

RESEARCH ARTICLE

Tryptic digestion of ubiquitin standards reveals an improved strategy for identifying ubiquitinated proteins by mass spectrometry

Nicholas J. Denis, Julian Vasilescu, Jean-Philippe Lambert*, Jeffrey C. Smith* and Daniel Figeys

Ottawa Institute of Systems Biology, University of Ottawa, Ottawa, ON, Canada

Ubiquitination plays an essential role in maintaining cellular homeostasis by regulating a multitude of essential processes. The ability to identify ubiquitinated proteins by MS currently relies on a strategy in which ubiquitinated peptides are identified by a 114.1 Da diglycine (GG) tag on lysine residues, which is derived from the C-terminus of ubiquitin, following trypsin digestion. In the following study, we report a more comprehensive approach for mapping ubiquitination sites by trypsin digestion and MS/MS analysis. We demonstrate that ubiquitination sites can be identified by signature peptides containing a GG-tag (114.1 Da) and an LRGG-tag (383.2 Da) on internal lysine residues as well as a GG-tag found on the C-terminus of ubiquitinated peptides. Application of this MS-based approach enabled the identification of 96 ubiquitination sites from proteins purified from human MCF-7 breast cancer cells, representing a 2.4-fold increase in the number of ubiquitination sites that could be identified over standard methods. Our improved MS-based strategy will aid future studies which aim to identify and/or characterize ubiquitinated proteins in human cells.

Received: June 4, 2006
Revised: September 21, 2006
Accepted: September 29, 2006

**Keywords:**

In-gel digestion / Mass spectrometry / MCF-7 / MG132 / Ubiquitin

1 Introduction

Ubiquitin (Ub) is a highly conserved 76 amino acid protein that is present in all eukaryotic cells. Ub is conjugated to cellular proteins through a conserved mechanism involving Ub activating (E1), conjugating (E2), ligating (E3), and the recently described E4 enzymes [1, 2]. This process, termed ubiquitination, couples the C-terminal glycine residue of Ub to a target lysine residue of the protein substrate, forming an isopeptide bond. Protein ubiquitination regulates many cel-

lular processes including transcription, endocytosis, cell-cycle control, signal transduction, stress response, DNA repair as well as proteasomal-mediated degradation [3–10]. Therefore, it is not surprising that many diseases are associated with abnormalities in the ubiquitination process [11–15].

To date, several methods have been reported for the analysis of ubiquitinated proteins [16–19]. Peng *et al.* [18] identified 1075 potentially ubiquitinated proteins and 110 ubiquitination sites in yeast by LC-MS/MS after fractionating peptides on a strong cation exchange (SCX) column. Vasilescu *et al.* [19] used immunoaffinity purification and SDS-PAGE separation to identify 70 ubiquitinated proteins from human breast cancer cells. Another approach was developed by Matsumoto *et al.* [20] which utilized immunoaffinity pur-

Correspondence: Dr. Daniel Figeys, Ottawa Institute of Systems Biology, 451 Smyth Road, Ottawa, ON, Canada K1H 8M5
E-mail: dfigeys@uottawa.ca
Fax: +1-613-562-5655

Abbreviations: GG, diglycine; Ub, ubiquitin

* Both these authors have contributed equally to this work.

ifications under native and denaturing conditions to identify 670 potentially ubiquitinated proteins and a total of 18 precise ubiquitination sites. All three studies identified ubiquitination sites using an approach which relies on trypsin-mediated proteolysis of ubiquitinated proteins, to yield peptides with internal lysine residues harboring a diglycine (GG) tag derived from the C-terminus of Ub (Fig. 1a) [17]. A recent study using model peptides demonstrated that trypsin digestion produces lysine residues with GG (114.1 Da) and LRGG (383.2 Da) remnants, making it theoretically possible to identify ubiquitinated proteins by MS/MS using these two mass tags [21].

In this study, it was our aim to validate that signature peptides can be identified by GG and LRGG tags on internal lysine residues, and to determine if trypsin can perform proteolysis at ubiquitinated lysine residues. If so, this would provide an additional mass tag on C-terminal lysine residues. To test this hypothesis, Ub and poly-Ub standards were separated by SDS-PAGE, excised and subjected to in-gel digestion, and LC-MS/MS analysis. Our results demonstrate that ubiquitination sites can be identified by both GG and LRGG tags on internal lysine residues, and that trypsin digestion can cleave after ubiquitinated lysine residues. To validate this strategy, we performed an immunopurification experiment using a protein extract from MCF-7 human breast cancer cells treated with the proteasome inhibitor MG132, and confirmed that ubiquitination sites from a complex biological sample could be identified using these mass tags. When compared to the standard MS-based approach, we found a significant increase in the number of

ubiquitination sites that could be identified. Our comprehensive MS-based strategy for mapping ubiquitination sites provides a new method that will aid the study of protein ubiquitination.

2 Materials and methods

2.1 Antibodies and reagents

A mouse mAb (FK2 clone) which has specificity toward mono- and poly-ubiquitinated proteins, but not free Ub, was obtained from Biomol (Plymouth Meeting, PA, USA). The proteasome inhibitor MG132 was obtained from Sigma-Aldrich and a 1 mM stock solution was prepared in DMSO. Mono- and poly-Ub standards were obtained from Biomol and were prepared at a concentration of 5 mg/mL in PBS. All other reagents, unless otherwise stated, were obtained from Fisher Scientific (Burlington, ON, Canada).

2.2 Cell culture

Human MCF-7 breast epithelial cancer cells were cultured at 37°C in 5% CO₂ in DMEM medium supplemented with 10% FBS, 1% non-essential amino acids, and 0.2% insulin in 150 mm tissue culture dishes. Upon reaching 80% confluence, MCF-7 cells were treated with 5 μM MG132 for 8 h. Cells were washed twice in PBS, harvested, and resuspended in modified RIPA lysis buffer (50 mM Tris-HCl, pH 7.4, 150 mM NaCl, 0.25% Na deoxycholate, 1% NP-40, and 1 mM

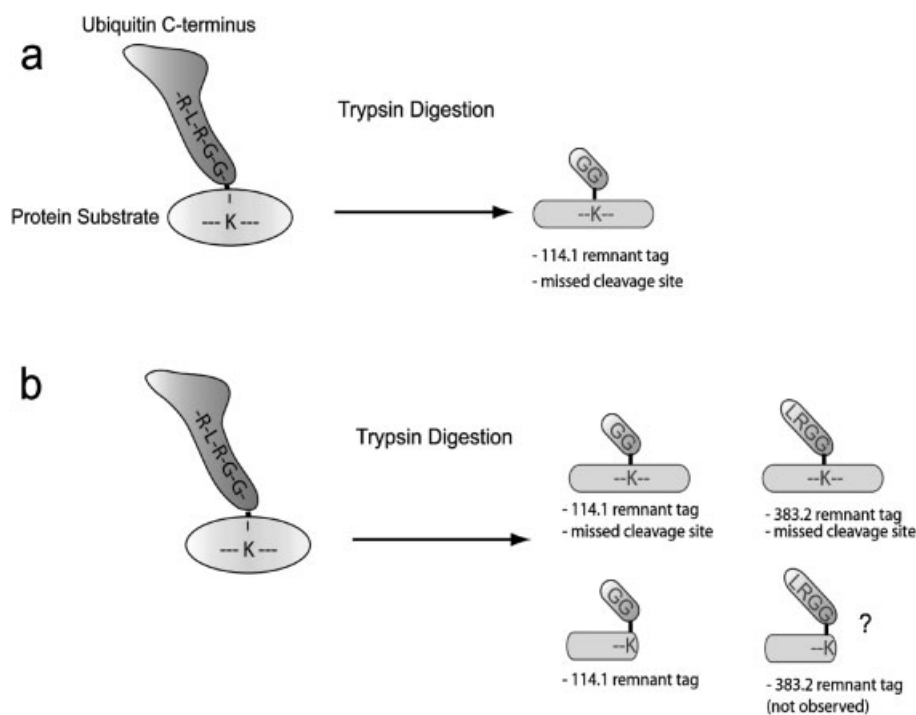


Figure 1. MS-based strategy for identifying ubiquitination sites. (a) The conventional approach for identifying ubiquitination sites relies on the identification of a 114.1 Da tag on internal lysine residues due to a GG remnant derived from Ub. (b) Our comprehensive strategy for identifying ubiquitination sites also includes the identification of a 383.2 Da tag, due to an LRGG remnant of Ub on internal lysine residues as well as a 114.1 Da tag on C-terminal lysine residues.

EDTA) containing a cocktail of protease inhibitors (Roche). Cell lysates were centrifuged for 10 min at 14 000 rpm to pellet cell debris. The supernatant was collected and protein concentration was determined by a Bradford protein assay kit from BioRad (Hercules CA, USA).

2.3 Immunoaffinity purification

The anti-Ub antibody was coupled to protein G-agarose beads as previously described [19]. Immunoaffinity purification of ubiquitinated proteins was performed using a cell lysate containing 250 µg of total protein. Cell lysates were precleared using 10 µL of protein G-agarose beads for 1 h at 4°C. Precleared cell lysates were then subjected to immunoaffinity purification by adding them to 10 µL anti-Ub beads. Samples were allowed to incubate for 4 h at 4°C on a rotator. Anti-Ub beads were washed 4 × with ten volumes of modified RIPA buffer and proteins were eluted with 25 µL of 2% SDS for 5 min at 95°C.

2.4 Protein processing and LC-MS/MS

Ub (5 µg) and poly-Ub (20 µg) were separated on NuPage 4–12% Bis-Tris precast gels (Invitrogen) and visualized using silver nitrate, as were immunoaffinity eluates from MCF-7 cells. Gel bands were excised and subjected to in-gel tryptic digestion as previously described [22]. Digestions were carried out for 16 h using sequencing grade trypsin obtained from Promega (Madison, WI, USA). Peptides from gel bands were extracted and hydrated in 20 µL of 5% formic acid. Peptides were then analyzed by LC-MS/MS as follows. Peptides were loaded at 5 µL/min onto a 200 µm × 50 mm precolumn packed with 5 µm YMC ODS-A C18 beads (Waters, Milford, MA, USA) using an Agilent 1100 series HPLC system (Agilent Technologies, Palo Alto, CA, USA). 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 an LTQ linear ion-trap mass spectrometer (Thermo-Electron, USA). MS/MS spectra were acquired in a data-dependant acquisition mode that automatically selected and fragmented the four most intense peaks from each MS spectrum generated. Peptide and MS/MS mass tolerances were set at ± 2 and 0.8 Da, respectively. MS/MS data were then analyzed and matched to human protein sequences in the NCBI database (nrdb) using the MASCOT database search engine (Matrix Science, UK) with carbamidomethyl as a fixed modification and oxidation, and ubiquitination (GG and LRGG remnant tags) as variable modifications. MS/MS spectra of ubiquitinated proteins were manually verified to confirm and map ubiquitination sites. A minimum peptide ion score cut-off of 20 as well as the presence of three consecutive *γ*- or *b*-ions were required.

3 Results and discussion

3.1 Identification of ubiquitination sites on a poly-Ub protein standard

Current MS-based approaches for identifying ubiquitination sites relies on the identification of signature peptides with an internal GG tag on ubiquitinated lysine residues following tryptic digestion (Fig. 1a). Our aim was to validate that signature peptides can be identified by GG and LRGG tags on internal lysine residues and to determine if tryptic digestion can cleave ubiquitinated lysine residues, as shown in Fig. 1b. To determine if tryptic digestion of ubiquitinated proteins produces both the GG and LRGG internal tags and the C-terminal GG tag, we used a poly-Ub standard containing isoforms of various chain lengths.

Ubiquitination of protein substrates often leads to the formation of various poly-Ub chains *in vivo* [1, 2, 4, 6], therefore, the use of poly-Ub for this experiment seemed a logical choice as it contains various ubiquitination sites. Poly- and mono-Ub standards were separated by SDS-PAGE (Fig. 2a), and gel bands corresponding to mono- and poly-Ub were excised and subjected to in-gel tryptic digestion. LC-MS/MS analysis of the tryptic peptides from poly-Ub resulted in the identification of ubiquitination sites on all seven lysine residues of Ub. No ubiquitination sites were identified from gel bands corresponding to mono-Ub (data not shown).

The conventional GG tag (Fig. 2b) was identified on all seven lysine residues of the poly-Ub standard, whereas the LRGG tag was identified on K11, K27, and K63 of the poly-Ub standard (Fig. 2c). Also of key importance was the finding that tryptic digestion of poly-Ub could occur after ubiquitinated lysine residues (Fig. 2d), as demonstrated by the identification of C-terminal GG tags on K27, K29, and K33. This is the first time that trypsin has been shown to cleave at ubiquitinated lysine residues, disproving the belief that ubiquitination sites always produce a missed cleavage site. It is important to note that MS/MS spectra were searched against the full NCBI human database and it is, therefore, highly unlikely that peptides identified with ubiquitination sites at their C-terminus were randomly matched to poly-Ub.

The proportion of ubiquitination sites identified by internal GG or LRGG tags and the C-terminal GG tag are shown in Fig. 3a. Although the majority of ubiquitination sites identified contained GG tags and missed cleavage sites, 23% of the total ubiquitination sites identified were found to contain an LRGG remnant tag, and 23% were identified to be cleaved at the ubiquitination site by trypsin. It is worth noting, however, that no C-terminal LRGG tags were identified. We suspect that the presence of these two additional amino acids (LR) may sterically block trypsin's active site, preventing cleavage at these lysine residues. Our results demonstrate that the identification of ubiquitination sites should not be limited to the analysis of GG tags on internal lysine residues as 46% more ubiquitination sites were identified when the additional mass tags were considered.

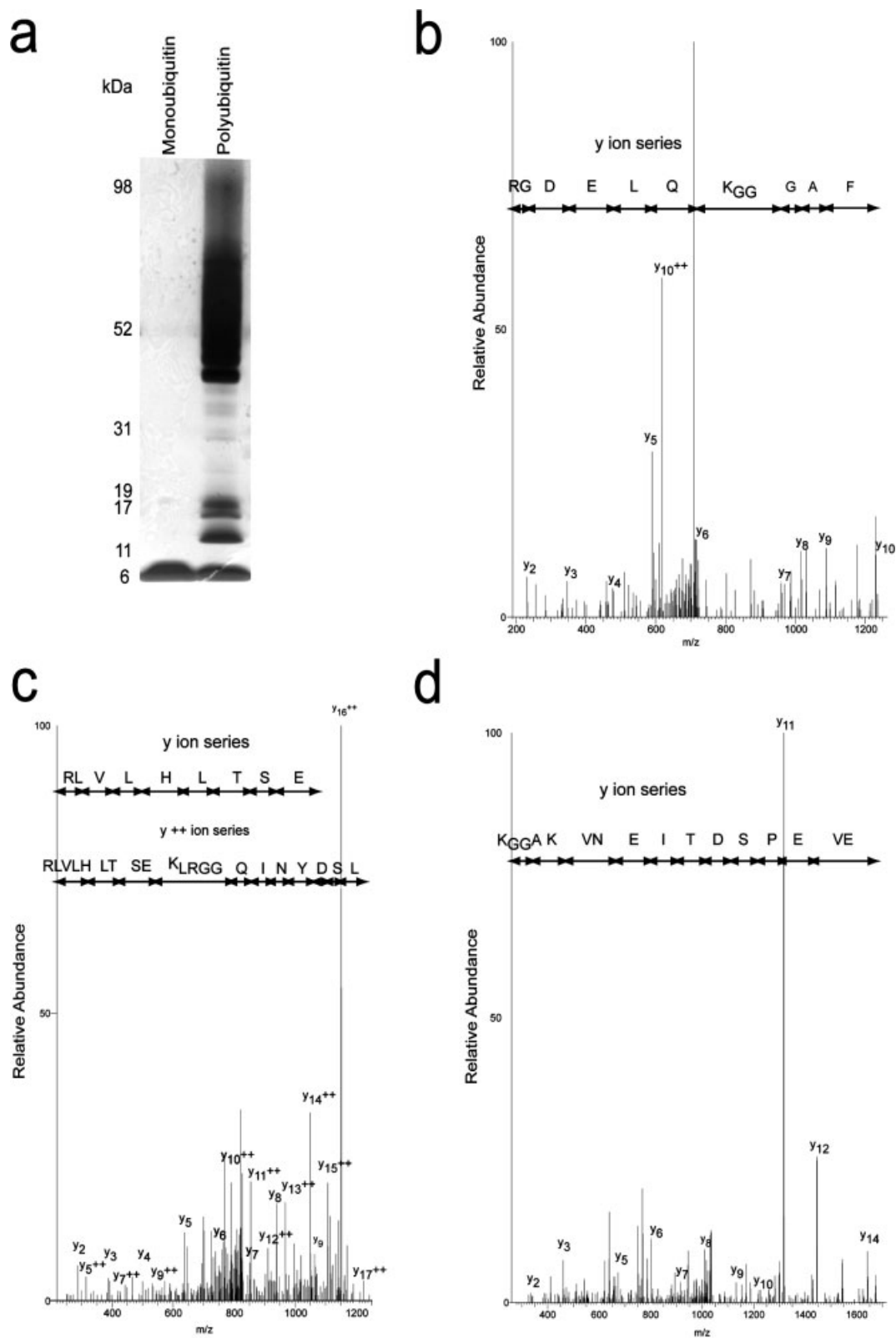


Figure 2. Ubiquitination sites on poly-Ub identified by MS/MS. (a) Mono- and poly-Ub standards were separated by SDS-PAGE and visualized by silver staining. (b–d) MS/MS spectra of ubiquitinated peptides from poly-Ub reveal the presence of an internal GG tag on K48 (ion score 54), an internal LRGG tag on K63 (ion score 43), and a C-terminal GG tag on K29 (ion score 60).

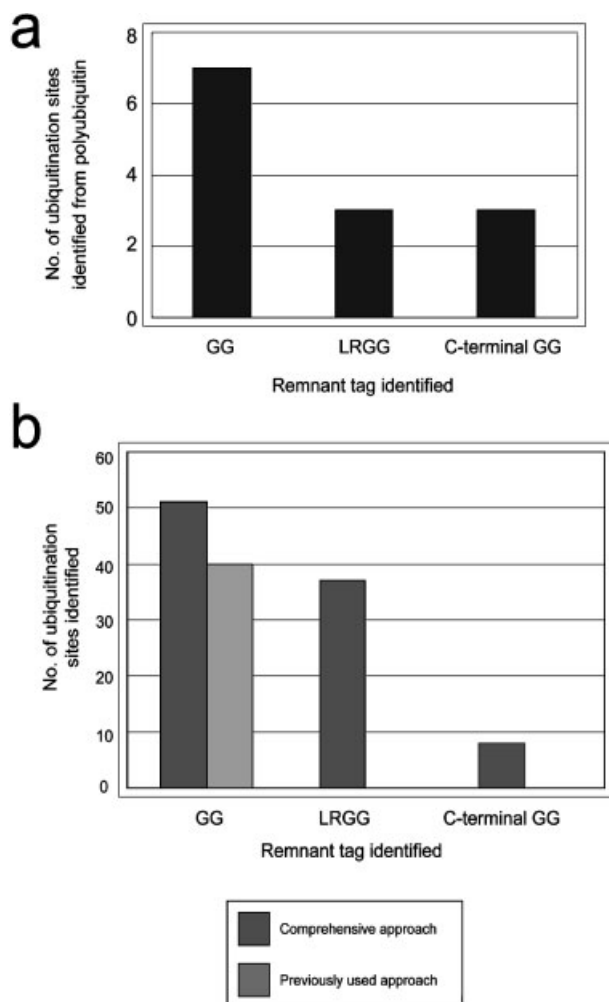


Figure 3. Proportional analysis of identified ubiquitination sites. (a) Bar graph showing the number of ubiquitination sites identified from a poly-Ub protein standard. (b) Bar graph showing the number of ubiquitination sites identified from immunopurified proteins from MCF-7 breast cancer cells.

3.2 Identification of ubiquitination sites from MCF-7 cell lysate

To test whether these additional mass tags could be used to identify ubiquitinated proteins from a complex biological sample, we analyzed a mixture of ubiquitinated proteins that were purified from MCF-7 breast cancer cells after treatment with the proteasome inhibitor MG132. MCF-7 cell lysates, containing 250 μ g of total protein, were subjected to immunopurification using an anti-Ub antibody, and sepa-

rated by SDS-PAGE. Proteins were subjected to in-gel tryptic digestion and LC-MS/MS analysis. MS/MS spectra of signature peptides containing the internal GG tag were identified by searching the NCBI human database. Using this approach, a total of 40 ubiquitination sites were identified. However, when the same MS/MS spectra were searched again using our comprehensive approach, that takes both GG and LRGG tags (Fig. 4a) into account, as well as the possibility of trypsin digestion at ubiquitinated lysine residues (Fig. 4b), a total of 96 ubiquitination sites were identified (Fig. 3b), resulting in a 2.4-fold increase in the number of sites that were identified over the previous strategies used to date. Of the 96 ubiquitination sites identified, 53% were identified through an internal GG remnant tag, while 39% were identified through a GG remnant tag found at the C-terminus of the peptide, which represents peptides that have undergone tryptic digestion at the ubiquitinated lysine residue; the remaining 8% of the ubiquitination sites were identified through an LRGG remnant tag. A complete list of the ubiquitination sites identified can be found in the Supplementary Data (Table S1).

These results demonstrate that our comprehensive strategy enabled the identification of an increased number of ubiquitination sites on proteins that were derived from a biologically-relevant sample. It should be noted that manual verification should be performed for all C-terminal ubiquitination sites to confirm that Asparagine (Asn) is not the following amino acid in the protein sequence, as this amino acid has the same mass as the GG tag.

4 Concluding remarks

In this study, we have confirmed that tryptic digestion of ubiquitinated proteins produces signature peptides with mass increases of 114.1 or 383.2 Da, due to GG or LRGG tags on specific lysine residues. We have also shown that ubiquitination sites can be found at the C-terminus of tryptic peptides, which demonstrates that trypsin can cleave after ubiquitinated lysine residues. Our comprehensive strategy for identifying ubiquitination sites can be applied to large-scale proteomic studies, and may result in an increased understanding of how the cell is regulated by ubiquitination and how the dysfunction of protein ubiquitination plays a role in the onset of various human diseases.

D. F. acknowledges funding from the Canada Research Chair, NSERC, MDS, and CFI.

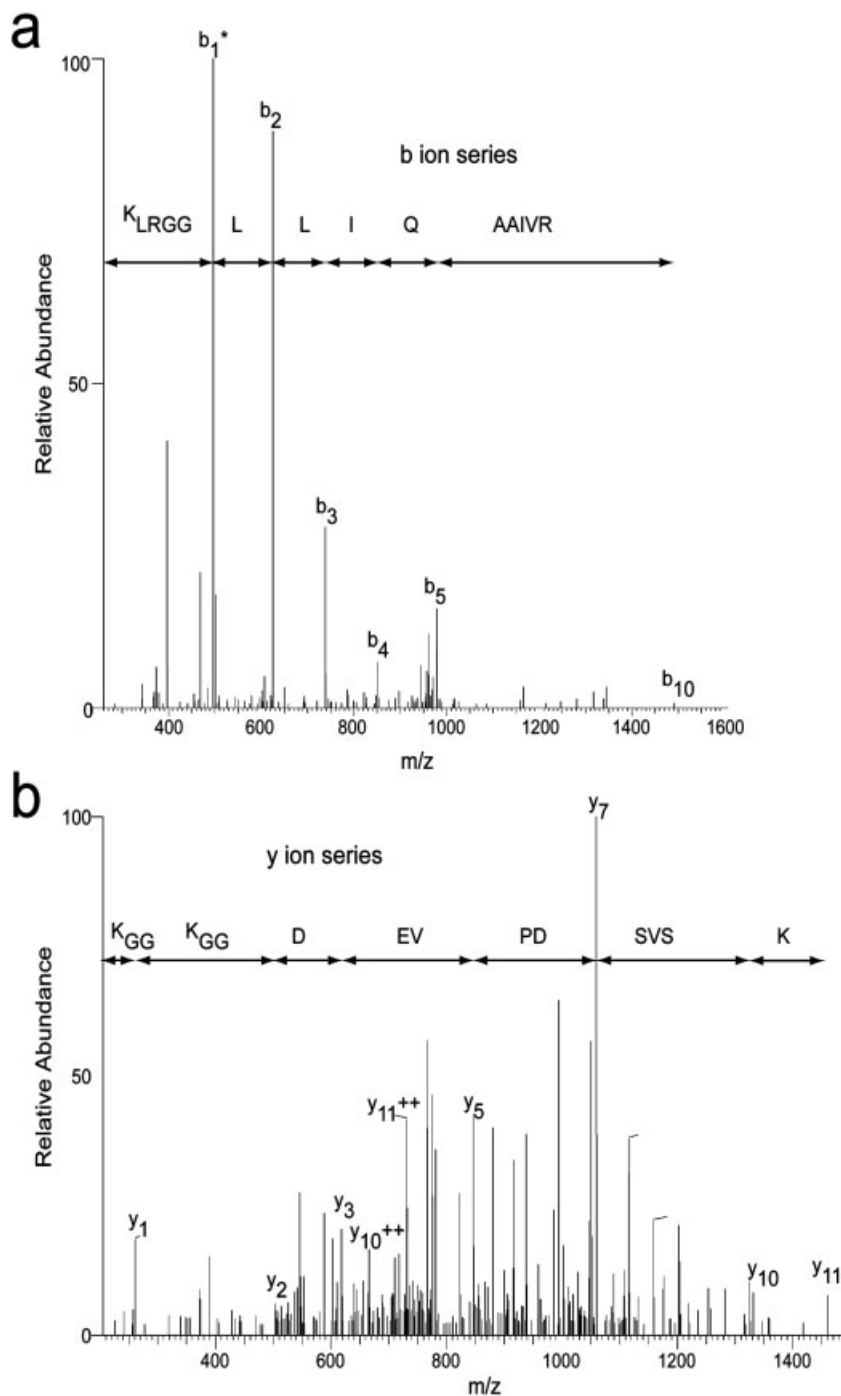


Figure 4. Ubiquitination sites on proteins from MCF-7 cells identified by MS/MS. (a) MS/MS spectra of a ubiquitinated peptide matching to a Ub ligase (Chain A of the Cul1-Rbx1-Skp1-F Boxskp2 Scf) containing an internal LRGG tag on K1 (ion score 21). b -ions that have lost an NH_3 group are denoted by b^* . (b) MS/MS spectra of a ubiquitinated peptide matching to the Fam44A protein containing an internal GG tag on K11 and a C-terminal GG tag on K12 (ion score 44).

5 References

- [1] Hochstrasser, M., *Annu. Rev. Genet.* 1996, 30, 405–439.
- [2] Hoppe, T., *Trends Biochem. Sci.* 2006, 30, 183–187.
- [3] Muratani, M., Tansey, W. P., *Nat. Rev. Mol. Cell Biol.* 2003, 4, 1–10.
- [4] Chen, Z. J., *Nat. Cell Biol.* 2005, 7, 758–765.
- [5] Welchman, R. L., Gordon, C., Mayer, R. J., *Nat. Rev. Mol. Cell Biol.* 2005, 6, 599–609.
- [6] Haglund, K., Dikic, I., *EMBO J.* 2005, 24, 3353–3359.
- [7] Eytan, E., Moshe, Y., Brauntstein, L., Hershko, A., *Proc. Natl. Acad. Sci. USA* 2006, 103, 2081–2086.
- [8] Thrower, J. S., Hoffman, L., Rechsteiner, M., Pickart, C. M., *EMBO J.* 2000, 19, 94–102.

- [9] d'Azzo, A., Bongiovanni, A., Nastasi, T., *Traffic* 2005, 6, 429–441.
- [10] Nandi, D., Tahiliani, P., Kumar, A., Chandu, D., *J. Biosci.* 2006, 31, 137–155.
- [11] Beckman, J. S., Maurer, F., Dolorenzi, M., Falquet, L., *Hum. Mutat.* 2005, 25, 507–512.
- [12] Lim, K.-L., Dawson, V. L., Dawson, T. M., *Neurobiol. Aging* 2006, 27, 524–529.
- [13] De Pril, R., Fischer, D. F., Van Leeuwen, F. W., *Neurobiol. Aging* 2006, 27, 515–523.
- [14] Ciechanover, A., *Neurology* 2006, 66, S7–S19.
- [15] Fuchs, S. Y., *Cancer Metastasis Rev.* 2005, 24, 329–338.
- [16] Chang, Y.-F., Cheng, C.-M., Chang, L.-K., Jong, Y.-J. *et al.*, *Biochem. Biophys. Res. Commun.* 2006, 342, 1022–1026.
- [17] Kirkpatrick, D. S., Denison, C., Gygi, S. P., *Nat. Cell Biol.* 2005, 7, 750–757.
- [18] Peng, J., Schwartz, D., Elias, J. E., Thoreen, C. C. *et al.*, *Nat. Biotechnol.* 2003, 21, 921–926.
- [19] Vasilescu, J., Smith, J. C., Ethier, M., Figeys, D., *J. Proteome Res.* 2005, 4, 2192–2200.
- [20] Matsumoto, M., Hatakeyama, S., Oyamada, K., Oda, Y. *et al.*, *Proteomics* 2005, 5, 1–7.
- [21] Esteban Warren, M. R., Parker, C. E., Mocanu, V., Klapper, D. *et al.*, *Rapid Commun. Mass Spectrom.* 2005, 19, 429–437.
- [22] Wilm, M., Shevchenko, A., Houthaeve, T., Breit, S. *et al.*, *Nature* 1996, 379, 466–469.