2, FLAGmga2Δlpky, FLAGmga2-2k, or FLAGMga2-3k as described in the supplemental information. The immunoprecipitated proteins were subjected to an *in vitro* ubiquitination assay using recombinant Rsp5p. Modified proteins were resolved by SDS-PAGE and transferred to a nitrocellulose membrane, and Western blotting was performed with anti-Ub and anti-FLAG antibodies. (b) Lysates from cells harboring the indicated expression constructs were subjected to immunoprecipitation with anti-FLAG antibodies. The immunoprecipitated proteins were separated by SDS-PAGE and transferred to a nitrocellulose membrane, and proteins were detected by Western blotting with anti-Ub and anti-FLAG antibodies. (c) Immunopurified Mga2p120–Mga2p90 complexes from cells harboring FLAGMGA2, FLAGmga2Δlpky, FLAGmga2-2k, or FLAGMga2-3k were liberated from the beads with FLAG peptide and used in *in vitro* pull-down assay with immobilized GSTRsp5p. Bound material was resolved by SDS-PAGE, transferred to a nitrocellulose membrane, and detected by Western blotting with anti-FLAG antibodies. The amount of GST-Rsp5p used in the pull-down assays was determined by Ponceau S staining of the blot prior to transfer. The left panel shows the input for each Mga2p protein and the asterisk denotes an alternatively processed Mga2p product. (d) Cells expressing HA-tagged Rsp5p and/or FLAG-tagged Mga2p proteins were lysed, and immunoprecipitations were performed with anti-FLAG antibody. Immunoprecipitated proteins were resolved by SDS-PAGE and transferred to nitrocellulose membranes, and the membranes were probed with either anti-HA or anti-FLAG antibodies. The left panel shows input levels of Mga2p and Rsp5p proteins in the extracts as determined by Western blotting.

**Figure 5.2 continued**

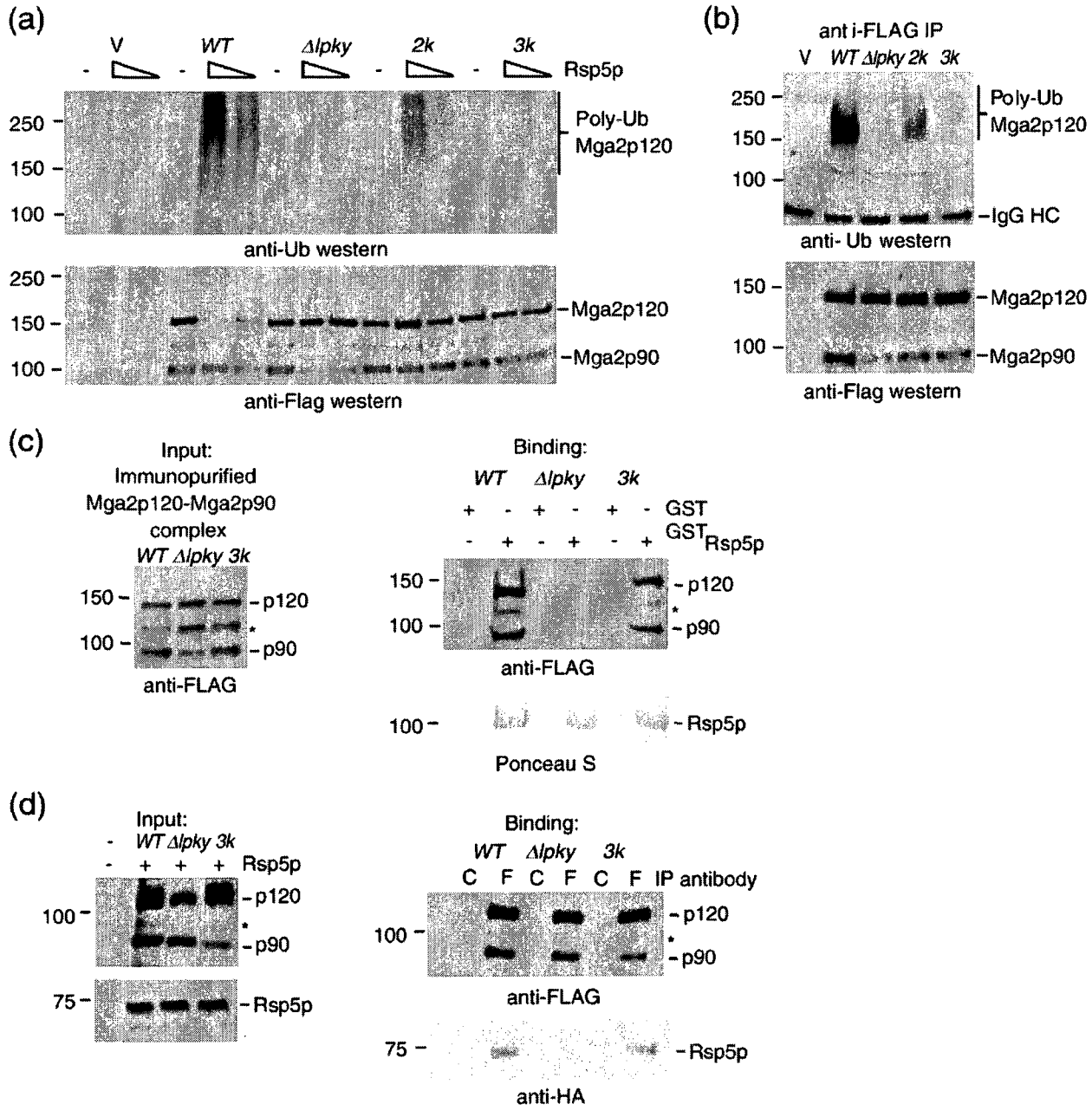


**Figure 5.3 Lysine 980, 983, and 985 of Mga2p120 mediate mobilization of Mga2p90 from membranes *in vitro* and *in vivo*.** (a) Left panel: Recombinant Cdc48p and $Cdc48p^{Npl4p/Ufd1p}$ complex was prepared, resolved by SDS-PAGE gel, and visualized by Coomassie blue staining to assess purity as described in the supplemental information. Right panel: Microsomes were isolated from cells containing FLAGMGA2, FLAGmga2pΔlpky, or FLAGmga2p-3k, and *in vitro* mobilization assays were performed as described previously12 with BSA, Cdc48p, or $Cdc48p^{Npl4p/Ufd1p}$ complex with or without apyrase. Following completion of the assay, the soluble and insoluble fractions were separated by ultracentrifugation and amounts of Mga2p products in the two fractions were determined by Western blotting with anti-FLAG antibodies. (b) Cells harboring FLAGMGA2, FLAGmga2pΔlpky, or FLAGmga2p-3k were fixed and localization of Mga2p in these cells was determined by indirect immunofluorescence using anti-FLAG mouse monoclonal antibody as described in the supplemental information. Cells were also co-stained with rabbit anti-Kar2p polyclonal antibodies. FLAG-tagged proteins were visualized by an Alexa Fluor 647 (red)-conjugated anti-mouse secondary antibody, while Kar2p was visualized by fluorescein isothiocyanate (green)-conjugated anti-rabbit secondary antibody. N, nuclear.

**Figure 5.3 continued**

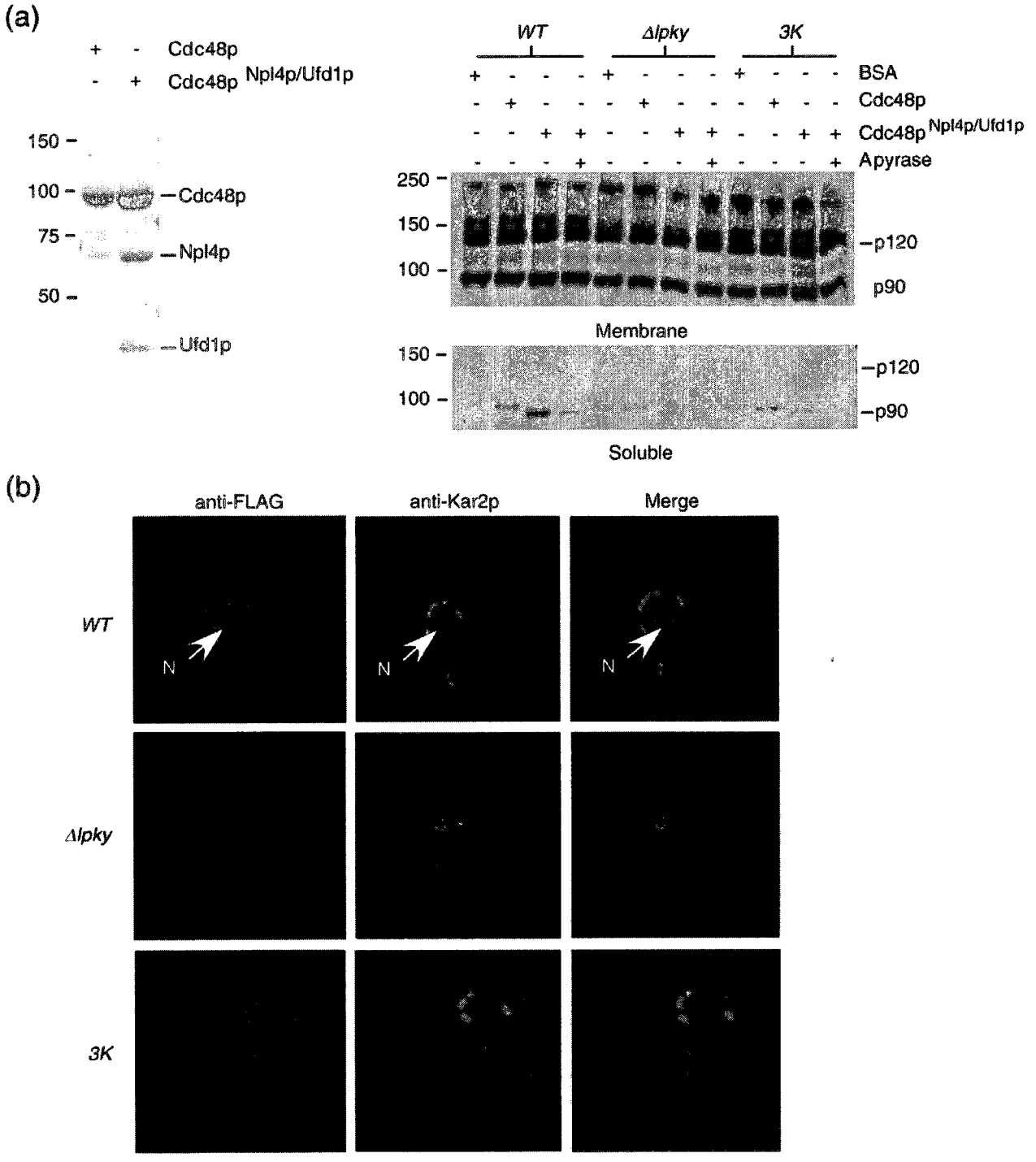
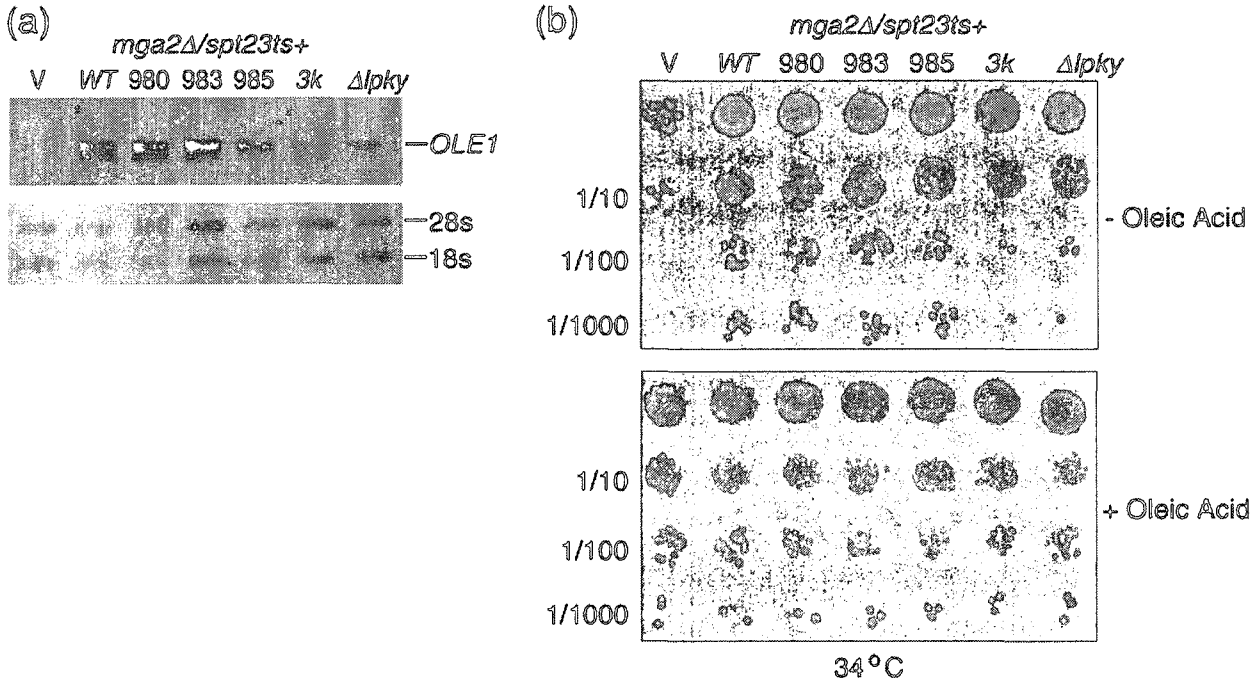


Figure 5.4 Lysine 980, 983, and 985 of Mga2p120 negatively affect its function *in vivo*. (a) mga2Δ spt23-ts cells were transformed with plasmids harboring MGA2, mga2-k980r, mga2-k983r, or mga2-k985r, mga2Δlpky, and mga2-3k. Cells were grown at 25 °C and then at 34 °C for 16 h. Cells were harvested, RNAwas prepared, and Northern blotting was carried out with radiolabeled OLE1. RNA loading was assessed by ethidium bromide staining of gel prior to transfer; levels of 28S and 18S ribosomal RNA are depicted in the bottom panel. (b) Cells transformed with the indicated constructs were grown at 25°C, serially diluted, and plated on synthetic dropout media with or without oleic acid. Plates were incubated at 34°C for 2–3 days.



## 5.7 References

(1) Hochstrasser, M. (1996). Ubiquitin-dependent protein degradation. Annu. Rev. Genet. 30, 405–439.

(2) Hershko, A. & Ciechanover, A. (1998). The ubiquitin system. Annu. Rev. Biochem. 67, 425–479.

(3) Chau, V., Tobias, J. W., Bachmair, A., Marriott, D., Ecker, D. J., Gonda, D. K. & Varshavsky, A. (1989). A multiubiquitin chain is confined to specific lysine in a targeted short-lived protein. Science, 243, 1576–1583.

(4) Haglund, K. & Dikic, I. (2005). Ubiquitylation and cell signaling. EMBO J. 24, 3353–3359.

(5) Zhang, S., Skalsky, Y. & Garfinkel, D. J. (1999). MGA2 or SPT23 is required for transcription of the delta9 fatty acid desaturase gene, OLE1, and nuclear membrane integrity in Saccharomyces cerevisiae. Genetics, 151, 473–483.

(6) Rape, M., Hoppe, T., Gorr, I., Kalocay, M., Richly, H. & Jentsch, S. (2001). Mobilization of processed, membrane-tethered SPT23 transcription factor by CDC48 (UFD1/NPL4), a ubiquitin-selective chaperone. Cell, 107, 667–677.

(7) Shcherbik, N., Zoladek, T., Nickels, J. T. & Haines, D. S. (2003). Rsp5p is required for ER bound Mga2p120 polyubiquitination and release of the processed/ tethered transactivator Mga2p90. Curr. Biol. 13, 1227–1233.

(8) Hoppe, T., Matuschewski, K., Rape, M., Schlenker, S., Ulrich, H. D. & Jentsch, S. (2000). Activation of a membrane-bound transcription factor by regulated ubiquitin/proteasome-dependent processing. Cell, 102, 577–586.

(9) Piwko, W. & Jentsch, S. (2006). Proteasome-mediated protein processing by bidirectional degradation initiated from an internal site. Nat. Struct. Mol. Biol. 13, 691–697.

(10) Schuberth, C. & Buchberger, A. (2008). UBX domain proteins: major regulators of the AAA ATPase Cdc48/p97. Cell. Mol. Life Sci. 65, 2360–2371.

(11) Halawani, D. & Latterich, M. (2006). p97: The cell's molecular purgatory? Mol. Cell, 22, 713–717.

(12) Shcherbik, N. & Haines, D. S. (2007). Cdc48p (Npl4p/Ufd1p) binds and segregates membraneanchored/tethered complexes via a polyubiquitin signal present on the anchors. Mol. Cell, 25, 385–397.

(13) Auld, K. L., Brown, C. R., Casolari, J. M., Komili, S. & Silver, P. A. (2006). Genomic association of the proteasome demonstrates overlapping gene regulatory activity with transcription factor substrates. Mol. Cell, 21, 861–871.

(14) Shcherbik, N., Kee, Y., Lyon, N., Huibregtse, J. M. & Haines, D. S. (2004). A single PXY motif located within the carboxyl terminus of Spt23p and Mga2p mediates a physical and functional interaction with ubiquitin ligase Rsp5p. J. Biol. Chem. 279, 53892–53898.

(15) Vasilescu, J., Smith, J. C., Zweitzig, D. R., Denis, N. J., Haines, D. S. & Figeys, D. (2008). The proteomic reactor facilitates the analysis of affinity-purified proteins by mass spectrometry: application for identifying ubiquitinated proteins in human cells. J. Mass Spectrom. 43, 296–304.

(16) Gupta, R., Kus, B., Fladd, C., Wasmuth, J., Tonikian, R., Sidhu, S. et al. (2007). Ubiquitination screen using protein microarrays for comprehensive identification of Rsp5 substrates in yeast. Mol. Syst. Biol. 3, 116.

(17) Kus, B., Gajadhar, A., Stanger, K., Cho, R., Sun, W., Rouleau, N. et al. (2005). A high throughput screen to identify substrates for the ubiquitin ligase Rsp5. J. Biol. Chem. 280, 29470–29478.

# Chapter 6: Differential proteomic screen to evidence proteins ubiquitinated upon mitotic exit in cell-free extract of *Xenopus laevis* embryos

## 6.1 Manuscript status and statement of author contributions

A version of this chapter has been published in the Journal of Proteome Research: *Bazile F, Gagne JP, Mercier G, Lo KS, Pascal A, Vasilescu J, Figeys D, Poirier GG, Kubiak JZ, Chesnel F. Differential proteomic screen to evidence proteins ubiquitinated upon mitotic exit in cell-free extract of Xenopus laevis embryo. Journal of Proteome Research. Vol. 7: 4701-4714, 2008.* FB, JPG, GM conceived, designed and performed most of the experiments. GM and KSL performed the mass spectrometry analysis. JV performed the data analysis to map protein ubiquitination sites. DF, JZK and FC edited the manuscript. FB wrote the manuscript.

## 6.2 Summary

Post-translational modification of proteins via ubiquitination plays a crucial role in numerous functions of the cell. Polyubiquitination is one of the key regulatory processes involved in the regulation of mitotic progression. Here we describe a differential proteomic screen dedicated to the identification of novel proteins ubiquitinated upon mitotic exit in cell-free extract of *Xenopus laevis* embryos. Mutated recombinant 6×His-tagged ubiquitin ($Ubi^{K48R}$) was added to mitotic extract from which we purified conjugated proteins, as well as associated proteins in nondenaturing conditions by cobalt affinity chromatography. Proteins eluted from $Ubi^{K48R}$ supplemented and control extracts were compared by LC-MS/MS analysis after

monodimensional SDS-PAGE. A total of 144 proteins potentially ubiquitinated or associated with them were identified. Forty-one percent of these proteins were shown to be involved in ubiquitination and/or proteasomal degradation pathway confirming the specificity of the screen. Twelve proteins, among them ubiquitin itself, were shown to carry a "GG" or "LRGG" remnant tag indicating their direct ubiquitination. Interestingly, sequence analysis of ubiquitinated substrates carrying these tags indicated that in *Xenopus* cell-free embryo extract supplemented with $Ubi^{K48R}$, the majority of polyubiquitination occurred through lysine-11 specific ubiquitin chain polymerization. The potential interest in this atypical form of ubiquitination as well as usefulness of our method in analyzing atypical polyubiquitin species is discussed.

## 6.3 Introduction

Ubiquitination is a post-translational protein modification involved in the regulation of different cell functions. The precise regulation of ubiquitination is correlated to the number of ubiquitin moieties involved in the modification. Polyubiquitination determines the half-life of modified proteins by targeting them to finely controlled proteolysis via the proteasome pathway, while monoubiquitination is known for its non-proteolytic signaling functions.[1] Ubiquitination is carried out by a series of enzymes (usually three; E1 activating, E2 conjugating and E3 ligating enzymes). The specificity of polyubiquitinated substrates relies largely on the third enzyme, the E3 ubiquitin ligase.[2]

Polyubiquitination is a key regulator of mitotic progression. Numerous E3 ligases are involved in polyubiquitination of proteins involved in regulation of the M-phase of the cell cycle;

however, a particular E3 ligase APC/C (Anaphase Promoting Complex/Cyclosome) plays a crucial role. The tuning of APC/C specificity is mediated via two different and exclusive regulatory subunits associating with the core complex: Cdc20 or Cdh1. APC/C-dependent polyubiquitination mediates the sequential degradation of mitotic proteins responsible for the correct sequence of events during the M-phase.[3,4]

Numerous substrates of polyubiquitination were identified during mitosis. Nek2A and cyclin A are the early substrates degraded in prophase.[5] Cyclins B1-B5 are degraded sequentially upon the metaphase/anaphase transition, which assures inactivation of MPF (M-phase Promoting Factor) and therefore M-phase exit.[6,7] Similarly, polyubiquitination and subsequent degradation of securin enables separation of sister chromatids and anaphase movement during the same period of M-phase.[8,9] Later substrates, such as Tome-1 or Aurora A, are degraded after the M-phase exit and throughout the G1 phase.[10,11] Polyubiquitination and degradation of Cdc20, the regulatory subunit of APC/C, is of special importance in M-phase regulation since it enables the middle-M-phase switch from Cdc20-APC/C- to Cdh1-APC/C-dependent targeting of proteins to be degraded.[12] Some substrates, such as cyclin B, are degraded first via Cdc20-APC/C and later via Cdh1-APC/C ubiquitin ligase-dependent pathways.[13] Therefore, polyubiquitination not only regulates the sequence of degradation of mitotic proteins, but also autoregulates the switch between two regulatory pathways.

The number of mitotic substrates of polyubiquitination being identified in different species and cellular contexts is still growing. There are likely other proteins behaving in this way that are yet to be identified. In addition, some substrates may differ between cell types and species. Recent data show that monoubiquitinated proteins, like histones (e.g. H2A), could also

be involved in regulation of mitotic progression.[14] Extensive identification of novel poly- and mono-ubiquitinated mitotic proteins could therefore reveal new mechanisms involved in the regulation of mitotic progression and enable a better understanding of mitotic regulation.

*Xenopus laevis* embryo cell-free extract is a model of choice to study the regulatory pathways of mitotic progression because of the perfect synchrony of a large quantity of cytoplasm accessible for biochemical analysis. In addition, embryonic mitosis, whose biochemical events can be restored in the cell-free cytoplasmic extract, is thought to be a simplified version of somatic mitosis. For example, the checkpoint controls are not executed during embryonic mitoses in *Xenopus* despite the presence of the necessary molecular machinery.[15] The mitotic substrates that are polyubiquitinated also differ in embryonic and somatic cells. Among the substrates of APC/C, the Nek2A protein was shown to be absent during early embryonic mitoses,[16] while Aurora A, although present, is not degraded because of the absence of APC/C regulator Cdh1.[17,18] Another fundamental feature of the cell-free *Xenopus* embryo extract is that proteasome inhibition during embryonic M-phase does not prevent MPF inactivation,[7] while it perfectly arrests somatic cells[19] or mouse oocytes[9] in M-phase. This could be linked to the potential presence of different substrates necessary to be degraded upstream from cyclin B in these three cell types.

The recent development of proteome analysis has provided a powerful tool especially suitable for the study of ubiquitination.[20-22] Subproteomics approaches using affinity probes that target ubiquitinated proteins have been used successfully in various models.[23-26] Optimized versions of this approach combine the screening of mono- and poly-ubiquitinated proteins to the identification of the modification site on the substrate's lysine residues.[27,28] Taking advantage of

a targeted proteomic approach, we performed a large-scale proteome analysis of ubiquitin-protein conjugates during the first embryonic M-phase in *Xenopus laevis*. Using cobalt-affinity chromatography, we designed a proteomic screen capable of identifying novel ubiquitin substrates. An exogenously expressed epitope tagged His(6×)-ubiquitin bearing a mutation at the 48$^{th}$ amino acid residue (Ubi$^{K48R}$) was used to limit ubiquitin chain assembly in mitotic cell-free extracts.[29] Although other lysine residues within ubiquitin can be involved in ubiquitin conjugation, K48 has been shown to be the most prominent conjugation site in the models studied so far.[30-34] Notably, Chau and collaborators demonstrated that Ubi$^{K48R}$ causes a 10-fold decrease in substrate degradation rates compared to wild-type ubiquitin.[30] This mutant protein dramatically slows down cyclin B degradation and MPF inactivation upon mitotic exit in cell-free *Xenopus* embryo extracts.[29] In addition, MPF stability increases when MG132 is present in the extract in addition to Ubi$^{K48R}$.

The goal of this manipulation in our system was to accumulate short-chain His(6×)-ubiquitinated target proteins upon prolonged mitotic exit and to isolate them from the extract. Proteins eluted from the cobalt affinity column were resolved by SDS-PAGE, in-gel trypsin digested and identified by liquid chromatography-tandem mass spectrometry (LC-MS/MS). Spectral data was searched for the characteristic ion fragmentation pattern corresponding to an ubiquitin signature, which is inferred from the presence of an ubiquitin remnant tag (GG or LRGG) generated from the tryptic digestion of a covalently linked ubiquitin moiety. This procedure enabled us to identify candidates for novel proteins potentially ubiquitinated during mitotic exit in *Xenopus* embryo cell-free extract.

## 6.4 Materials and methods

Frogs. *Xenopus laevis* females were purchased from NASCO (Fort Atkinson, WI). Drugs. Proteasome inhibitor MG132 was purchased from Biomol (Plymouth Meeting, PA). Other chemicals were obtained either from Sigma or MP Biochemicals unless otherwise stated.

Egg Collection and Activation. Females were injected subcutaneously with human chorionic gonadotropin (500-600 IU per female; Organon, Puteaux, France) and kept overnight at 21°C in 110 mM NaCl. Unfertilized eggs collected from "overnight lay" were dejellied with 2% L-cysteine pH 7.8 in XB buffer (100 mM KCl, 1 mM $MgCl_2$, 50 μM $CaCl_2$, 10 mM HEPES, 50 mM sucrose, pH 7.6), washed in XB, treated for 1.5 min with 0.5 μg/mL calcium ionophore A23187 and then extensively washed in XB. Activated eggs were then incubated in XB at 21°C.

Cell-Free Extracts. Cytoplasmic extracts from calcium ionophore-activated embryos before the first embryonic mitosis were prepared according to a classical protocol of Murray[35] modified as previously described.[36,37] Briefly, embryos were cultured at 21°C in XB for 60 min post activation. They were transferred into appropriate tubes (5 mL ultraclear centrifuge tubes; Beckman Coulter, Roissy, France) containing 0.5 mL of XB supplemented with 0.1 mM AEBSF, aprotinin, leupeptin, pepstatin, chymostatin (10 μg/mL each) and 25 μg/mL cytochalasin D and packed through a short spin at 100g. After removal of any excess XB medium, embryos were subjected to two consecutive centrifugations: a crushing spin, 10 000g for 10 min at 4°C and a clarification spin of the supernatant 10 000g for 10 min at 4°C in which cytochalasin D, AEBSF, aprotinin, leupeptin, pepstatin and chymostatin were again added. The resulting low-speed supernatants were then reincubated at 21°C for 60 min and aliquots were taken out and either

frozen in liquid nitrogen and stored at -70 °C (for cobalt resin chromatography) or mixed with sample buffer,[38] heated at 85°C for 5 min, and stored at -20°C (for Western blot analyses).

Metal Ion Affinity Chromatography. Samples of 200 μL of extract were diluted in NP40 buffer (20 mM Tris-HCl, pH 7.5, 150 mM NaCl, 10% glycerol, 50 mM NaF, 50 mM β-glycerophosphate, 0.2% NP40) supplemented with 0.5 mM sodium orthovanadate, 0.5 mM AEBSF and 10 μg/μL leupeptin, aprotinin and pepstatin, and centrifuged through a 30-kDa filter (Nanosep 30K OMEGA) at 10 000g and 4°C for 60 min. A second filtration was performed with filter loaded with 400 μL of NP40 buffer. The solution remaining on the filter was diluted 50 times with NP40 buffer. The mixture was then incubated overnight at 4°C with 30 μL of equilibrated cobalt resin (TALON Metal Affinity Resin; Clontech Laboratories, Inc. Mountain View, CA). Following centrifugation at 10 000g at 4°C for 20 s to pellet the resin, it was washed five times with the washing buffer (20 mM Tris-HCl, pH 7.5, 150 mM NaCl, 10% glycerol, 50 mM NaF, 50 mM β-glycerophosphate, 0.2% NP40, 20 mM imidazole). The last wash was performed with a buffer only containing 20 mM Tris-HCl, pH 7.5, and 150 mM NaCl. Then the resin was eluted with 300 μL of imidazole buffer (20 mM Tris-HCl, pH 7.5, 150 mM NaCl, 10% glycerol, 50 mM NaF, 50 mM β-glycerophosphate, 0.2% NP40, 500 mM imidazole). The eluted fraction was concentrated on 10 kDa filter (Nanosep 10K OMEGA) by a 8000g centrifugation at 4°C to obtain a final volume of 30 μL. Then, 15 μL of loading buffer[38] was added and the mixture was heated for 5 min at 85°C.

Sample Preparation for MS Analysis. Protein samples were resolved on a 4-12% Bis-Tris Criterion XT Precast gradient gel (Bio-Rad) with XT MOPS running buffer. The gel was then stained with SYPRO Ruby fluorescent protein stain (Bio-Rad, Hercules, CA) according to the

manufacturer's instructions. Image acquisition was made using a CCD-based Chemi-Imager 4000 imaging system and AlphaEase software 3.3 (Alpha Innotech Corporation, San Leandro, CA). Protein lanes were cut into 80 gel slices using a disposable lanepicker (The Gel Company, San Francisco, CA). Gel slices were pooled two by two into 96-well plates. In-gel protein digests were performed on a MassPREP liquid handling station (Waters, Milford, MA) according to the manufacturer's specifications and using sequencing-grade modified trypsin (Promega, Madison, WI). Peptide extracts were dried out using a SpeedVac.

LC-MS/MS Method. The samples were analyzed using a NanoLC 1100 series system (Agilent, Palo Alto, CA) coupled with a hybrid quadrupole-TOF mass spectrometer (QSTAR xl, MDS Sciex, Concord, Canada) equipped with a nanoelectrospray ionization source (Proxeon, Odense, Denmark) fitted with a 10-μm-i.d. fused-silica tip (FS360-20-10-D, New Objective, Woburn, MA). The dried samples were resuspended in 12 μL of a 0.1% trifluoroacetic acid (TFA) in water solution containing 20 fmol/μL of two synthetic peptides, used as internal standards to monitor the chromatographic and detection performances. Five microliters were loaded on a 300 μm i.d. × 5 mm Zorbax trapping column (300SB-C18, 5 μm, Agilent). Chromatographic separation was achieved on a 75 μm i.d. × 10 cm Biobasic C18 Integrafrit column (IFC75-BI-10, New Objective) with a linear gradient of 13-32% B (0.1% formic acid in acetonitrile) against A (0.1% formic acid in water) in 43 min, at 250 nL/min. The total run time was 70 min. MS data was acquired automatically using Analyst QS 1.1 software (MDS SCIEX). An information dependent acquisition (IDA) method was used to fragment the three most intense peaks, which were then put on an exclusion list for 4 min.

Protein Databases. A FASTA compatible *Xenopus* protein database (43 835 entries) has been generated from the Uniref100 protein sequence database available on the Universal Protein Resource Web site.[39] A Perl (version 5.8.8) script was used to retrieve and format all sequences related to *Xenopus* genus (NCBI taxonomy database ID:8353). A general database (All taxa) that contains nonidentical sequences from GenBank, CDS translations, PDB (Protein Data Bank), Swiss-Prot, PIR (Protein Information Resource), and PRF (Protein Research Foundation) has been adapted from the NCBI (National Center for Biotechnology Information) nr database (nonredundant).

Database Searching. Tandem mass spectra were extracted by Mascot.dll version 1.6b16 (Applied Biosystems). Charge state deconvolution and deisotoping process were not performed. All MS/MS samples were analyzed using Mascot (Matrix Science, London, U.K.; version 2.1.04) and X!Tandem (www-.thegpm.org; version 2006.04.01.2). X!Tandem and Mascot were set up to search the *Xenopus* or the All taxa databases. Mascot and X!Tandem were searched with a fragment ion mass tolerance of 0.20 Da and a parent ion tolerance of 0.20 Da. Iodoacetamide derivative of cysteine was specified in Mascot and X!Tandem as a fixed modification. Deamidation of asparagines and glutamine, methylation and dimethylation of lysine and arginine, oxidation of methionine, and acetylation of lysine and the N-terminus were specified in Mascot and X!Tandem as variable modifications. Scaffold (version 01_06_05, Proteome Software, Inc., Portland, OR) was used to validate MS/MS based peptide and protein identifications. Peptide identifications were accepted if they could be established at greater than 80.0% probability as specified by the Peptide Prophet algorithm.[40] Protein identifications were accepted if they could be established at greater than 90.0% probability and contained at least 2 identified peptides. Protein probabilities were assigned by the Protein Prophet algorithm.[41]

Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone were grouped to satisfy the principles of parsimony.

Identification of Ubiquitination Sites. Several studies were conducted to attempt the precise identification of polyubiquitination sites within proteins by tandem mass spectrometry. However, a re-evaluation of the ion score cutoffs to achieve acceptable false positive rates for protein identifications performed by Vasilescu and colleagues[42] have tempered the rather impressive rate of ubiquitination site mapping reported in some studies. They found that, although these cutoff scores were generally useful for achieving low false positive rates for protein identifications, they may not be suitable for mapping post-translational ubiquitination modifications. Moreover, the choice of high accuracy and resolution mass spectrometers such as the QSTAR that was used for the present study is critical to achieve acceptable false positive rates. Therefore, the search for ubiquitination sites was based on this recent study that described the use of a parsing tool to extract signature peptides. Briefly, MS/MS data was searched against the *Xenopus* protein database using the Mascot search engine. Carbamidomethylation of cysteine residues was selected as a fixed modification. GG and LRGG ubiquitination signatures of lysine and oxidation of methionine residues were selected as variable modifications. Peptide mass tolerances were set to 100 ppm and fragment ion mass tolerances were set to 0.2 Da. Mascot outputs were formatted and parsed using the reported parser criteria for acceptable ubiquitination sites.[42] Signature peptides were then manually validated to ensure that all ubiquitination sites were mapped to internal lysine residues.

## 6.5 Results and discussion

Fine Control of Mitotic Progression and Cyclin B Behavior. We have previously shown that 6×His-tagged $Ubi^{K48R}$ substantially slows down cyclin B degradation, MCM4 protein dephosphorylation and histone H1 kinase inactivation in mitotic cell-free embryo extract, therefore delaying M-phase exit.[29] Additional inhibition of the proteolytic activity of the proteasome with MG132 amplifies this delay.[29] We used the latter experimental system to search for other proteins that could behave similarly to cyclin B subsequent to 6×His-$Ubi^{K48R}$ conjugation. First, we analyzed whether cyclin B2 is indeed polyubiquitinated in the presence of both 6×His-$Ubi^{K48R}$ and MG132 in the extract. The samples collected during the slowed down mitotic exit were incubated with cobalt resin and the eluates were immunoblotted, revealed with anticyclin B2 antibody and compared to the sample taken at the beginning of the incubation. Cyclin B2 is visible as an up-shifted phosphorylated band at 45 kDa and two discrete higher molecular weight additional bands corresponding to the molecular weight of mono- and di-ubiquitinated forms of this protein (Figure 6.1).

Similar samples were collected from mitotic extract after incubation of control and 6×His-$Ubi^{K48R}$ and MG132-supplemented extracts after 50 min and the evolution of mitotic marker proteins (MCM4 phosphorylation status and cyclin B2 abundance) was followed by Western blotting (Figure 6.2). Sibling samples were thereafter subjected to cobalt affinity chromatography. 6×His-$Ubi^{K48R}$ pulled-down cobalt-associated proteins were processed for proteomic analysis (see Materials and Methods). Sypro Ruby staining of the gel shows clear quantitative and qualitative differences in the control and experimental samples (Figure 6.3). This suggested that 6×His-$Ubi^{K48R}$ pull-down actually enables efficient isolation of

polyubiquitinated proteins and their associated partners since the pull-down was performed in native conditions. Differential LC-MS/MS analysis of in-gel digested bands excised from the gel enabled the identification of 144 Xenopus proteins specifically bound to the resin (Figure 6.4). "Gene Ontology" (GO) analysis of the protein data set enabled us to classify these proteins according to their established or putative functions (Figure 6.4, Table 6.1).

The Differentially Identified Candidate Ubiquitinated Proteins. The two major categories correspond either to proteins of the ubiquitin/proteasome pathway (59 out of 144: 41%) or to ribosomal subunits (30 out of 144: 20.8%). Among the proteins belonging to the ubiquitin cycle we found ubiquitin itself (Table 6.1; protein no. 15), the ubiquitin-activating enzyme E1 (proteins nos. 1 or 2), ubiquitin-conjugating enzymes E2s (proteins nos. 36, 59,...), ubiquitin ligases E3s (proteins nos. 4, 12,...) as well as subunits of the 26S proteasome and interestingly Rpn10/S5A (protein no. 23). The latter was shown before to bind K48-linked polyubiquitinated substrates.[43] Ubiquitinyl hydrolases (proteins nos. 5, 105,...) were also purified from the mitotic extract as well as carrier proteins like VCP/p97 (protein no. 75) or Rad23 (protein no. 72) that interact with ubiquitinated substrates and target them to the proteasome for degradation.[44,45] The remaining 38% of purified proteins are either involved in DNA/RNA metabolism, signal transduction, cell cycle and division or in a variety of other functions ("other": Figure 6.4).

In the context of the present study aimed at identifying new embryonic mitotic players, the proteins assigned with a role in (or related to) cell cycle and division are of particular interest and represent 8.3% (12 out of 144) of the purified proteins. Some have already been described as mono/polyubiquitinated (PCNA; Table 6.1; no. 126)[46] or polyubiquitinated and degraded by the proteasome during mitotic exit (plk1; Table 6.1; no. 126)[47]. Others, like Npl4 (Table 6.1; no. 63)

and VCP/p97 (Table 6.1; no. 75) have been reported during M-phase to associate together in order to present ubiquitinated mitotic substrates to the 26S proteasome for degradation.[48] In the latter case, we should have expected to identify Ufd1 as well since Npl4 requires Ufd1 to associate with VCP/p97.[49-51] As Ufd1 complete sequence is present in *Xenopus laevis* databases (MGC68571), its absence in our screen may result from a loss during the purification procedure or an insufficient recovery.

Other proteins known to be involved in mitotic regulation (e.g., CDK11, MCM7A, PP2A and Orc3-like) were neither described as ubiquitinated nor associated with ubiquitinated ones so far. We did not find any mitotic cyclin (A or B) that are expected to be sequentially ubiquitinated and proteolyzed during M-phase, but purified their common partner, cdc2a (Table 6.1; no. 92). Cyclin partners of cdc2a are polyubiquitinated to different degree (e.g. Figure 6.1 for cyclin B2) and thus migrate in the gel as dispersed bands contrarily to the enzymatic component of the complex of cyclins A, B1, B2, B3, B4 which migrates as a single, more abundant band which facilitates identification. Finally, few cell-cycle related proteins identified here such as CDK11 (Table 6.1; no. 47), MCM7A (no. 68), PP2A (no. 124) or orc3-like p81 (no. 143) need to be further studied in order to know whether they are directly ubiquitinated or associated with proteins of the ubiquitin/proteasome machinery.

At first glance, the identification of such a low number of known cell cycle-involved candidates can be surprising. However, ubiquitination is a very dynamic process involving both E3 ubiquitin ligases and deubiquitinases (DUBs) and M-phase players that have been already evidenced as ubiquitinated are usually modified during a very narrow time window. Even the addition of MG132 in our assay to limit degradation of the ubiquitinated substrates probably did

not prevent DUB action limiting the level of ubiquitinated substrates. CDK11, which appears in our data set, is of particular interest. The 58-kDa isoform (C-terminus part of p110 CDK11) has been recently shown to be specifically expressed during M-phase in mammalian somatic cells and is required for its completion.[52-54] Its regulation during mitosis, however, has not yet been defined. As we identified only C-terminal peptides, it seems that only the short, p58 CDK11 form is expressed during embryonic M-phase in *Xenopus laevis*. It is tempting to speculate that down-regulation of its expression upon mitotic exit observed in human somatic cells (Regis Giet, personal communication) may occur through ubiquitination and subsequent degradation by the proteasome.

The list of 144 proteins we present here contains, however, both ubiquitinated proteins and members of complexes associated with them, as the screen was performed in non-denaturing conditions. Not all of them may be specifically ubiquitinated during mitosis since we are unable to distinguish such proteins from those which are modified via ubiquitination constitutively or in a cell cycle independent manner, for example, meiosis-specific substrates polyubiquitinated and degraded gradually following oocyte activation. At least 6 oocyte and/or embryo specific proteins were purified. Among them, zyg-11 seems an interesting candidate. It is a substrate recognition subunit within the Cullin2-based E3 ubiquitin ligase[55] which allows meiotic anaphase II progression in *Caenorhabditis elegans*.[56] Other such candidates are zona pellucida proteins (ZP A-C) that have been suggested in other animal models to be ubiquitinated upon fertilization.[57,58] We found, however, no evidence in the literature for the ubiquitination of two other oocyte proteins identified in our screen, namely, CPSF-5 (cleavage and polyadenylation specificity factor; Table 6.1: protein no. 98) and EP45 precursor (protein no. 62). It is likely that

they belong to a group of oocyte-specific proteins that might be degraded following the entry into the embryonic development.

Specificity of *Xenopus laevis* Protein Identifications. The successful use of tandem mass spectrometry (MS/MS) to annotate complex proteomics samples predominantly depends on whether the proteins exist in the database that is being searched. Being conscious that the *Xenopus laevis* protein database might be incomplete because of partial genome coverage, MS/MS data was run against a nonredundant protein database (NCBI nr) which is a composite of several protein databases where entries with absolutely identical sequences have been merged (3 525 863 entries). The idea behind this approach is to identify Xenopus homologous proteins in other taxa that could have been missed from the original database because of partial proteome coverage (Table 6.2). As expected, evolutionary conserved proteins known to be associated with the ubiquitin machinery are found in other organisms, but the fairly limited number (30; Table 6.2) of non-Xenopus proteins identified using the All taxa database indicates that the Xenopus database is satisfactory for the identification of a wide range of Xenopus specific proteins.

Identification of Ubiquitinated Proteins via Detection of Ubiquitin Moieties. In contrast to protein identification probabilities that can be modeled from a standard statistical distribution of correct and incorrect matches within a large data set like we performed with Scaffold, the precise identification of ubiquitination sites within proteins requires an alternative approach. Scaffold calculates protein probabilities from peptide probabilities. Peptide probabilities are combined to estimate the protein probability, even if these probabilities are low. This allows for a more complete use of the data set. However, an important factor to determine protein probabilities is the prevalence of multiple peptides hits for a protein identification. In our data

set, most of the proteins are identified by multiple peptides, so the Scaffold algorithm downgrades one-hit protein identifications. In a shotgun proteomics approach, the peptide coverage for identified proteins is low,[59] so the probability of identifying a peptide that bears a ubiquitination signature is even lower. The probability of identifying a protein with multiple peptides in which one of these will be a polyubiquitinated peptide is very unlikely. Thus, Scaffold is most appropriate to identify statistically relevant proteins in a large data set but too stringent to account for unlikely events, such as the identification of a one-hit peptide that harbors a ubiquitination signature. A Scaffold graph of protein probability calculation based on the peptides identified by LC-MS/MS is presented in Figure 6.5. Scaffold is very conservative, so we can see that, even if one-hit peptides are identified with high probabilities, the corresponding protein probabilities will not meet our 90% protein probability cutoff. Although they may have acceptable quality spectra, these one-hit proteins are excluded from the protein identification data set presented in Table 6.1.

To directly ascertain the ubiquitination of lysine residues within all identified peptides, we searched for peptide sequences bearing "GG" and "LRGG" moieties indicating unequivocally ubiquitination sites[21] using a recently reported parser tool.[42] We found 13 modified peptides that could be assigned to 12 proteins, including ubiquitin itself (Table 6.3). The majority of ubiquitin peptides carried such moieties at lysine 11 (TLTGKTITLEVEPSDTIENVK), while the remaining ones were modified at lysine 48 (LIFAGKQLEDGR) (see Table 6.3 and Figure 6.6A,B). We identified a single low ion score peptide with a LRGG moeity at lysine 29 (AKIQDK), but we excluded it after manual inspection. Interestingly, an ubiquitin signature at lysine 63 was not identified. This result suggests an unexpected hierarchy of frequency of lysines participating in polyubiquitin in our

experimental conditions in contrast to previous data obtained in different species and cells.[30] However, new data shows potential roles for K11-mediated ubiquitination, suggesting that this kind of atypical ubiquitination may indeed play some specific functions. Kirkpatrick and colleagues have shown that the major mitotic E3 ligase, APC/C, can *in vitro* polyubiquitinate cyclin B1 through K11-, K48- and K63-linked chains.[60] Recently, Jin and collaborators have demonstrated that human APC/C, when associated with the ubiquitin-conjugating enzyme UbcH10, targets substrates for proteasomal degradation preferentially through K11-linked polyubiquitination.[61] It seems therefore that, in Xenopus embryo cell-free mitotic extracts, K63-linked polyubiquitin chain elongation could be a very minor event. In contrast, K11-linked polyubiquitination is either a major process in this context, or is favored by the addition of an excess of free 6×His-Ubi$^{K48R}$. In both cases, these observations strengthen the conclusions of Jin and collaborators about the physiological importance of K11-linked polyubiquitination which was so far underestimated.[61] Alternatively, we cannot totally rule out that canonical ubiquitination sites (K48 and K63) can be more efficiently deubiquitinated by endogenous deubiquitinases in the extract.

Among the proteins identified as carriers of GG and LRGG motifs of ubiquitin, besides ubiquitin itself (endogenous or K48R exogenous mutant), we found two ubiquitin-conjugating enzymes E2: ubiquitin carrier protein (Table 6.3; no. Q7SZ88) and E2D2 (Ubc4/5 homologue; no. P62840). These enzymes are known to carry a polyubiquitin chain on a conserved catalytic cysteine that is then transferred by an E3 ligase to a substrate. However, Ravid and Hochstrasser recently reported that these conjugating enzymes can be targeted to proteasomal degradation under the cysteine-linked autoubiquitinated form or after the polyubiquitin chain has been transferred to a lysine side chain.[62] We found Xenopus ubiquitin carrier protein to be

ubiquitinated on lysine-92 within the $IC_{87}LDILK_{92}DK$ peptide nearby the catalytic cysteine-87, which suggests the possible auto-ubiquitination of this enzyme as demonstrated by Ravid and Hochstrasser.[62] In contrast, the ubiquitination site of E2D2 identified by us was K128 (Table 6.3 and Figure 6.6C), whereas the conserved catalytic cysteine is at position 85. But in this case as in few E2s, there is no lysine in position +5 from the cysteine. It is therefore necessary to determine whether ubiquitination on lysine-128 also occurs via autocatalysis (by the transfer of the chain from the cysteine-85; *in silico* structure analysis of XlE2D2 with Swiss PDBviewer suggests that it is physically possible) and whether it is physiologically relevant during mitosis.

None of the remaining proteins displayed in Table 6.3 have been reported to be ubiquitinated. However, some of them deserve to be further analyzed as they belong to the ubiquitin machinery or may be implicated in M-phase regulation. The chaperonin TCP-1 is known as a component of a complex associated with the COP-1 E3 ligase,[63] XRP2 is a GTPase activating protein for Arf-like 3 (Arl3), the knock-down of which inhibits cytokinesis,[64] Xenopus putative SSRM1 (no. Q2VPI0) which is described in many species as a splicing coactivator would also participate in sister chromatid cohesion as it physically and functionally interacts with cohesin in *C. elegans*.[65] Understanding the nature (mono- or poly-ubiquitination) and the eventual mitotic role of this modification on these proteins is an important issue.

Finally, the presence of organ-specific developmental proteins that have not been shown to be expressed during oocyte maturation or early embryonic stages illustrates some limitations that might be encountered when dealing with low identification scores (Table 6.3: no. Q6GQD9: mab21-like 1; no.Q2VPL1: Usher syndrome 1C protein homologue; no. Q5EC17: ventricular myosin heavy chain (vMHC)). Although poorly characterized, these proteins may be false

positive identifications. This is particularly likely for vMHC which is not expressed before formation of heart chambers during Xenopus organogenesis.[66]

Limits and Further Usefulness of the Screen. Some proteins effectively ubiquitinated during mitosis may lack in this list because of the stringency of our screen. Proteins with naturally high affinity to cobalt were discarded from the list since they were also present in high quantity in the control sample. Also quantitatively minor ubiquitinated substrates, which could play a crucial role in mitosis, may have been omitted in our screen because of their poor representation and low number of identified peptides. In addition, the list does not contain nuclear DNA-associated proteins since the nuclear fraction is discarded upon cell-free extract production. Identification of substrates of ubiquitination among M-phase regulating proteins is of great importance for a better understanding of mitotic regulation. Knowledge of the molecules and pathways involved in M-phase progression is the first step to design novel anticancer drugs and therapies. Proteasome inhibitors have proven to be potent anticancer drugs.[67] Ubiquitinated substrates may constitute even more specific therapeutic targets. Eventual identification of these proteins in our list will be possible after validation of the chosen candidates and analysis of their type of ubiquitination and their role in M-phase regulation.

## 6.6 Conclusion

In conclusion, our screen enabled us to present a list of 144 candidate proteins which are either ubiquitinated or associated with complexes in which some members are ubiquitinated during mitotic exit. We evidenced already known cell-cycle as well as oocyte/embryo-specific proteins whose ubiquitination relevance and role need to be elucidated. Detection of "GG" and "LRGG" motifs shows that (i) at least 9 of the candidates are effectively ubiquitinated and (ii) K11- and K48-linked polyubiquitination predominate in *Xenopus laevis* cell-free extract upon mitotic exit in our experimental conditions.

**Figure 6.1 Cyclin B2 is mono- and di-ubiquitinated in mitotic *Xenopus laevis* cell-free extract in the presence of 6×His-Ubi$^{K48R}$ and MG132.**

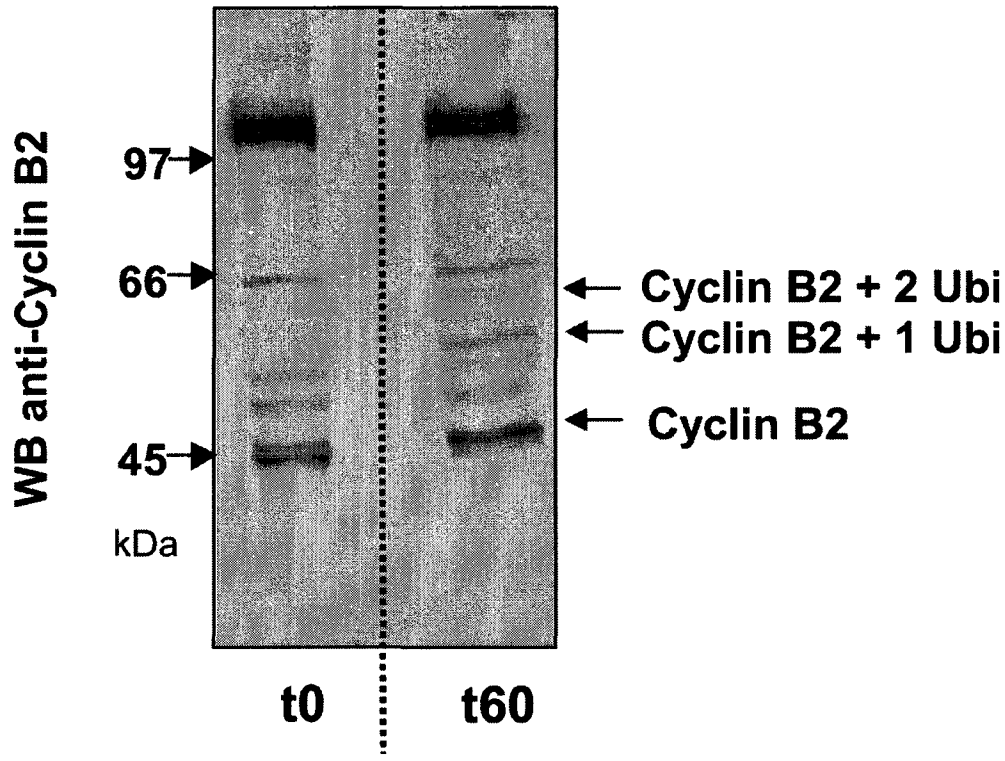


**Figure 6.2 MCM4 status and cyclin B2 abundance during the M-phase progression in the cell-free extract.** Note that cyclin B2 Western blot in the bottom panel (t0 and t60) corresponds to the input of the WB shown in Figure 6.1.

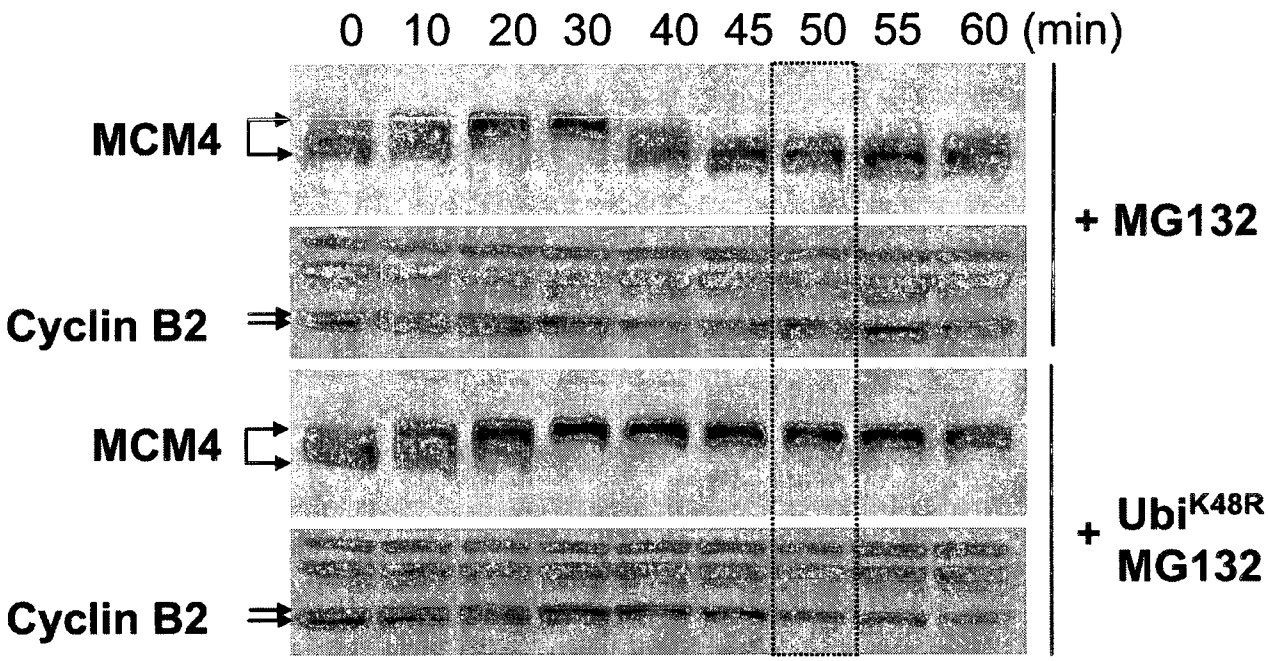


**Figure 6.3 Comparative protein profile from cobalt affinity chromatography of Ubi$^{K48R}$-containing cell-free *Xenopus* embryonic mitotic extracts versus control extracts.** Eluted proteins were resolved on SDS-PAGE and revealed with SYPRO Ruby protein gel stain.

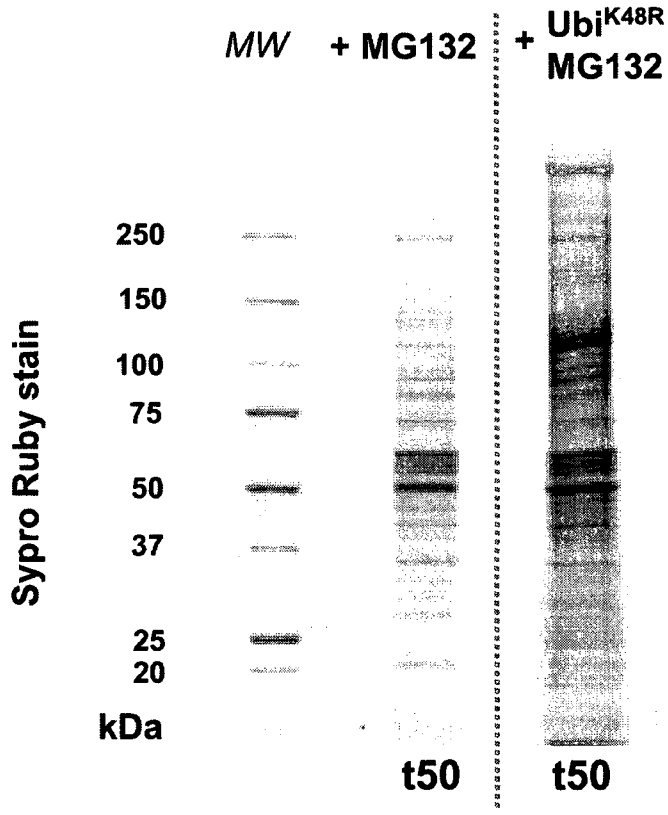


Figure 6.4 Distribution of protein functions based on Gene Ontology (GO) classification of putative polyubiquitinated proteins and ubiquitin-associated protein.

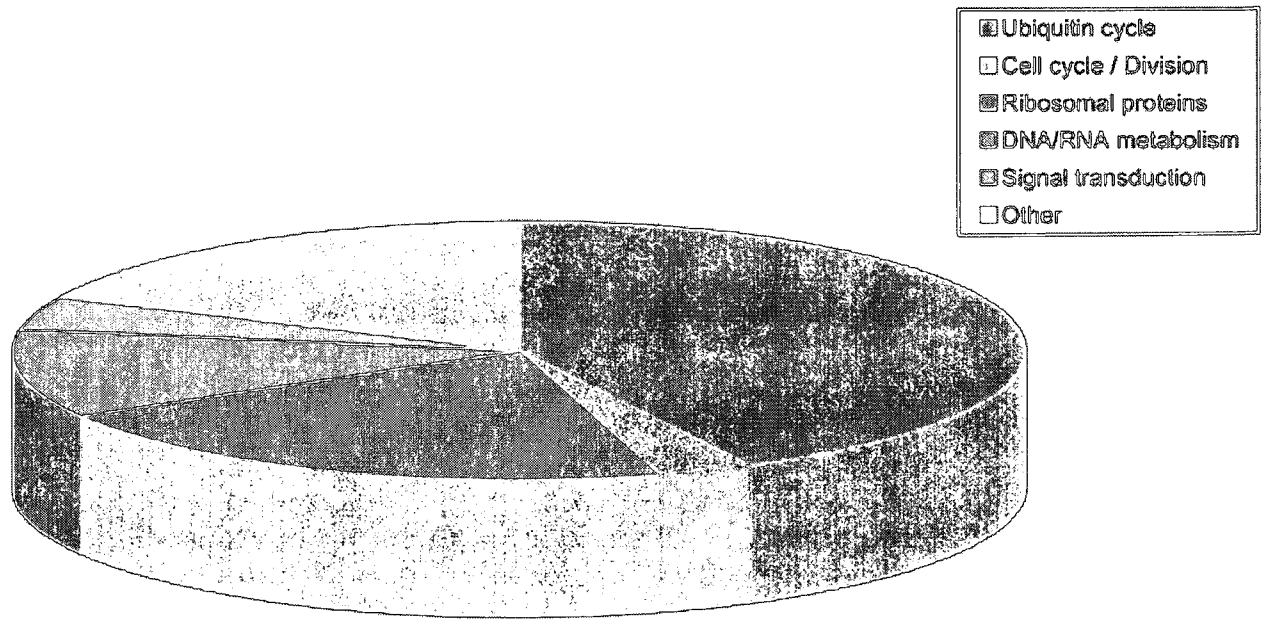


**Figure 6.5 Protein probabilities derived from peptides as calculated by Scaffold.** The graph shows for each sample how the peptide probabilities are combined to estimate the protein probability.

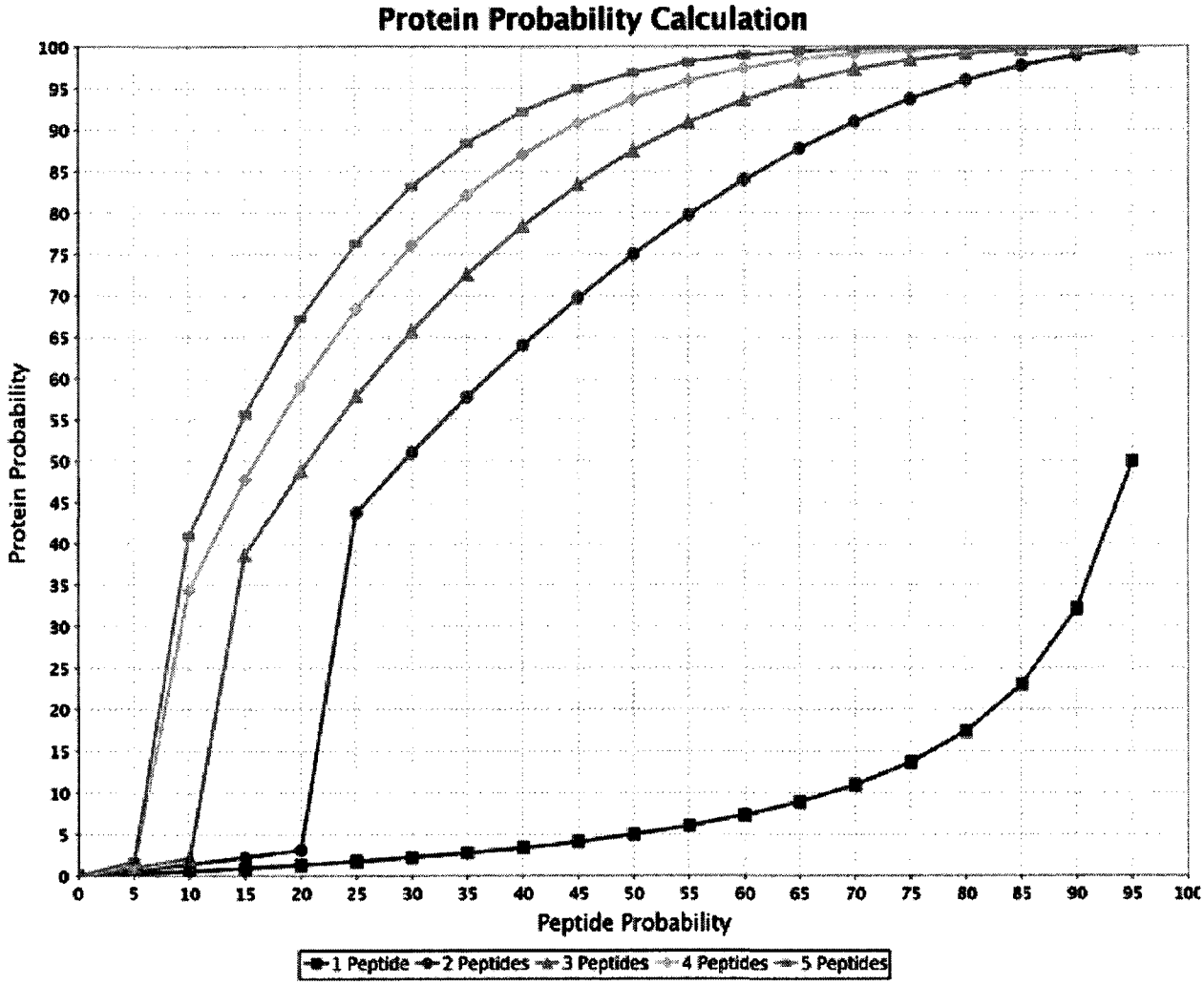
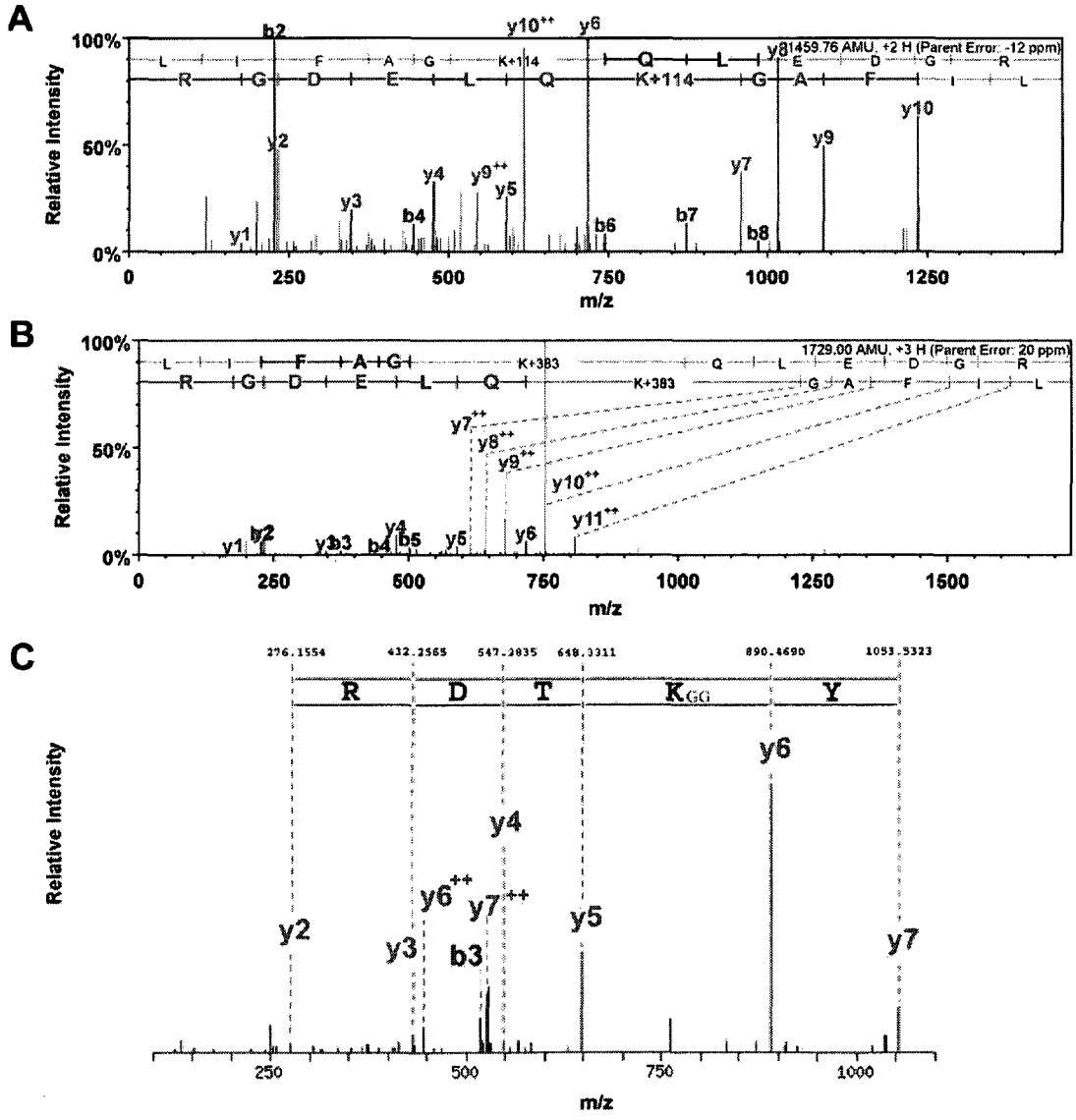


**Figure 6.6 Polyubiquitination signatures on peptides identified by MS/MS analysis from affinity-purified Ubi$^{K48R}$ supplemented cell-free extracts.** Ubiquitination can be identified in MS/MS spectra by two signatures, GG and LRGG. For some peptides, only one signature is observed; for other peptides, both signatures are observed. Modification of polyubiquitin on K48 is shown here with a GG (K+114 on spectrum A) and a LRGG (K+383 on spectrum B) signature for the peptide LIFAG<u>K</u>QLEDGR. Another example of GG signature is shown on peptide IY<u>K</u>TDREK which corresponds to K128 of ubiquitin-conjugating enzyme E2 D2 (spectrum C).



## Table 6.1 Ubiquitin-Conjugated Substrates Identified by Tandem Mass Spectrometry in Mitotic *Xenopus laevis* Cell-Free Extracts.

| | protein name | accession numbers[a] | unique peptides Ctrl | unique peptides $Ubi^{K48R}$ | gene ontology | function |
|---|---|---|---|---|---|---|
| 1 | Ubiquitin activating enzyme | Q9DEE8 | 0 | 48 | GO:0004839 | ubiquitin activating enzyme activity |
| 2 | Ube1-prov protein | Q801S7 | 0 | 48 | GO:0004839 | ubiquitin activating enzyme activity |
| 3 | Tcp1-A-prov protein | Q6NTQ5 | 5 | 34 | GO:0006457 | protein folding |
| 4 | Ube3a-prov protein | Q7ZTL9 | 2 | 33 | GO:0004842 | ubiquitin-protein ligase activity |
| 5 | Isopeptidase T | Q6U7I9 | 0 | 31 | GO:0004221 | ubiquitin thiolesterase activity |
| 6 | MGC114631 protein | Q498L1 | 3 | 26 | GO:0006511 | ubiquitin-dependent protein catabolic process |
| 7 | Putative uncharacterized protein (26S proteasome regulatory complex component) | Q6NTP6 | 3 | 23 | GO:0006511 | ubiquitin-dependent protein catabolic process |
| 8 | MGC79063 protein (26S proteasome regulatory complex component) | Q6DD69 | 0 | 23 | GO:0003872 | 6-phosphofructokinase activity |
| 9 | 26S protease regulatory subunit 7 | Q68ER4 | 1 | 19 | GO:0006511 | ubiquitin-dependent protein catabolic process |
| 10 | Pros26.4-prov protein | Q7SYS3 | 0 | 17 | GO:0006511 | ubiquitin-dependent protein catabolic process |
| 11 | Cullin-1 | Q1KMU1 | 0 | 17 | GO:0007049 | cell cycle |
| 12 | E3 ubiquitin-protein ligase HECTD1 | UPI000069E240 | 0 | 17 | GO:0004842 | ubiquitin-protein ligase activity |
| 13 | MGC114813 protein | Q4QR39 | 0 | 16 | GO:0006511 | ubiquitin-dependent protein catabolic process |
| 14 | Proteasome subunit alpha type | Q8AVD2 | 2 | 15 | GO:0006511 | ubiquitin-dependent protein catabolic process |
| 15 | His6-UbiK48R/UBC-prov protein | Q7SY79 | 1 | 15 | GO:0006511 | ubiquitin-dependent protein catabolic process |
| 16 | 26S protease regulatory subunit 6A-A | O42587 | 1 | 15 | GO:0006511 | ubiquitin-dependent protein catabolic process |
| 17 | 40S ribosomal protein S4 | P49401 | 2 | 14 | GO:0003735 | structural constituent of ribosome |
| 18 | MGC81323 protein (Alpha tubulin) | Q6NRV3 | 0 | 13 | GO:0005198 | structural molecule activity |
| 19 | Proteasome subunit alpha type | Q58E21 | 1 | 12 | GO:0006511 | ubiquitin-dependent protein catabolic process |
| 20 | XIZPB | Q91673 | 0 | 12 | GO:0032190 | acrosin binding |
| 21 | Rps11-prov protein | Q7SZ77 | 0 | 12 | GO:0003735 | structural constituent of ribosome |
| 22 | MGC82894 protein | Q8GNC1 | 0 | 12 | GO:0008180 | signalosome complex |
| 23 | Proteasome subunit xrpn10 | Q588Y7 | 2 | 11 | GO:0006511 | ubiquitin-dependent protein catabolic process |
| 24 | Proteasome subunit alpha type | Q3KPN6 | 1 | 11 | GO:0006511 | ubiquitin-dependent protein catabolic process |
| 25 | Psmd12 protein (PINT motif protein: Proteasome, Int-6, Nip-1 and TRIP-15) | Q8AXA3 | 0 | 11 | GO:0006511 | ubiquitin-dependent protein catabolic process |
| 26 | Proteasome subunit alpha type | Q66JJ6 | 1 | 10 | GO:0006511 | ubiquitin-dependent protein catabolic process |
| 27 | LOC495277 protein | Q5XGL4 | 1 | 9 | GO:0006511 | ubiquitin-dependent protein catabolic process |
| 28 | LOC496005 protein | Q5PQ24 | 1 | 9 | GO:0006511 | ubiquitin-dependent protein catabolic process |
| 29 | 20S proteasome alpha5 subunit | Q68A89 | 1 | 9 | GO:0006511 | ubiquitin-dependent protein catabolic process |
| 30 | XIZPC protein | Q6INF5 | 0 | 9 | GO:0032190 | acrosin binding |
| 31 | MGC82808 protein | Q6GNH1 | 0 | 9 | GO:0003735 | structural constituent of ribosome |
| 32 | P44s10-prov protein | Q7ZYE1 | 0 | 9 | GO:0006511 | ubiquitin-dependent protein catabolic process |
| 33 | Proteasome (Prosome, macropain) 26S subunit, non-ATPase, 7 | Q28EL2 | 0 | 9 | GO:0006511 | ubiquitin-dependent protein catabolic process |
| 34 | MGC97603 protein | Q5BI72 | 1 | 8 | GO:0006511 | ubiquitin-dependent protein catabolic process |
| 35 | Proteasome (Prosome macropain) subunit alpha type 2 | Q28GV0 | 1 | 8 | GO:0006511 | ubiquitin-dependent protein catabolic process |
| 36 | Hip2-prov protein | Q8AVM8 | 0 | 8 | GO:0004842 | ubiquitin-protein ligase activity |
| 37 | Proteasome subunit beta type 4 precursor | P28024 | 0 | 8 | GO:0006511 | ubiquitin-dependent protein catabolic process |
| 38 | Psmb1-prov protein | Q7ZYL1 | 0 | 8 | GO:0006511 | ubiquitin-dependent protein catabolic process |

## Table 6.1 continued

| | protein name | accession numbers[a] | unique peptides Ctrl | unique peptides $Ubi^{K.ε88}$ | gene ontology | function |
|---|---|---|---|---|---|---|
| 39 | Hypothetical protein | Q3B8B3 | 1 | 7 | GO:0006511 | ubiquitin-dependent protein catabolic process |
| 40 | Hypothetical protein | Q3KQC9 | 0 | 7 | GO:0003735 | structural constituent of ribosome |
| 41 | Nucleosome assembly protein 1 p56B | Q4U0Y5 | 0 | 7 | GO:0006334 | nucleosome assembly |
| 42 | MGC82306 protein (Ribosomal protein S13) | Q6NTT2 | 0 | 7 | GO:0003735 | structural constituent of ribosome |
| 43 | MGC85310 protein | Q66KY6 | 1 | 6 | GO:0003735 | structural constituent of ribosome |
| 44 | MGC80364 protein | Q6GQ40 | 0 | 6 | GO:0006511 | ubiquitin-dependent protein catabolic process |
| 45 | Abcf2-prov protein | Q7ZWW5 | 0 | 6 | GO:0016887 | ATPase activity |
| 46 | Hsp90beta protein | Q6AZV1 | 0 | 6 | GO:0006986 | response to unfolded protein |
| 47 | MGC80275 protein | Q6DE25 | 0 | 6 | GO:0004674 | protein serine/threonine kinase activity |
| 48 | Adenylyl cyclase-associated protein | Q6DE60 | 0 | 6 | GO:0007186 | G-protein coupled receptor protein signaling pathway |
| 49 | MGC68571 protein | Q6P704 | 0 | 6 | GO:0006511 | ubiquitin-dependent protein catabolic process |
| 50 | Ubiquitin carrier protein | Q4QR35 | 0 | 6 | GO:0006511 | ubiquitin-dependent protein catabolic process |
| 51 | Baculoviral IAP repeat-containing protein 6 | UPI00006A2EEF | 0 | 6 | GO:0004840 | ubiquitin conjugating enzyme activity |
| 52 | Proteasome subunit beta type 5 precursor | UPI00006A170A | 0 | 6 | GO:0006511 | ubiquitin-dependent protein catabolic process |
| 53 | LOC495025 protein | Q5XHE2 | 0 | 6 | GO:0006511 | ubiquitin-dependent protein catabolic process |
| 54 | MGC89305 protein | Q68ER8 | 1 | 5 | GO:0003735 | structural constituent of ribosome |
| 55 | Ribosomal protein S15 | Q5RJJ6 | 1 | 5 | GO:0003735 | structural constituent of ribosome |
| 56 | Coronin homologue | Q9PWG7 | 1 | 5 | GO:0015629 | actin cytoskeleton |
| 57 | Sdcbp-prov protein | Q6DFR3 | 1 | 5 | GO:0005515 | protein binding |
| 58 | MGC86287 protein | Q6AX35 | 0 | 5 | GO:0006511 | ubiquitin-dependent protein catabolic process |
| 59 | Ubiquitin-conjugating enzyme E2N | Q6P349 | 0 | 5 | GO:0004842 | ubiquitin-protein ligase activity |
| 60 | Proliferating cell nuclear antigen | Q6PF20 | 0 | 5 | GO:0030337 | DNA polymerase processivity factor activity |
| 61 | 60s ribosomal protien rpl12 prov protein | Q28FK6 | 0 | 5 | GO:0003735 | structural constituent of ribosome |
| 62 | Estrogen-regulated protein EP45 precursor | Q00387 | 0 | 5 | GO:0004867 | serine-type endopeptidase inhibitor activity |
| 63 | LOC494816 protein | Q63ZR4 | 0 | 5 | GO:0008270 | zinc ion binding |
| 64 | 40S ribosomal protein S3a | Q642T2 | 0 | 5 | GO:0003735 | structural constituent of ribosome |
| 65 | MGC53234 protein | Q7ZXD6 | 0 | 5 | GO:0007155 | cell adhesion |
| 66 | LOC443632 protein (Similar to eukaryotic translation initiation factor 3) | Q6GPA3 | 0 | 5 | GO:0003743 | translation initiation factor activity |
| 67 | MGC115523 protein | Q498F1 | 0 | 4 | GO:0003735 | structural constituent of ribosome |
| 68 | DNA replication licensing factor mcm7-A (CDC47p) | O42591 | 0 | 4 | GO:0030174 | regulation of DNA replication initiation |
| 69 | Proteasome subunit beta type | Q68EX4 | 0 | 4 | GO:0006511 | ubiquitin-dependent protein catabolic process |
| 70 | Rps9-prov protein (Ribosomal protein S4) | Q7ZYU4 | 0 | 4 | GO:0003735 | structural constituent of ribosome |
| 71 | Ribosomal protein S13 | Q28I74 | 0 | 4 | GO:0003735 | structural constituent of ribosome |
| 72 | MGC115064 protein | Q505M5 | 0 | 4 | GO:0006511 | ubiquitin-dependent protein catabolic process |
| 73 | MGC85348 protein | Q66KX3 | 0 | 4 | GO:0003735 | structural constituent of ribosome |
| 74 | MGC131131 protein | Q3B8H2 | 0 | 4 | GO:0006511 | ubiquitin-dependent protein catabolic process |
| 75 | Valosin-containing protein | Q6GL04 | 0 | 4 | GO:0017111 | nucleoside-triphosphatase activity |
| 76 | 40S ribosomal protein S27 | Q5D091 | 0 | 4 | GO:0003735 | structural constituent of ribosome |
| 77 | MGC82841 protein | Q6GNF4 | 0 | 4 | GO:0003735 | structural constituent of ribosome |

## Table 6.1 continued

| | protein name | accession numbers[a] | unique peptides Ctrl | unique peptides $Ubi^{K-ε-GG}$ | gene ontology | function |
|---|---|---|---|---|---|---|
| 78 | LOC485148 protein | Q5XGT1 | 0 | 4 | GO:0008180 | signalosome complex |
| 79 | Probable ATP-dependent RNA helicase DDX6 | UPI000069E2CF | 0 | 4 | GO:0003724 | RNA helicase activity |
| 80 | Canx-prov protein | Q7ZXJ1 | 0 | 4 | GO:0005509 | calcium ion binding |
| 81 | Rpl10-prov protein (Ribosomal protein L10e) | Q7ZXK4 | 0 | 4 | GO:0003735 | structural constituent of ribosome |
| 82 | MGC68500 protein (Similar to hyaluronic acid binding protein 4) | Q6PB22 | 0 | 4 | GO:0004252 | serine-type endopeptidase activity |
| 83 | Retinoblastoma-associated factor 600 | UPI00006A2087 | 0 | 3 | GO:0004842 | ubiquitin-protein ligase activity |
| 84 | Wu:fj06d02 protein | Q3B8I6 | 0 | 3 | GO:0003746 | translation elongation factor activity |
| 85 | LOC496146 protein (similar to homeobox prox 1 protein) | Q5M9B0 | 0 | 3 | GO:0003704 | specific RNA polymerase II transcription factor activity |
| 86 | Eukaryotic translation initiation factor 2B | Q28D58 | 0 | 3 | GO:0003746 | translation elongation factor activity |
| 87 | Phosphofructokinase, platelet | Q6P4Y8 | 0 | 3 | GO:0003872 | 6-phosphofructokinase activity |
| 88 | Ribosomal protein S5 | Q28EJ8 | 0 | 3 | GO:0003735 | structural constituent of ribosome |
| 89 | Hypothetical protein LOC549795 (Similar to Mitogen-activated protein kinase 12) | UPI0000509D47 | 0 | 3 | GO:0004674 | protein serine/threonine kinase activity |
| 90 | Ribosomal protein L22 | Q28IL6 | 0 | 3 | GO:0003735 | structural constituent of ribosome |
| 91 | Ubiquitin carrier protein | Q6DFK8 | 0 | 3 | GO:0006511 | ubiquitin-dependent protein catabolic process |
| 92 | Cell division control protein 2-A | P35567 | 0 | 3 | GO:0051301 | cell division |
| 93 | Putative uncharacterized protein | Q6DJQ0 | 0 | 3 | GO:0003735 | structural constituent of ribosome |
| 94 | Protein VAC14 homologue | Q68F38 | 0 | 3 | GO:0008047 | enzyme activator activity |
| 95 | Ubiquitin carrier protein | Q5I005 | 0 | 3 | GO:0006511 | ubiquitin-dependent protein catabolic process |
| 96 | Ubiquitin-conjugating enzyme X | P56616 | 0 | 3 | GO:0004842 | ubiquitin-protein ligase activity |
| 97 | Ubiquitin carrier protein | Q7ZX78 | 0 | 3 | GO:0004842 | ubiquitin-protein ligase activity |
| 98 | Cleavage and polyadenylation specific factor 5 | Q28I89 | 0 | 3 | GO:0016787 | hydrolase activity |
| 99 | UPI000069FA76 (similar to ubiquitin-activating enzyme E1) | UPI000069FA76 | 0 | 3 | GO:0004840 | ubiquitin conjugating enzyme activity |
| 100 | XIZPA protein | Q6AX05 | 0 | 3 | GO:0032190 | acrosin binding |
| 101 | Y-box-binding protein 2-A | P21574 | 0 | 3 | GO:0009386 | translational attenuation |
| 102 | Vacuolar protein sorting 28 | Q28GA6 | 0 | 3 | GO:0043162 | ubiquitin-dependent protein catabolic process via the multivesicular body pathway |
| 103 | MGC83258 protein | Q6NTI4 | 0 | 3 | GO:0004842 | ubiquitin-protein ligase activity |
| 104 | MGC85404 protein | Q66KV2 | 0 | 3 | GO:0003735 | structural constituent of ribosome |
| 105 | Ubiquitin specific protease 10 | Q6DII4 | 0 | 3 | GO:0006511 | ubiquitin-dependent protein catabolic process |
| 106 | Polyprotein | Q91674 | 0 | 3 | GO:0004252 | serine-type endopeptidase activity |
| 107 | Retinoblastoma-associated factor 600 | UPI00006A2083 | 0 | 3 | GO:0004842 | ubiquitin-protein ligase activity |
| 108 | MGC83924 protein | Q6IRM8 | 0 | 3 | GO:0003735 | structural constituent of ribosome |
| 109 | *LOC495281 protein* | Q5XG62 | 0 | 3 | GO:0004842 | ubiquitin-protein ligase activity |
| 110 | Hepatocyte growth factor-regulated tyrosine kinase substrate | Q28CS1 | 0 | 3 | GO:0006886 | intracellular protein transport |
| 111 | Probable ubiquitin carboxyl-terminal hydrolase FAF-X | UPI000069E1D9 | 0 | 2 | GO:0006511 | ubiquitin-dependent protein catabolic process |
| 112 | MGC81067 protein (Similar to hyaluronic acid binding protein 4) | Q6NRY1 | 0 | 2 | GO:0004252 | serine-type endopeptidase activity |
| 113 | THO complex subunit 2 (Tho2) | UPI000069DB7B | 0 | 2 | GO:0006368 | RNA elongation from RNA polymerase II promoter |
| 114 | MGC86445 protein (similar to transport protein Sec31) | Q66KF3 | 0 | 2 | GO:0006888 | ER to Golgi vesicle-mediated transport |

## Table 6.1 continued

| | protein name | accession numbers[a] | unique peptides Ctrl | unique peptides $Ubi^{K-\varepsilon-R}$ | gene ontology | function |
|---|---|---|---|---|---|---|
| 115 | Cytosine-specific methyltransferase | P79922 | 0 | 2 | GO:0006306 | DNA methylation |
| 116 | MGC83233 protein | Q4KLW5 | 0 | 2 | GO:0006511 | ubiquitin-dependent protein catabolic process |
| 117 | Proteasome subunit alpha type | Q5XGP0 | 0 | 2 | GO:0006511 | ubiquitin-dependent protein catabolic process |
| 118 | Proteasome (Prosome, macropain) subunit, beta type, 4 | Q5RIU8 | 0 | 2 | GO:0006511 | ubiquitin-dependent protein catabolic process |
| 119 | Nucleosome assembly protein 1-like 1 | Q6NVB1 | 0 | 2 | GO:0006334 | nucleosome assembly |
| 120 | 40S ribosomal protein S10 | Q07254 | 0 | 2 | GO:0003735 | structural constituent of ribosome |
| 121 | Ribophorin II | Q5I083 | 0 | 2 | GO:0018279 | protein amino acid N-linked glycosylation via asparagine |
| 122 | MGC82844 protein | Q6GNF2 | 0 | 2 | GO:0003735 | structural constituent of ribosome |
| 123 | Ribosomal protein L14 | Q28CV5 | 0 | 2 | GO:0003735 | structural constituent of ribosome |
| 124 | Protein phosphatase 2 (Formerly 2A), regulatory subunit B (PR 52) | Q5XGH9 | 0 | 2 | GO:0007165 | signal transduction |
| 125 | Glucose-6-phosphate dehydrogenase | Q28DI9 | 0 | 2 | GO:0006006 | glucose metabolic process |
| 126 | Serine/threonine-protein kinase PLK1 | P62205 | 0 | 2 | GO:0004674 | protein serine/threonine kinase activity |
| 127 | MGC81784 protein | Q6GMP0 | 0 | 2 | GO:0006561 | proline biosynthetic process |
| 128 | 60S ribosomal protein L10a | Q7ZYS8 | 0 | 2 | GO:0003735 | structural constituent of ribosome |
| 129 | LOC494780 protein (similar to FIP1 like 1) | Q63ZL7 | 0 | 2 | GO:0006397 | mRNA processing |
| 130 | Cullin-2 (CUL-2) | UPI000069F166 | 0 | 2 | GO:0007049 | cell cycle |
| 131 | Glyceraldehyde 3-phosphate dehydrogenase | Q7ZY52 | 0 | 2 | GO:0006096 | glycolysis |
| 132 | Hypothetical protein LOC496549 (DEAD (Asp-Glu-Ala-Asp) box polypeptide 1) | Q5XH91 | 0 | 2 | GO:0008026 | ATP-dependent helicase activity |
| 133 | MGC78885 protein | Q6AZV3 | 0 | 2 | GO:0003735 | structural constituent of ribosome |
| 134 | 40S ribosomal protein S15a. [*Xenopus tropicalis*] | UPI00004D3DA5 | 0 | 2 | GO:0003735 | structural constituent of ribosome |
| 135 | WD repeat-containing protein 82-A | Q640I6 | 0 | 2 | GO:0006511 | ubiquitin-dependent protein catabolic process |
| 136 | LOC100101273 protein | Q66KV6 | 0 | 2 | GO:0003735 | structural constituent of ribosome |
| 137 | Ubiquitin conjugating enzyme E2 | Q76EZ3 | 0 | 2 | GO:0006511 | ubiquitin-dependent protein catabolic process |
| 138 | MGC69020 protein (similar to zyg-11 homologue B (*C. elegans*)-like isoform 1) | Q6PF36 | 0 | 2 | GO:0005488 | binding |
| 139 | Proteasome subunit alpha type | Q6PE96 | 0 | 2 | GO:0006511 | ubiquitin-dependent protein catabolic process |
| 140 | MGC86316 protein | Q6AZL9 | 0 | 2 | GO:0003735 | structural constituent of ribosome |
| 141 | p53-associated parkin-like cytoplasmic protein (putative E3 ubiquitin ligase) | UPI00006A1BE7 | 0 | 2 | GO:0006511 | ubiquitin-dependent protein catabolic process |
| 142 | Proteasome subunit beta type | Q6NRC2 | 0 | 2 | GO:0006511 | ubiquitin-dependent protein catabolic process |
| 143 | Origin recognition complex associated protein p81 | O93513 | 0 | 2 | GO:0006260 | DNA replication |
| 144 | Sec23 homologue B | Q28DM5 | 0 | 2 | GO:0016192 | vesicle-mediated transport |

[a] UPI (Universal Protein Index) and Swiss-Prot accession numbers.

## Table 6.2 Assignment of Homologue Ub-Conjugated Substrates in Non-*Xenopus* Taxa.

| | protein name | accession numbers | unique peptides Ctrl | unique peptides $Ubl^{LysR}$ | gene ontology | function |
|---|---|---|---|---|---|---|
| 1 | ARF-binding protein 1 [*Homo sapiens*] | gil56417899 | 0 | 35 | GO:0006886 | intracellular protein transport |
| 2 | Predicted: similar to ubiquitin protein ligase E3A [*Bos taurus*] | gil119913502 | 0 | 19 | GO:0004842 | ubiquitin-protein ligase activity |
| 3 | Conserved hypothetical protein [*Chaetomium globosum* CBS 148.51] | gil116181620 | 1 | 8 | GO:0006511 | ubiquitin-dependent protein catabolic process |
| 4 | Tubulin beta chain (Beta tubulin) [*Paracentrotus lividus*] | gil135489 | 0 | 7 | GO:0005200 | structural constituent of cytoskeleton |
| 5 | Predicted: similar to E3 ubiquitin-protein ligase HECTD1 [*Gallus gallus*] | gil118091807 | 0 | 7 | GO:0004842 | ubiquitin-protein ligase activity |
| 6 | Predicted: similar to 26S proteasome subunit p44.5 [*Monodelphis domestica*] | gil126313851 | 0 | 6 | GO:0006511 | ubiquitin-dependent protein catabolic process |
| 7 | Unnamed protein product [*Mus musculus*] | gil74141261 | 0 | 6 | GO:0006511 | ubiquitin-dependent protein catabolic process |
| 8 | Predicted: similar to heat shock protein 8 [*Mus musculus*] | gil94387119 | 0 | 6 | GO:0006986 | response to unfolded protein |
| 9 | Arachin Ahy-4 [*Arachis hypogaea*] | gil57669861 | 0 | 6 | GO:0045735 | nutrient reservoir activity |
| 10 | Predicted: similar to Actin, cytoplasmic 2 (Gamma-actin) [*Rattus norvegicus*] | gil109507063 | 1 | 5 | GO:0005200 | structural constituent of cytoskeleton |
| 11 | Tubulin alpha 6 [*Homo sapiens*] | gil14389309 | 0 | 5 | GO:0005200 | structural constituent of cytoskeleton |
| 12 | 40S ribosomal protein S18 [*Ictalurus punctatus*] | gil41018101 | 0 | 5 | GO:0003735 | structural constituent of ribosome |
| 13 | Predicted: similar to KIAA1734 protein [*Gallus gallus*] | gil118089629 | 0 | 3 | GO:0004842 | ubiquitin-protein ligase activity |
| 14 | Predicted: similar to proteasome [*Monodelphis domestica*] | gil126278093 | 0 | 3 | GO:0006511 | ubiquitin-dependent protein catabolic process |
| 15 | LBA [*Mus musculus*] | gil10180266 | 0 | 3 | GO:0016197 | endosome transport |
| 16 | 26S proteasome regulatory chain 4 [validated] [*Homo sapiens*] | gil345717 | 0 | 3 | GO:0006511 | ubiquitin-dependent protein catabolic process |
| 17 | Predicted: similar to carbamoyl-phosphate synthetase 2 [*Rattus norvegicus*] | gil109477973 | 0 | 2 | GO:0004087 | carbamoyl-phosphate synthase activity |
| 18 | Thyroid hormone receptor interactor 12 [*Homo sapiens*] | gil10863903 | 0 | 2 | GO:0004842 | ubiquitin-protein ligase activity |
| 19 | Predicted: similar to E3 ubiquitin protein ligase [*Macaca mulatta*] | gil109087122 | 0 | 2 | GO:0004842 | ubiquitin-protein ligase activity |
| 20 | Predicted: hypothetical protein [*Danio rerio*] | gil125843812 | 0 | 2 | GO:0006397 | mRNA processing |
| 21 | Polyubiquitin [*Cricetulus griseus*] | gil2627133 | 0 | 2 | GO:0006464 | protein modification process |
| 22 | Chain A, VcpP97 complexed with ADP [*Mus musculus*] | gil62738726 | 0 | 2 | GO:0006986 | response to unfolded protein |
| 23 | Ribosomal protein S10 [*Solea senegalensis*] | gil124300805 | 0 | 2 | GO:0003735 | structural constituent of ribosome |
| 24 | Rps13 protein [*Mus musculus*] | gil15029927 | 0 | 2 | GO:0003735 | structural constituent of ribosome |
| 25 | 40S ribosomal protein S2 [*Cricetulus griseus*] | gil1173239 | 0 | 2 | GO:0003735 | structural constituent of ribosome |
| 26 | HSP 90-beta [*Bos taurus*] | gil118601868 | 0 | 2 | GO:0006986 | response to unfolded protein |
| 27 | Ubiquitin-conjugating enzyme [*Homo sapiens*] | gil10444495 | 0 | 2 | GO:0004842 | ubiquitin-protein ligase activity |
| 28 | Predicted: similar to Hornerin [*Homo sapiens*] | gil113412358 | 0 | 2 | GO:0005509 | calcium ion binding |
| 29 | Ribosomal protein S27 [*Homo sapiens*] | gil4506711 | 0 | 2 | GO:0003735 | structural constituent of ribosome |
| 30 | Rcd1 (required for cell differentiation) homologue 1 [*Mus musculus*] | gil10946722 | 0 | 2 | GO:0019221 | cytokine and chemokine mediated signaling |

[a] All observed MS/MS spectra were subjected to search against the NCBI nonredundant (NCBInr) data bank.

## Table 6.3 Identification of Ubiquitination Sites (K) in Ubiquitin-Conjugated Proteins.

| peptide | Mascot ion score (min–max) | number of GG spectra | number of LRGG spectra | protein identification | accession number |
|---|---|---|---|---|---|
| TLTG**K**TITLEVEPSDTIENVK | 35–90 | 26 | 14 | Ubiquitin | P62972 |
| LIFAG**K**QLEDGR | 22–65 | 16 | 16 | Ubiquitin/Ubi$^{K48R}$ | P62972 |
| LLA**K**VR | 21–39 | 2 | 0 | Putative uncharacterized protein | Q56A66 |
| IEDA**K**R | 36 | 1 | 0 | Protein XRP2 (Retinitis pigmentosa 2) | Q8AVX5 |
| KN**K**NITK | 35 | 1 | 0 | Putative uncharacterized protein | A5PKM7 |
| W**K**DIVLR | 34 | 1 | 0 | MGC86445 protein | Q66KP3 |
| KAAIS**K**SIR | 34 | 1 | 0 | Protein mab-21-like 1 | Q6GQD9 |
| NL**K**WIGLDLLNGKPR | 33–33 | 0 | 3 | T-complex protein 1, alpha subunit | Q6NTQ5 |
| Q**KK**LLK | 30 | 1 | 0 | LOC494754 protein | Q2VPI0 |
| EV**K**LDR | 30 | 1 | 0 | MGC131293 protein | Q2VPL1 |
| ICLDIL**K**DK | 28 | 1 | 0 | Ubiquitin carrier protein | Q7SZ88 |
| EALE**K**SEARR | 24 | 1 | 0 | Ventricular myosin heavy chain | Q5EC17 |
| IY**K**TDREK | 23 | 1 | 0 | Ubiquitin-conjugating enzyme E2 D2 | P62840 |

[a] GG (+114) and LRGG (+383) ubiquitin remnants tags were targeted as variable modifications in MASCOT. Only peptides with internal lysine modifications are listed.

## 6.7 References

(1) Sun, L.; Chen, Z. J. The novel functions of ubiquitination in signaling. Curr. Opin. Cell Biol. 2004, 16 (2), 119–126.

(2) Fang, S.; Weissman, A. M. A field guide to ubiquitylation. Cell. Mol. Life Sci. 2004, 61 (13), 1546–1561.

(3) Pfleger, C. M.; Lee, E.; Kirschner, M. W. Substrate recognition by the Cdc20 and Cdh1 components of the anaphase-promoting complex. Genes Dev. 2001, 15, 2396–2407.

(4) Castro, A.; Bernis, C.; Vigneron, S.; Labbe´, J. C.; Lorca, T. The anaphase-promoting complex: a key factor in the regulation of cell cycle. Oncogene 2005, 24 (3), 314–325.

(5) Hames, R. S.; Wattam, S. L.; Yamano, H.; Bacchieri, R.; Fry, A. M. APC/C-mediated destruction of the centrosomal kinase Nek2A occurs in early mitosis and depends upon a cyclin A-type D-box. EMBO J. 2001, 20, 7117–7127.

(6) Hochegger, H.; Klotzbucher, A.; Kirk, J.; Howell, M.; le Guellec, K.; Fletcher, K.; Duncan, T.; Sohail, M.; Hunt, T. New B-type cyclin synthesis is required between meiosis I and II during Xenopus oocyte maturation. Development 2001, 128, 3795–3807.

(7) Chesnel, F.; Bazile, F.; Pascal, A.; Kubiak, J. Z. Cyclin B dissociation from CDK1 precedes its degradation upon MPF inactivation in mitotic extracts of Xenopus laevis embryos. Cell Cycle 2006, 5 (15), 1687–1698.

(8) Hagting, A.; Den Elzen, N.; Vodermaier, H. C.; Waizenegger, I. C.; Peters, J. M.; Pines, J. Human securin proteolysis is controlled by the spindle checkpoint and reveals when the APC/C switches from activation by Cdc20 to Cdh1. J. Cell Biol. 2002, 157, 1125–1537.

(9) Terret, M. E.; Wassmann, K.; Waizenegger, I.; Maro, B.; Peters, J. M.; Verlhac, M. H. The meiosis I-to-meiosis II transition in mouse oocytes requires separase activity. Curr. Biol. 2003, 13 (20), 1797–1802.

(10) Arlot-Bonnemains, Y.; Klotzbucher, A.; Giet, R.; Uzbekov, R.; Bihan, R.; Prigent, C. Identification of a functional destruction box in the Xenopus laevis aurora-A kinase pEg2. FEBS Lett. 2001, 508, 149–152.

(11) Ayad, N. G.; Rankin, S.; Murakami, M.; Jebanathirajah, J.; Gygi, S.; Kirschner, M. W. Tome-1, a trigger of mitotic entry, is degraded during G1 via the APC. Cell 2003, 113, 101–113.

(12) Baker, D. J.; Dawlaty, M. M.; Galardy, P.; van Deursen, J. M. Mitotic regulation of the anaphase-promoting complex. Cell. Mol. Life Sci. 2007, 64 (5), 589–600.

(13) Raff, J. W.; Jeffers, K.; Huang, J. Y. The roles of Fzy/Cdc20 and Fzr/Cdh1 in regulating the destruction of cyclin B in space and time. J. Cell Biol. 2002, 157, 1139–1149.

(14) Joo, H. Y.; Zhai, L.; Yang, C.; Nie, S.; Erdjument-Bromage, H.; Tempst, P.; Chang, C.; Wang, H. Regulation of cell cycle progression and gene expression by H2A deubiquitination. Nature 2007, 449, 1068–1072.

(15) Masui, Y.; Wang, P. Cell cycle transition in early embryonic development of Xenopus laevis. Biol Cell. 1998, 90 (8), 537–548.

(16) Uto, K.; Nakajo, N.; Sagata, N. Two structural variants of Nek2 kinase, termed Nek2A and Nek2B, are differentially expressed in Xenopus tissues and development. Dev. Biol. 1999, 208, 456–464.

(17) Lorca, T.; Castro, A.; Martinez, A. M.; Vigneron, S.; Morin, N.; Sigrist, S.; Lehner, C; Doree, M; Labbe, J. C. Fizzy is required foractivation of the APC/cyclosome in Xenopus egg extracts. EMBO J. 1998, 17 (13), 3565–3575.

(18) Castro, A.; Arlot-Bonnemains, Y.; Vigneron, S.; Labbe′, J. C.; Prigent, C.; Lorca, T. APC/Fizzy-Related targets Aurora-A kinase for proteolysis. EMBO Rep. 2002, 3 (5), 457–462.

(19) Potapova, T. A.; Daum, J. R.; Pittman, B. D.; Hudson, J. R.; Jones, T. N.; Satinover, D. L.; Stukenberg, P. T.; Gorbsky, G. J. The reversibility of mitotic exit in vertebrate cells. Nature 2006, 440 (7086), 954–958.

(20) Denison, C.; Kirkpatrick, D. S.; Gygi, S. P. Proteomic insights into ubiquitin and ubiquitin-like proteins. Curr. Opin. Chem. Biol. 2005, 9 (1), 69–75.

(21) Kirkpatrick, D. S.; Denison, C.; Gygi, S. P. Weighing in on ubiquitin: the expanding role of mass-spectrometry-based proteomics. Nat. Cell Biol. 2005, 7 (8), 750–757.

(22) Xu, P.; Peng, J. Dissecting the ubiquitin pathway by mass spectrometry. Biochim. Biophys. Acta 2006, 1764 (12), 1940–1947.

(23) Peng, J.; Schwartz, D.; Elias, J. E.; Thoreen, C. C.; Cheng, D.; Marsischky, G.; Roelofs, J.; Finley, D.; Gygi, S. P. A proteomics approach to understanding protein ubiquitination. Nat. Biotechnol. 2003, 21 (8), 921–926.

(24) Kirkpatrick, D. S.; Weldon, S. F.; Tsaprailis, G.; Liebler, D. C.; Gandolfi, A. J. Proteomic identification of ubiquitinated proteins from human cells expressing His-tagged ubiquitin. Proteomics 2005, 5 (8), 2104–2111.

(25) Matsumoto,M.; Hatakeyama, S.; Oyamada, K.; Oda, Y.; Nishimura, T.; Nakayama, K. I. Large-scale analysis of the human ubiquitinrelated proteome. Proteomics 2005, 5 (16), 4145–4151.

(26) Vasilescu, J.; Smith, J. C.; Ethier, M.; Figeys, D. Proteomic analysis of ubiquitinated proteins from human MCF-7 breast cancer cells by immunoaffinity purification and mass spectrometry. J. Proteome Res. 2005, 4 (6), 2192–2200.

(27) Warren, M. R.; Parker, C. E.; Mocanu, V.; Klapper, D.; Borchers, C. H. Electrospray ionization tandem mass spectrometry of model peptides reveals diagnostic fragment ions for protein ubiquitination. Rapid Commun. Mass Spectrom. 2005, 19 (4), 429–437.

(28) Denis, N. J.; Vasilescu, J.; Lambert, J. P.; Smith, J. C.; Figeys, D. Tryptic digestion of ubiquitin standards reveals an improved strategy for identifying ubiquitinated proteins by mass spectrometry. Proteomics 2007, 7 (6), 868–874.

(29) Bazile, F.; Pascal, A.; Karaiskou, A.; Chesnel, F.; Kubiak, J. Z. Absence of reciprocal feedback between MPF and ERK2 MAP kinase in mitotic Xenopus laevis embryo cell-free extract. Cell Cycle 2007, 6 (4), 489–496.

(30) Chau, V.; Tobias, J. W.; Bachmair, A.; Marriott, D.; Ecker, D. J.; Gonda, D. K.; Varshavsky, A. A multiubiquitin chain is confined to specific lysine in a targeted short-lived protein. Science 1989, 243 (4898), 1576–1583.

(31) Finley, D.; Sadis, S.; Monia, B. P.; Boucher, P.; Ecker, D. J.; Crooke, S. T.; Chau, V. Inhibition of proteolysis and cell cycle progression in a multiubiquitination-deficient yeast mutant. Mol. Cell. Biol. 1994, 14 (8), 5501–5509.

(32) Jeon, H. B.; Choi, E. S.; Yoon, J. H.; Hwang, J. H.; Chang, J. W.; Lee, E. K.; Choi, H. W.; Park, Z. Y.; Yoo, Y. J. A proteomics approach to identify the ubiquitinated proteins in mouse heart. Biochem. Biophys. Res. Commun. 2007, 357 (3), 731–736.

(33) Maor, R.; Jones, A.; Nu¨hse, T. S.; Studholme, D. J.; Peck, S. C.; Shirasu, K. Multidimensional protein identification technology (MudPIT) analysis of ubiquitinated proteins in plants. Mol. Cell. Proteomics 2007, 6 (4), 601–610.

(34) Ikeda, F.; Dikic, I. Atypical ubiquitin chains: new molecular signals. EMBO Reports 2008, 9 (6), 536–42.

(35) Murray, A. W. Cell cycle extracts. Methods Cell Biol. 1991, 36, 581– 605.

(36) Chesnel, F.; Vignaux, F.; Richard-Parpaillon, L.; Huguet, A.; Kubiak, J. Z. Differences in regulation of the first two M-phases in Xenopus laevis embryo cell-free extracts. Dev. Biol. 2005, 285, 358–375.

(37) Chesnel, F.; Bazile, F; Pascal, A.; Kubiak, J. Z. Cyclin B2/cyclindependent kinase1 dissociation precedes CDK1 Thr-161 dephosphorylation upon M-phase promoting factor inactivation in Xenopus laevis cell-free extract. Int. J. Dev. Biol. 2007, 51, 297–305.

(38) Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 1970, 227, 680–685.

(39) The UniProt Consortium. The Universal Protein Resource (Uni-Prot). Nucleic Acids Res. 2007, 35, D193-D197.

(40) Keller, A.; Nesvizhskii, A. I.; Kolker, E.; Aebersold, R. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. Anal. Chem. 2002, 74 (20), 5383–5392.

(41) Nesvizhskii, A. I.; Keller, A.; Kolker, E.; Aebersold, R. A statistical model for identifying proteins by tandem mass spectrometry. Anal. Chem. 2003, 75 (17), 4646–4658.

(42) Vasilescu, J.; Smith, J. C.; Zweitzig, D. R.; Denis, N. J.; Haines, D. S.; Figeys, D. Systematic determination of ion score cutoffs based on calculated false positive rates: application for identifying ubiquitinated proteins by tandem mass spectrometry. J. Mass Spectrom. 2008, 43 (3), 296–304.

(43) Deveraux, Q.; Ustrell, V.; Pickart, C.; Rechsteiner, M. A 26 S protease subunit that binds ubiquitin conjugates. J. Biol. Chem. 1994, 269 (10), 7059–7061.

(44) Dai, R. M.; Li, C. C. Valosin-containing protein is a multi-ubiquitin chain-targeting factor required in ubiquitin-proteasome degradation. Nat. Cell Biol. 2001, 3 (8), 740–744.

(45) Rao, H.; Sastry, A. Recognition of specific ubiquitin conjugates is important for the proteolytic functions of the ubiquitin-associated domain proteins Dsk2 and Rad23. J. Biol. Chem. 2002, 277, 11691–11695.

(46) Hoege, C.; Pfander, B.; Moldovan, G. L.; Pyrowolakis, G.; Jentsch, S. RAD6-dependent DNA repair is linked to modification of PCNA by ubiquitin and SUMO. Nature 2002, 419, 135–141.

(47) Lindon, C.; Pines, J. Ordered proteolysis in anaphase inactivates Plk1 to contribute to proper mitotic exit in human cells. J. Cell Biol. 2004, 164 (2), 233–241.

(48) Cheeseman, I. M.; Desai, A. Cell division: AAAtacking the mitotic spindle. Curr. Biol. 2004, 14 (2), R70–R72.

(49) Bruderer, R. M.; Brasseur, C.; Meyer, H. H. The AAA ATPase p97/ VCP interacts with its alternative co-factors, Ufd1-Npl4 and p47, through a common bipartite binding mechanism. J. Biol. Chem. 2004, 279 (48), 49609–49616.

(50) Pye, V. E.; Beuron, F.; Keetch, C. A.; McKeown, C.; Robinson, C. V.; Meyer, H. H.; Zhang, X.; Freemont, P. S. Structural insights into the p97-Ufd1-Npl4 complex. Proc. Natl. Acad. Sci. U.S.A. 2007, 104 (2), 467–472.

(51) Lass, A.; McConnell, E.; Fleck, K.; Palamarchuk, A; Wojcik, C. Analysis of Npl4 deletion mutants in mammalian cells unravels new Ufd1-interacting motifs and suggests a regulatory role of Npl4 in ERAD. Exp. Cell Res. 2008, 314 (14), 2715–2723.

(52) Petretti, C.; Savoian, M.; Montembault, E.; Glover, D. M.; Prigent, C.; Giet, R. The PITSLRE/CDK11p58 protein kinase promotes centrosome maturation and bipolar spindle formation. EMBO Rep. 2006, 7 (4), 418–424.

(53) Hu, D.; Valentine, M.; Kidd, V. J.; Lahti, J. M. CDK11(p58) is required for the maintenance of sister chromatid cohesion. J Cell Sci. 2007, 120, 2424–2434.

(54) Yokoyama, H.; Gruss, O. J.; Rybina, S.; Caudron, M.; Schelder, M.; Wilm, M.; Mattaj, I. W.; Karsenti, E. Cdk11 is a RanGTP-dependent microtubule stabilization factor that regulates spindle assembly rate. J. Cell Biol. 2008, 180 (5), 867–875.

(55) Vasudevan, S.; Starostina, N. G.; Kipreos, E. T. The Caenorhabditis elegans cell-cycle regulator ZYG-11 defines a conserved family of CUL-2 complex components. EMBO Rep. 2007, 8 (3), 279–286.

(56) Liu, J.; Vasudevan, S.; Kipreos, E. T. CUL-2 and ZYG-11 promote meiotic anaphase II and the proper placement of the anteriorposterior axis. Development 2004, 131, 3513–3525.

(57) Sutovsky, P.; Manandhar, G.; McCauley, T. C.; Caamano, J. N.;Sutovsky, M.; Thompson, W. E.; Day, B. N. Proteasomal interference prevents zona pellucida penetration and fertilization in mammals. Biol. Reprod. 2004, 71 (5), 1625–1637.

(58) Yi, Y. J.; Manandhar, G.; Oko, R. J.; Breed, W. G.; Sutovsky, P. Mechanism of sperm-zona pellucida penetration during mammalian fertilization: 26S proteasome as a candidate egg coat lysin. Soc. Reprod. Fertil. Suppl. 2007, 63, 385–408.

(59) Peng, J. Evaluation of proteomic strategies for analyzing ubiquitinated proteins. BMB Rep. 2008, 41 (3), 177–183.

(60) Kirkpatrick, D. S.; Hathaway, N. A.; Hanna, J.; Elsasser, S.; Rush, J.; Finley, D.; King, R. W.; Gygi, S. P. Quantitative analysis of in vitro ubiquitinated cyclin B1 reveals complex chain topology. Nat. Cell Biol. 2006, 8, 700–710.

(61) Jin, L.; Williamson, A.; Banerjee, S.; Philipp, I.; Rape, M. Mechanism of ubiquitin-chain formation by the human anaphase-promoting complex. Cell 2008, 133 (4), 653–665.

(62) Ravid, T.; Hochstrasser, M. Autoregulation of an E2 enzyme by ubiquitin-chain assembly on its catalytic residue. Nat. Cell Biol. 2007, 9 (4), 422–427.

(63) Yi, C.; Li, S.; Wang, J.; Wei, N.; Deng, X. W. Affinity purification reveals the association of WD40 protein constitutive photomorphogenic1 with the hetero-oligomeric TCP-1 chaperonin complex in mammalian cells. Int. J. Biochem. Cell Biol. 2006, 38 (7), 1076–1083.

(64) Zhou, C.; Cunningham, L.; Marcus, A. I.; Li, Y.; Kahn, R. A. Arl2 and Arl3 regulate different microtubule-dependent processes. Mol. Biol. Cell 2006, 17 (5), 2476–2487.

(65) McCracken, S.; Longman, D.; Marcon, E.; Moens, P.; Downey, M.; Nickerson, J. A.; Jessberger, R.; Wilde, A.; Blencowe, B. J. Proteomic analysis of SRm160-containing complexes reveals a conserved association with cohesin. J. Biol. Chem. 2005, 280 (51), 42227–42236.

(66) Garriock, R. J.; Meadows, S. M.; Krieg, P. A. Developmental expression and comparative genomic analysis of Xenopus cardiac myosin heavy chain genes. Dev. Dyn. 2005, 233 (4), 1287–1293.

(67) Nencioni, A.; Grunebach, F.; Patrone, F.; Ballestrero, A.; Brossart, P. Proteasome inhibitors: antitumor effects and beyond. Leukemia 2007, 21 (1), 30–36.

# Chapter 7: Discussion

## 7.1 Thesis summary

In the second chapter of my thesis, a novel method for identifying ubiquitinated proteins from a complex biological sample, such as a whole cell lysate, using a combination of immunoaffinity purification and liquid chromatography-tandem mass spectrometry analysis (LC-MS/MS) was described. Application of this method enabled the confident identification of 70 ubiquitinated proteins from the human MCF-7 breast cancer cell line after overnight treatment with the proteasome inhibitor MG132. Proteins identified included Ub-ligating enzymes, subunits of the 26S proteasome complex, heat shock proteins, transport proteins, DNA repair proteins, transcription and translation elongation factors, and several proteins involved in cell cycle regulation, apoptosis, and signal transduction. These findings represented the first proteomic analysis of native ubiquitinated proteins from a genetically unmodified mammalian cell line.

In the third chapter of my thesis, an application of a microfluidic protein processing device, termed the Proteomic Reactor, for enzymatic digestion of proteins for subsequent LC-MS/MS analysis was described. Use of the Proteomic Reactor enabled 27 ubiquitination sites to be precisely mapped on 21 proteins and enabled the identification of 58 candidate ubiquitinated proteins from control and VCP depleted H1299 lung adenocarcinoma cells. Differences in ubiquitination were also observed when two populations of H1299 cells (i.e. -/+ VCP depletion) were compared. Several proteins designated as candidates were confirmed to be ubiquitinated by

subsequent immunoprecipitation and Western blotting experiments. Our results demonstrated that the Proteomic Reactor facilitates the analysis of ubiquitinated proteins and may be used for the study of other post-translational modifications.

In the fourth chapter of my thesis, a method for determining peptide ion score cutoffs that permit the confident identification of ubiquitinated proteins by MS/MS was described. For initial experiments, we analysed multi-Ubiquitin (Ub) chains with quadrupole time-of-flight and quadrupole ion trap mass spectrometers in combination with the Mascot database search engine. We found that standard ion score cutoffs were not sufficiently stringent as significant false positive rates (FPR) were observed. We also found that FPR and false negative rates varied significantly depending on the cutoff score used and that appropriate cutoff scores could only be determined following a systematic evaluation of FPR. When a range of cutoff scores was used for the analysis of complex mixtures of ubiquitinated proteins obtained from human cell lysates by immunopurification with an anti-Ub antibody, unacceptably high FPR were observed. As a final set of experiments, MS/MS data obtained for the analysis of immunopurified ubiquitinated proteins was searched against five protein databases of increasing size and it was determined that FPR for ubiquitinated proteins are affected by the size of the protein database that is searched.

In the fifth chapter of my thesis, a specific application of the gel-based method I developed for identifying ubiquitinated proteins was described for Mga2p, an endoplasmic reticulum (ER)-localized transcription factor that is released from the ER membrane by a unique Ub-dependent mechanism. We identified, using an *in vitro* ubiquitinated Mga2p120–Mga2p90 complex, that lysine residues 983 and 985 contained within the carboxy-terminal domain of Mga2p120 are Rsp5p-directed Ub-conjugation sites. Mutation of these residues as well as

proximally located lysine 980 resulted in suppression of Mga2p120 ubiquitination *in vitro* and *in vivo*, inefficient liberation of Mga2p90 by $Cdc48p^{Npl4p/Ufd1p}$ *in vitro*, and ER membrane retention of Mga2p in cells. Moreover, mga2Δ/spt23ts harboring Rsp5p binding and conjugation mga2 mutants expressed low OLE1 (an Mga2p90 target gene) transcripts and displayed reduced growth. We concluded that residues 980, 983, and 985 are targets of Rsp5p-induced polyubiquitination and mediate $Cdc48p^{Npl4p/Ufd1p}$-dependent Mga2p90–Mga2p120 separation and Mga2p90 mobilization.

Finally, in the sixth chapter of my thesis, a specific application of the method I developed for determining peptide ion score cutoffs that permit the confident identification of ubiquitinated proteins by MS/MS was described for a proteomic analysis of *Xenopus laevis* embryos. Mutated recombinant 6×His-tagged ubiquitin ($Ubi^{K48R}$) was added to mitotic cell-free embryo extracts from which Ub-conjugated proteins were purified, as well as associated proteins, in nondenaturing conditions by cobalt affinity chromatography. Proteins eluted from $Ubi^{K48R}$ supplemented and control extracts were compared by LC-MS/MS analysis after monodimensional SDS-PAGE. A total of 144 proteins potentially ubiquitinated or associated with them were identified. Forty-one percent of these proteins were shown to be involved in ubiquitination and/or proteasomal degradation pathway confirming the specificity of the screen. Twelve proteins, among them ubiquitin itself, were shown to carry a GG or LRGG tag indicating their direct ubiquitination. Sequence analysis of ubiquitinated substrates carrying these tags indicated that in *Xenopus* cell-free embryo extract supplemented with $Ubi^{K48R}$, the majority of polyubiquitination occurred through lysine-11 specific ubiquitin chain polymerization.

## 7.2 Future perspectives

Several follow-up studies could be performed based on the results presented in this thesis. Since the gel-based method discussed in chapter two was successfully used to map Rsp5p-directed Ub-conjugation sites on Mga2p, a reasonable set of future experiments would involve the study of Mga2p's homologue, Spt23p, using a similar approach. Spt23p encodes a p90 product that also undergoes Rsp5p-dependent release from the endoplasmic reticulum membrane and possesses redundant OLE1 transactivation function. It is expected that the mapping of Ub-conjugation sites on Spt23p and assessment of the *in vitro* and *in vivo* effects of mutating these lysine residues would be fairly straightforward. However, if initial attempts to identify Ub-conjugation sites on Spt23p fail using the current methodology, several modifications to the protocol could be made to improve its sensitivity given the availability of new affinity purification methods and mass spectrometry-based technologies.

As an example, the current affinity purification protocol for capturing ubiquitinated proteins employs an affinity resin consisting of an anti-Ub antibody (clone FK2) coupled to protein-G agarose beads. Since the article's publication in 2005, an improved affinity resin employing commercially available magnetic protein-G Dynabeads (Invitrogen) has been successfully employed and reported by our group and others. Similar to protein-A/G agarose beads, an antibody of interest (i.e. anti-Ub antibody) can be covalently coupled to the protein-A/G Dynabeads with dimethyl pimelimidate (DMP). However, the use of the Dynabeads avoids two problems associated with agarose beads that occur during the washing step; 1) the loss of beads and sample; and 2) unwanted remaining washing buffer. These problems are avoided because all of the washing steps take place in a single microfuge tube with no centrifugations

since the Dynabeads are retained on the walls of the tube using a magnet. Other benefits of the Dynabeads include reduced background contamination, since fewer proteins bind non-specifically to the beads, as well as reduced sample processing times. It would therefore be of interest to determine the improvements in sensitivity that could be achieved by incorporating this technology into our purification protocol for ubiquitinated proteins.

Another set of follow-up experiments that could be performed would involve the study of selected proteins from the group of candidate ubiquitinated proteins discussed in the third and sixth chapters to map their specific Ub-conjugation sites. Specifically, proteins among the group of 58 candidates identified in control and VCP-depleted H1299 lung adenocarcinoma cells from Chapter 3 (Group A) and among the group of 144 candidates identified in *Xenopus laevis* cell-free embryo extracts from Chapter 6 (Group B). Among the candidates from Group A, Sequestosome-1 represents an ideal target because it contains an ubiquitin-associated (UBA) domain, is involved in proteasomal degradation of ubiquitinated proteins, and was shown by us to be ubiquitinated by immunoprecipitation and western blotting. Among the candidates from Group B, several proteins assigned with a role in (or related to) cell cycle and division are ideal targets, including PCNA, plk1, Np14, VCP, CDK11, MCM7A, PP2A, Orc3-like p81 and cdc2a. CDK11, for example, is of particular interest since the C-terminus part of p110 CDK11 has been shown to be specifically expressed during M-phase in mammalian somatic cells and is required for its completion. For the remaining proteins in Groups A and B, it would be of interest to specifically search for the presence of a UBA domain within their primary structure, since this 40 amino acid motif is known bind both mono- and poly-Ub *in vitro* and has been proposed to limit Ub chain elongation and to target ubiquitinated proteins to the 26S proteasome for degradation.

A further set of experiments which represents a significant undertaking, but would have implications in biomarker discovery and the potential to aid in predicting and diagnosing diseases, involves the identification and profiling of ubiquitinated proteins in biological fluids, such as blood plasma or serum. Although no studies have reported the identification of ubiquitinated proteins in human plasma, we recently confirmed the presence of ubiquitinated proteins in mouse plasma (from both normal and diseased samples) by immunoprecipitation and western blotting. We believe that the identification of plasma ubiquitinated proteins using the methods discussed in this thesis will be feasible once a suitable albumin depletion system has been incorporated into our purification protocol. This represents a critical step towards this goal since albumin is the most abundant protein in plasma and impairs our ability to identify lower-abundance proteins of interest. Currently, several commercially available albumin and immunoglobulin (IgG) depletion kits are available which might serve this purpose. Once the identity of specific ubiquitinated proteins are confirmed in plasma, it would also be of interest to determine whether ubiquitination of these proteins occurs in the blood and if so, what purpose and function it might serve. Interaction partners of ubiquitinated proteins in plasma could also be identified by co-immunoprecipitation and LC-MS/MS analysis, which might reveal which E3 ubiquitin ligases they associate or interact with. To confirm whether E1 activating, E2 conjugating and E3 ligating enzymes are functional in plasma, *in vitro* ubiquitination assays could be performed where the identified enzymes and substrates are added in the presence of epitope-tagged Ub. Ubiquitinated substrates would be visualized by western blotting using an antibody that recognizes the epitope tag. Immunopurification using the same antibody followed by LC-MS/MS analysis would then be employed to identify their Ub-conjugation sites.

**Appendix: Curriculum vitae**

Contact information

Name: Julian Vasilescu

Mailing address: c/o BMI Dept., 451 Smyth Road, Ottawa, Ontario, K1H 8M5, Canada

Email: jvasiles@uottawa.ca or julian_vas@hotmail.com

Education

Doctor of Philosophy – Biochemistry (2006-2010)

University of Ottawa, Ottawa, Canada

Master of Science – Experimental Medicine (2002-2004)

University of British Columbia, Vancouver, Canada

Bachelor of Science – Biochemistry (1998-2002)

McGill University, Montreal, Canada

Publications

1) Gyamera-Acheampong C, Vasilescu J, Figeys D, Mbikay M. PCSK4-null sperm display enhanced protein tyrosine phosphorylation and ADAM2 proteolytic processing during in vitro capacitation. Fertility and Sterility. 2009. [Epub ahead of print Mar 31, 2009].

2) Zhou H, Hou W, Denis NJ, Zhou H, Vasilescu J, Figeys D. Glycoproteomic Reactor for Human Plasma. Journal of Proteome Research. Vol 8: 556-566, 2009.

3) Bhattacharya S, Shcherbik N, Vasilescu J, Smith JC, Figeys D, Haines DS. Identification of lysine residues within the carboxy-terminus of Mga2p120 that are targets of Rsp5p-dependent ubiquitination and mediate mobilization of interaction Mga2p90. Journal of Molecular Biology. Vol. 385: 718-725, 2009.

4) Bazile F, Gagne JP, Mercier G, Lo KS, Pascal A, Vasilescu J, Figeys D, Poirier GG, Kubiak JZ, Chesnel F. Differential proteomic screen to evidence proteins ubiquitinated upon mitotic exit in cell-free extract of Xenopus laevis embryo. Journal of Proteome Research. Vol. 7: 4701-4714, 2008.

5) Mathivanan S, (144 other authors), Vasilescu J, Veenstra TD, Vidal-Taboada JM, Vihinen M, Wait R, Wang X, Wiemann S, Wu B, Xu T, Yates JR, Zhong J, Zhou M, Zhu Y, Zuerbig P, Pandey A. Human Proteinpedia: A portal for sharing and integration of human protein data. Nature Biotechnology. Vol. 26: 164-167, 2008.

6) Vasilescu J, Smith JC, Zweitzig D, Denis NJ, Haines DS, Figeys D. Systematic determination of ion score cutoffs based on calculated false positive rates: Application for identifying ubiquitinated proteins by tandem mass spectrometry. Journal of Mass Spectrometry. Vol. 43: 296-304, 2008.

7) Denis NJ, Vasilescu J, Lambert JP, Smith JC, Figeys D. Tryptic digestion of ubiquitin standards reveals an improved strategy for identifying ubiquitinated proteins by mass spectrometry. Proteomics. Vol. 7: 868-874. 2007.

8) Ewing RM, (19 other authors), Vasilescu J, Abu-Farha M,Lambert JP, Duewel HS, Stewart II, Kuehl B, Hogue K, Colwill K, Gladwish K, Muskat B, Kinach R, Adams SL, Moran MF, Morin GB, Topaloglou T, Figeys D. Large-scale mapping of human protein-protein interactions by mass spectrometry. Molecular Systems Biology. Vol. 3: 1-17, 2007.

9) Vasilescu J, Zweitzig D, Denis NJ, Smith JC, Ethier M, Haines DS, Figeys D. The proteomic reactor facilitates the analysis of affinity-purified proteins by mass spectrometry: application for identifying ubiquitinated proteins in human cells. Journal of Proteome Research. Vol. 6: 298-305, 2007.

10) Lee JW, Figeys D, Vasilescu J. Biomarker assay translation from discovery to clinical studies in cancer drug development: quantification of emerging protein biomarkers. Advances in Cancer Research. Vol. 96: 269-298, 2006.

11) Vasilescu J, Figeys D. Mapping protein-protein interactions by mass spectrometry. Current Opinion in Biotechnology. Vol. 17: 394-399, 2006.

12) Ethier M*, Lambert JP*, Vasilescu J*, Figeys D. Analysis of protein interaction networks using mass spectrometry compatible techniques. Analytica Chimica Acta. Vol. 564: 10-18, 2006. *equal contribution

13) Vasilescu J, Smith JC, Ethier M, Figeys D. Proteomic analysis of ubiquitinated proteins by immunoaffinity purification and mass spectrometry. Journal of Proteome Research. Vol. 4: 2192-2200, 2005.

14) Vasilescu J, Guo X, Kast J. Identification of protein-protein interactions using in vivo cross-linking and mass spectrometry. Proteomics. Vol. 4: 3845-3854, 2004.

15) Monneret G, Li H, Vasilescu J, Rokach J, Powell WS. 15-deoxy-delta(12,14)-prostaglandins D2 and J2 are potent activators of human eosinophils. Journal of Immunology. Vol. 168: 3563-3569, 2002.