Chapter 27

Identification of Endogenous SUMO1 Accepter Sites by Mass Spectrometry

He-Hsuan Hsiao, Erik Meulmeester, and Henning Urlaub

Abstract

Posttranslational modification (PTM) by the covalent conjugation of small ubiquitin-like modifier (SUMO) plays an important role in many biological processes, such as cell cycle progression, transcriptional regulation, subcellular transport, and other processes. An in-depth understanding of the function of SUMOylation requires the discovery of SUMO accepter sites. However, identification of endogenous SUMO-conjugated sites in higher eukaryotes by MS-based proteomic strategies is hampered by the low abundance of SUMO conjugates, the large tryptic fragments of SUMO1 or SUMO2/3 and the inability to match MS/MS spectra by protein database search engine. In this chapter, we describe a powerful method to overcome at least some of these challenges. To identify SUMO acceptor sites in endogenous SUMO1 conjugated protein, the SUMO1 conjugates are purified by immunoprecipitation with anti-SUMO1 antibodies followed by SDS-PAGE separation and in-gel tryptic digestion. The resulting peptides are either performed using standard data dependent acquisition (DDA) for protein identification or high mass DDA to enhance the sensitivity of detection on the LTQ-Orbitrap mass spectrometer. Finally, a Web-based database tool, ChopNSpice, coupled with a protein database search engine is introduced to ease the identification of SUMO1 attachment sites. Although this method was initially used to identify SUMO1 accepter sites, it can be readily adapted to study SUMO2/3 conjugates or even other Ubiquitin-like modifiers.

Key words: ChopNSpice, Liquid chromatography, Mass spectrometry, Posttranslational modification, Proteomics, Small ubiquitin-like modifier

1. Introduction

The ability to identify posttranslational modifications (PTMs) of proteins and to exactly localize the sites of attachment is a critical step for understanding the biological meaning of modifications. Recently, mass spectrometry (MS) has emerged as a powerful technology for protein mapping studies, peptide and protein sequence

Katrin Marcus (ed.), *Quantitative Methods in Proteomics*, Methods in Molecular Biology, vol. 893,
DOI 10.1007/978-1-61779-885-6_27,

analysis, and identification of PTMs in proteins (for summary see Chapter 7), including phosphorylation (see Chapters 21 and 22), glycosylation (see Chapter 23), oxidation (see Chapter 24), nitrosylation (see Chapter 25) acetylation, ubiquitylation (see Chapter 26) and SUMOylation (1–4). However, the enormous complexity of the proteome and the large dynamic range of protein expression impede the analysis of PTMs by MS. The small ubiquitin-like modifier (SUMO) proteins are members of the ubiquitin-like family of proteins that are reversible posttranslational modifiers. Covalent conjugation of SUMO to its targets requires the attachment of the C-terminal glycine residue of SUMO to the ε-amino group of a lysine residue in the target protein. The modification of the target by SUMO subsequently regulates protein–protein interaction, or intra-molecular interactions, emerging as an important regulatory mechanism in a diverse range of cellular processes (5–9).

Global proteomic approaches to identify novel SUMO substrates from yeast *Saccharomyces cerevisiae* have begun to emerge in recent years (10–13). These studies generally rely upon affinity purification of SUMOylated proteins from yeast *S. cerevisiae*, followed by the identification of SUMO accepter sites using mass spectrometric techniques. The presence of amino acid sequence (–EQIGG) on the modified lysine residue after digestion of the SUMO conjugates by trypsin leads to a 484-Da increase in the mass of the peptide that can be readily measured by mass spectrometry. The MS/MS spectra of these SUMO-conjugated peptides can be identified by a protein database search that considers for this mass shift (484 Da) on the modified lysine residue. Unfortunately, such strategies are difficult to be applied for the identification of SUMO accepter sites in mammalians due to the larger distance of Arg or Lys residue to the C terminus of mammalian SUMO-1, -2, and -3. As a consequence, the large SUMO-conjugated peptides from mammalians that are generated after trypsin digestion impede their identification by database searches. Identification of such large conjugated peptides using MS/MS spectra is only possible as the peptides of substrates within the protein database are also modified by SUMO sequence. ChopNSpice (4) is a powerful software tool to automatically generate SUMO-conjugated sequences of proteins in silico that is compatible with current proteomics search engines such as Mascot (14) or SEQUEST (15) (see Chapter 28) for the identification of SUMO accepter sites.

In this chapter, we describe a method for identifying SUMOylation sites isolated from HeLa-S3 cells using LTQ-Orbitrap mass spectrometry. The general workflow for globally mapping human endogenous SUMO1 accepter sites is depicted in Fig. 1. Briefly, the purification of human SUMO1 conjugates is carried out by an immunoprecipitation with α-SUMO1 antibodies

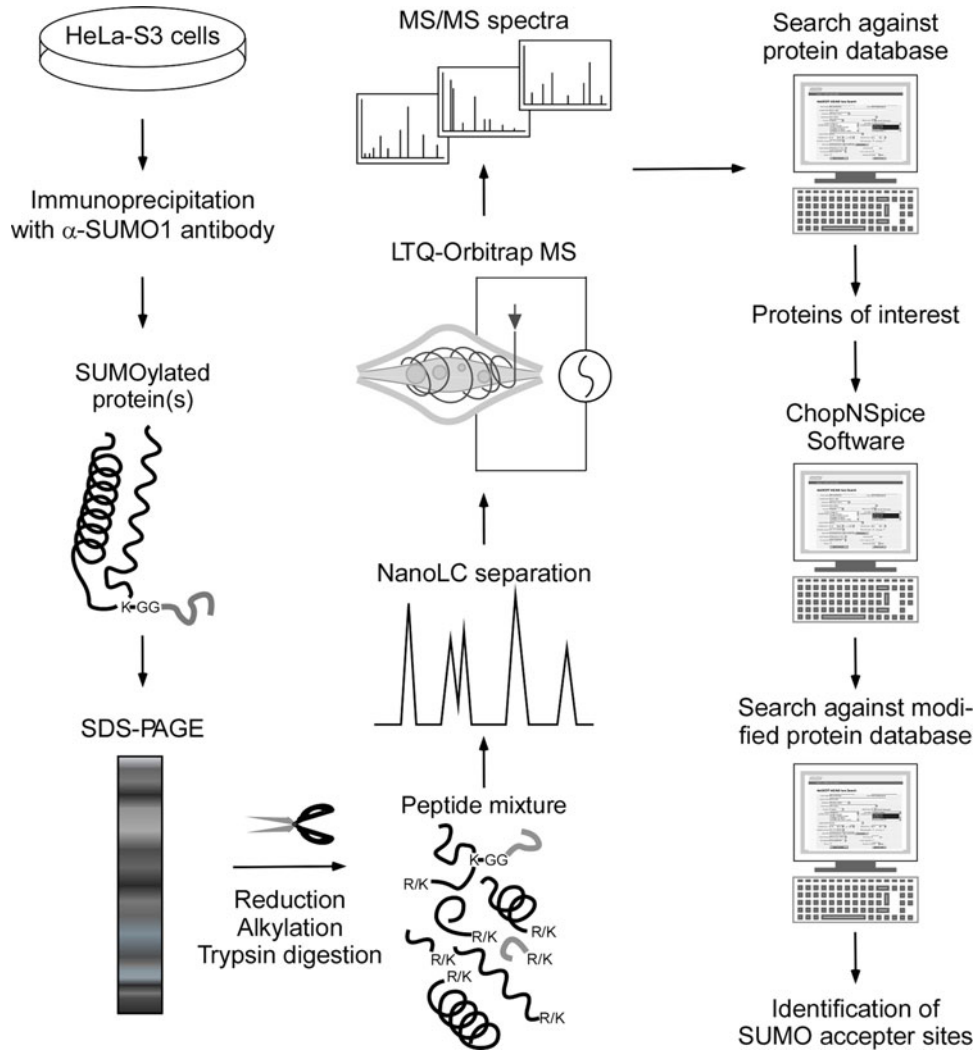

Fig. 1. Strategy for mapping endogenous human SUMO1 attachment sites.

from HeLa-S3 cells. The SUMOylated proteins are then separated by SDS-PAGE with Coomassie blue staining and cut equally into 20 slices. Each slice is reduced, alkylated, and subsequently digested with trypsin. Half of tryptic peptides are separated by reverse-phase nanoLC chromatography and detected with LTQ-Orbitrap mass spectrometer with standard data dependent acquisition (DDA). The resulting MS/MS spectra are searched against for protein identification. The proteins of interest found in the individual bands are performed a second run of MS/MS acquisition under high mass conditions, where only those precursor ions exceeding the mass of the SUMO1 C-terminal tryptic fragment (2,154 Da) are selected for sequencing. Both the MS/MS analyses

are merged into single peak list file for further protein database search. The proteins of interest are put into ChopNSpice software to create virtual SUMOylated protein sequences. The merging peak list file is resubmitted to search against modified protein database for identification of SUMO-conjugated sites.

2. Materials

2.1. Immunoprecipitation of SUMO1 Conjugates

1. Joklik's medium (Sigma-Aldrich, St. Louis, MO).
2. Fetal bovine serum (Invitrogen, Carlsbad, CA).
3. Antibiotics: Penicillin or Streptomycin stock solution (Merck, Darmstadt, Germany).
4. PBS: store at 4°C.
5. *N*-ethylmaleimide (NEM, Sigma-Aldrich).
6. TB: 10 mM PIPES, 55 mM manganese chloride ($MnCl_2$), 15 mM calcium chloride ($CaCl_2$), 250 mM potassium chloride (KCl).
7. Triton X-100 (Merck).
8. Adenosine triphosphate (ATP, Sigma-Aldrich).
9. Protease inhibitors (Sigma-Aldrich).
10. 0.45-μm filter (Millipore, Billerica, MA).
11. Monoclonal α-SUMO1 antibodies (GMP1, Zymed, South San Francisco, CA).
12. Protein G-agarose (GE Healthcare, Chalfont, UK).
13. Washing buffer: 20 mM NaP (pH 7.4), 150 mM NaCl, 1% Triton, 0.5% sodium deoxycholate, 0.1% SDS, 10 mM NEM.
14. Elution buffer: 2× SDS sample buffer.

2.2. Tryptic Digest of SUMO1 Conjugates and Extraction of SUMO1-Conjugated Peptides

1. Ultrapure water (LiChrosolv®).
2. Acetonitrile (ACN) (LiChrosolv®).
3. Ammonium bicarbonate (ABC) buffer: 25 mM ABC (Sigma-Aldrich), pH 8.5.
4. Dithiothreitol (DTT) solution: 25 mM DTT in 25 mM ABC, pH 8.5. Prepared freshly before use.
5. Iodoacetamide (IAA) solution: 50 mM IAA in 25 mM ABC, pH 8.5. Prepared freshly before use (see Note 1).
6. Trypsin stock solution: 0.1 μg/μL sequencing-grade endoprotease trypsin (Promega Corp., Madison, WI, sequencing grade) in 25 mM ABC, pH 8.5.
7. Extraction buffer: 50% ACN, 5% trifluoroacetic acid (TFA).

2.3. Determination of SUMO1 Conjugates Using NanoLC Coupled with Mass Spectrometry

1. Agilent 1100 HPLC system (Agilent Technologies, Santa Clara, CA).
2. Linear Ion Trap-Orbitrap mass spectrometer integrated with a nanoelectrospray ion source (LTQ-Orbitrap MS, Thermo Fisher Scientific, Waltham, MA).
3. HPLC mobile phase A (loading buffer): 0.1% formic acid (FA) in water.
4. HPLC mobile phase B: 0.1% formic acid and 95% ACN in water.
5. C18 trap column: 150 μm × 1.5 cm, 5 μm C18-AQ resin (Dr. Maisch GmbH, Ammerbuch-Entringen, Germany) in 360 μm O.D. fused silica capillary.
6. C18 analytical column: 75 μm × 10 cm, 5 μm C18-AQ resin (Dr. Maisch GmbH) in 360 μm O.D. fused silica emitter capillary.

2.4. Data Analysis and Identification of SUMOylation Site

1. Software to convert MS/MS spectra from the raw data to peaklist file (mgf file) using BioworksBrowser 3.3.1 SP1 (Thermo Fisher Scientific).
2. Software to identify SUMO conjugates, such as Mascot 2.2.07 (Matrix Science, London, UK).
3. Software, ChopNSpice, to automatically generate SUMO-conjugated sequences of proteins in silico (freely available online http://chopnspice.gwdg.de).

3. Methods

3.1. General Practice

To identify the modification for the ubiquitin-like proteins of interest requires the purification of the entire pool of conjugates, in our case SUMO1, from the cell line of choice. Subsequently the conjugates are separated by SDS-PAGE, digested with trypsin, the peptide components of that mixture are analyzed using nanoLC-Orbitrap mass spectrometry and finally the SUMO1-conjugated sites are identified using the Mascot search engine integrated with ChopNSpice software.

3.2. Immunoprecipitation of SUMO1 Conjugates

1. Maintain HeLa-S3 cells in Joklik's medium containing 10% fetal bovine serum and antibiotics.
2. Wash HeLa-S3 cells (1×10^9) twice with PBS containing 10 mM NEM.
3. Lyse in 2 pellet volumes of TB containing 0.1% Triton and 10 mM ATP supplemented with protease inhibitors and 10 mM NEM.

4. Centrifuge 16,000 × *g* for 15 min at 4°C.
5. Filter through a 0.45-μm filter.
6. Incubate the supernatant with 100 μg of monoclonal α-SUMO1 antibodies (GMP1) for 2 h at 4°C.
7. Centrifuge 16,000 × *g* for 15 min at 4°C.
8. Incubate the supernatant with protein G-agarose resins for 2 h at 4°C.
9. Wash the resins four times with 750 μL of washing buffer.
10. Elute bound proteins with 30 μL of elution buffer.
11. Separate the eluted proteins by SDS-PAGE followed by Coomassie blue staining (see Note 2 and Chapter 4).

3.3. Tryptic Digest of SUMO1 Conjugates and Extraction of SUMO1-Conjugated Peptides

1. Excise entire lane from SDS-PAGE into 20 gel slices (see Note 3).
2. Cut each slice into pieces of 1 × 1 mm using a scalpel and a spatula and transfer these pieces into 0.5-ml Eppendorf tube for tryptic digestion.
3. Wash the gel pieces with 400 μL ABC, spin the gel pieces down and remove all liquid.
4. Add 400 μL ACN to shrink the gel pieces, spin the gel pieces down and remove all liquid.
5. Add 60 μL DDT solution to reduce the disulfides and incubate in a thermomixer for 1 h at 37°C.
6. Spin the gel pieces down and remove all liquid.
7. Add 400 μL ACN to shrink the gel pieces, spin the gel pieces down and remove all liquid.
8. Add 60 μL IAA solution to alkylate the reduced cysteines and incubate in a thermomixer for 1 h at 37°C.
9. Repeat steps 6 and 7.
10. Add 60 μL trypsin solution (0.01 μg/μL) to digest proteins and incubate in a thermomixer at least for 16 h at 37°C (see Note 4).
11. Add 100 μL ACN to shrink the gel pieces, spin the gel pieces down and transfer all liquid to the other 0.5-ml Eppendorf tube.
12. Add 60 μL extraction buffer, vortex for 10 s, sonicate for 1 min, spin the gel pieces down, and transfer all liquid to the corresponding Eppendorf tube.
13. Repeat step 12 three times.
14. Add 60 μL ACN to shrink the gel pieces, spin the gel pieces down and transfer all liquid to the corresponding Eppendorf tube.
15. Repeat step 14.

16. Dry down the supernatant with a SpeedVac.
17. Store the dried sample at –20°C for further MS analysis.

3.4. Determination of SUMO1 Conjugates Using NanoLC Coupled with Mass Spectrometry

All NanoLC-MS/MS analyses were performed on a LTQ-Orbitrap XL mass spectrometer equipped with a nanoLC ESI source, connected to an Agilent 1100 LC system with an autosampler under the software control of Xcalibur 2.0.7.

3.4.1. Nano Liquid Chromatography

1. Dissolve the tryptic peptides with 4 μL of 0.1% FA in 50% ACN, vortex for 10 s and sonicate for 5 min (see Note 5).
2. Add 16 μL of 0.1% FA, vortex for 10 s and sonicate for 5 min.
3. NanoLC conditions: Following autosampler injection, the gradient was first held at 100% mobile phase A for 5 min, raised a 0.5 min linear gradient from 0 to 7.5% mobile phase B, a 37 min linear gradient from 7.5 to 37.5% mobile phase B, a 0.5 min linear gradient from 37.5 to 80% mobile phase B, 7 min at 80% mobile phase B, a 0.5 min linear gradient to 100% mobile phase A and finally held at 100% mobile phase A for 9.5 min.

3.4.2. Mass Spectrometry

In this step, parts of tryptic peptides separated by nanoLC are introduced into LTQ-Orbitrap MS through nanoelectrospray ion source. For protein identification, the eluted peptide ions are analyzed in a standard data-dependent acquisition. First, the m/z values of the peptides are detected and the most five intense ions are selected for collision-induced dissociation (CID) MS/MS, in which a peptide is fragmented in turn. The MS/MS spectra contain the peptide sequence information and are used to search protein database for protein identification. In addition, the remains of tryptic peptides are performed with high mass data-dependent acquisition. Under this condition, only peptides with a mass exceeding SUMO are triggered for MS/MS acquisition. This approach is highly sensitive for the detection and sequencing of larger SUMO conjugated-peptides.

1. Operate the LTQ-Orbitrap MS in the standard data-dependent acquisition to automatically trigger MS/MS scan by using the Tune and Xcalibur 2.0.7. software.
2. Set the following parameters in the Tune page:
   (a) Spray voltage at 1.8 kV.
   (b) Heated capillary temperature at 150°C.
   (c) Accumulation target value at 1×10^6 and 1×10^4 for MS and MS/MS scan, respectively.
   (d) Maximum injection time at 500 ms and 250 ms for MS and MS/MS scan, respectively.

3. Set the following parameters in the Xcalibur software:
   (a) A lock-mass ion from ambient air at m/z 445.120025 for internal calibration.
   (b) Full scan MS in the m/z range from 350 to 2,000 with a resolution $R = 30{,}000$ at m/z 400.
   (c) Five most intense multiply charged ions to be measured in the MS/MS scan.
   (d) Set the threshold for MS/MS scan at 1,000 counts.
   (e) Set normalized collision-induced dissociation (CID) collision energy at 37.5% for MS/MS in LTQ.
   (f) An activation $q = 0.25$.
   (g) Activation time = 30 ms.
   (h) Dynamic exclusion of up to 500 precursor ions for 60 s upon MS/MS scan and exclusion mass width of 10 ppm.
4. For high mass data-dependent acquisition, the settings are identical to standard data-dependent acquisition in addition to the followed parameters in the Xcalibur software.
   (a) Set an additional three microscans in LTQ MS/MS scan.
   (b) The mass range for selecting MS-data-dependent masses was set 2,154–1,000,000 for SUMO1-conjugated peptides, and the option of m/z value as masses had to be enabled (see Note 6).

3.5. Data Analysis and Identification of SUMOylation Site

Peak list files (mgf format) are generated from MS raw data by Bioworks browser, containing precursor masses and their corresponding product ion masses and intensities. The mgf file is used to match peptide sequence information by the Mascot search engine for protein identification. The proteins of interest are put into ChopNSpice software to create a new protein database. The search is repeated against the modified protein database and the SUMO acceptor sites can be identified.

1. Convert MS/MS spectra from the raw LC/MS/MS datafile to single mgf file using Bioworks browser (see Note 7).
2. Submit mgf file to Mascot search engine for protein identification with the following parameters: trypsin specificity; allow up to two missed cleavages; consider methionine oxidation and cysteine carboxyamidomethylation as variable modifications; instrument setting ESI-TRAP; mass tolerance of 10 ppm in MS mode and 0.8 Da in MS/MS mode.
3. The sequence of the protein of interest was manually saved as a Fasta file.

4. Submit the protein Fasta file to ChopNSpice software (http://chopnspice.gwdg.de) to create a new Fasta file with the following parameters:
   (a) Spice species: h. sapiens.
   (b) Spice: SUMO1.
   (c) Spice site: KX.
   (d) Spice mode: once per fragment.
   (e) Included unmodified fragments in output: yes.
   (f) Enzyme: Trypsin-N-term K or R, not C-term P.
   (g) Protein miscleavages min: 0.
   (h) Protein miscleavages max: 3.
   (i) Miscleavages in spice: 0–1.
   (j) Included minor end fragments: yes.
   (k) Output formatting: FASTA : Single protein sequence.
   (l) Cleavage marker: Mark all cleaved sites ("J").
   (m) Retain comments in FASTA format: yes.
   (n) Line breaks in FASTA output: yes.
5. Resubmit mgf file to Mascot search engine against a new Fasta file generated by ChopNSpice for identification of SUMO1-conjugated sites with the following parameters: enzyme cleaved at J at N- and C terminus; allow zero missed cleavages; consider methionine oxidation and cysteine carboxyamidomethylation as variable modifications; instrument setting ESI-TRAP; mass tolerance of 10 ppm in MS mode and 0.8 Da in MS/MS mode (see Note 8).

4. Notes

1. IAA is light sensitive and has to be stored in the dark after preparation.
2. A small aliquot of SUMO1 conjugates may be analyzed at this stage by SDS-PAGE followed by immunoblotting with α-SUMO1 antibodies (GMP1) to monitor the effectiveness of the purification.
3. Cutting the whole lane ensures that the entire sample will be analyzed.
4. The optimum trypsin amount is highly dependent on protein quantity. Experimenter has to adjust the amount accordingly.
5. The potential problem for analysis SUMO1 conjugates by MS arises from the adhesion between Eppendorf tube and for large

SUMO1-conjugated peptides. High percentage of organic solvent and sonication are essential to redissolve those large peptides.

6. Increase of additional microscans can improve the quality of MS/MS spectrum for mapping SUMO1-conjugated sites via database searching and high mass data-dependent acquisition enhance the sensitivity to identify the acceptor sites from more complex sample.
7. Mascot does not support to search MS/MS fragment ions with a charge state higher than 2, resulting in larger SUMO1-conjugated peptides with charge state of +4 or +5 show a low Mascot score or are not identified at all. MS/MS spectra recorded in the FT analyzer/detector of the Orbitrap MS with sufficient resolution for charge-state recognition have to use software tool Raw2msn (16) to deconvolute the higher charge stages of the fragment ions in the raw data to singly charged fragment ions for Mascot search that in turn increase the identification.
8. A new enzyme allowed cutting J at both C and N terminus have to be created in Mascot server. The file "quant_subs.pl" in the folder (mascot\cgi\) must be changed from $J \geq 0$ to $J \geq 0.05$ in line 3653. In addition, SUMO-conjugated sites also can be identified by using other search engines e.g., SEQUEST, X!Tandem, etc.

References

1. Mann M, Ong SE, Gronborg M et al (2002) Analysis of protein phosphorylation using mass spectrometry: deciphering the phosphoproteome. Trends Biotechnol 20:261–268
2. Kim SC, Sprung R, Chen Y et al (2006) Substrate and functional diversity of lysine acetylation revealed by a proteomics survey. Mol Cell 23:607–618
3. Meierhofer D, Wang X, Huang L, Kaiser P (2008) Quantitative analysis of global ubiquitination in HeLa cells by mass spectrometry. J Proteome Res 7:4566–4576
4. Hsiao HH, Meulmeester E, Frank BT et al (2009) "ChopNSpice," a mass spectrometric approach that allows identification of endogenous small ubiquitin-like modifier-conjugated peptides. Mol Cell Proteomics 8: 2664–2675
5. Seeler JS, Dejean A (2003) Nuclear and unclear functions of SUMO. Nat Rev Mol Cell Biol 4:690–699
6. Johnson ES (2004) Protein modification by SUMO. Annu Rev Biochem 73:355–382
7. Hay RT (2005) SUMO: a history of modification. Mol Cell 18:1–12
8. Geiss-Friedlander R, Melchior F (2007) Concepts in sumoylation: a decade on. Nat Rev Mol Cell Biol 8:947–956
9. Meulmeester E, Melchior F (2008) Cell biology: SUMO. Nature 452:709–711
10. Panse VG, Hardeland U, Werner T et al (2004) A proteome-wide approach identifies sumoylated substrate proteins in yeast. J Biol Chem 279:41346–41351
11. Denison C, Rudner AD, Gerber SA et al (2005) A proteomic strategy for gaining insights into protein sumoylation in yeast. Mol Cell Proteomics 4:246–254
12. Wohlschlegel JA, Johnson ES, Reed SI, Yates JR 3rd (2004) Global analysis of protein sumoylation in *Saccharomyces cerevisiae*. J Biol Chem 279:45662–45668

13. Hannich JT, Lewis A, Kroetz MB et al (2005) Defining the SUMO-modified proteome by multiple approaches in *Saccharomyces cerevisiae*. J Biol Chem 280:4102–4110
14. Perkins DN, Pappin DJ, Creasy DM, Cottrell JS (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis 20:3551–3567
15. Eng JK, Mccormack AL, Yates JR (1994) An approach to correlate tandem mass-spectral data of peptides with amino-acid-sequences in a protein database. J Am Soc Mass Spectrom 5:976–989
16. Olsen JV, de Godoy LM, Li G et al (2005) Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. Mol Cell Proteomics 4:2010–2021