Mass Spectrometric Analyses of Post-Translationally Modified Proteins

Dissertation

For the Award of the Degree

“Doctor of Philosophy” (Ph.D.)

Division of Mathematics and Natural Sciences

of the Georg-August-Universität Göttingen

Submitted by

He-Hsuan Hsiao

From Taipei, Taiwan

Göttingen 2010

Members of Thesis Committee

Dr. Henning Urlaub (Instructor)

Bioanalytical Mass Spectrometry Group

Max Planck Institute for Biophysical Chemistry

Prof. Dr. Ralf Ficner (Reviewer)

Department of Molecular Structural Biology

Georg-August-University Göttingen

Prof. Dr. Frauke Melchior (Reviewer)

ZMBH

University of Heidelberg

Date of Oral Examination: August 9, 2010

Affidavit

I hereby declare that this dissertation "Mass Spectrometric Analyses of Post-Translationally Modified Proteins" has been written independently and without unauthorized assistance. This dissertation has not been submitted elsewhere for any academic award or qualification.

He-Hsuan Hsiao

July, 2010

Göttingen, Germany

Acknowledgement

First of all, I would like to express my deepest thankfulness to Dr. Henning Urlaub for the opportunity to accomplish my PhD in his laboratory and for funding and supervising this work. Thanks for thinking for me and working wholeheartedly on my manuscripts for publication among his busiest time. He is an greatly kind man and provides many chances to his students for being engaged in their scientific researches.

I thank Prof. Dr. Ralf Ficner of the University of Göttingen and Prof. Dr. Frauke Melchior of University of Heidelberg for serving on my thesis committee, and giving the suggestion of my work.

I am grateful to my colleagues – Uwe Pleßmann, Monika Raabe, Johanna Lehne, Mads Grønborg, Florian Richter, Carla Schmidt, Miroslav Nikolov, Katharina Kramer, Ilian Atanassov, Romina Hofele and Ling Yun. Your expertises in biology, mass spectrometry and bioinformatics have broaden my horizon. It is a joy working with all of you.

Thanks to my collaborators – Prof. Dr. Reinhard Lührmann and the people in his laboratory who have directly contributed to the phosphoproteomic project in spliceosome; Prof. Dr. Jürgen Wienands and Thomas Oellerich on the SLP65 project; Prof. Dr. Markus Wahl and Xiao Luo on the NusB-S10-RNA interaction project; Prof. Dr. Frauke Melchior, Dr. Erik Meulmeester and Benedikt Frank on the SUMO project.

Finally, I would especially like to thank my parent, brothers, girlfriend Chia-Yan Wu and all my friends in Taiwan for being the stone of support that I have built my life upon. I share my success to all of the wonderful people who have supported, assisted, and loved me during my PhD study at University of Göttingen. Without these support, I would not have achieved so much.

Table of Contents

Members of Thesis Committee……I

Declaration……II

Acknowledgement……III

Table of Contents……IV

List of Abbreviations……VIII

Summary……1

Chapter 1 - General Introduction……3

1.1 Proteomics……3

1.2 Proteomic Analysis Techniques……4

1.3 Separation Technology……4

1.3.1 Sodium dodecyl sulfate polyacrylamide gel electrophoresis……5

1.3.2 Two-dimensional gel electrophoresis……5

1.3.3 Chromatography……6

1.3.4 Affinity-Based Enrichment for Sub-Proteome……6

1.4 Mass Spectrometry……7

1.4.1 Ion Sources……8

1.4.2 Mass Analyzer……9

1.4.2.1 Quadrupole Mass Analyzer……10

1.4.2.2 Ion Trap Mass Analyzer……10

1.4.2.3 Time-Of-Flight Mass Analyzer……11

1.4.2.4 Fourier-Transform Ion-Cyclotron Resonance Mass Analyzer……11

1.4.2.5 Orbitrap Mass Analyzer……11

1.4.3 Tandem Mass Spectrometry……12

1.5 Data Analysis……13

Chapter 2 - A High-Throughput Method for Phosphopeptide Enrichment of Spliceosomal Proteins and Its Application……15

2.1 Summary……15

2.2 Introduction……15

2.3 Experiment Sections……19

2.3.1 Materials……19

2.3.2 A Crude Mixture of Nuclear snRNP Particles, Individual U snRNP, Spliceosomal Complexes and SR Proteins Purification……19

2.3.3 Ethanol Precipitation……..……19
2.3.4 In-Solution Digestion……...19
2.3.5 In-Gel Digestion……20
2.3.6 In-House TiO_2 Microspin Column Fabrication……20
2.3.7 Comparison of Different TiO_2 Enrichment Procedure……20
2.3.8 NanoLC-ESI and -MALDI Mass Spectrometry Analysis……21
2.3.9 Interpretation of Tandem Mass Spectra……...22
2.4 Result and Discussion……22
2.4.1 TiO_2 Microspin Column……...22
2.4.2 Sensitivity of the TiO_2 Microspin Column……23
2.4.3 Optimizing Phosphopeptide Enrichment with TiO_2 microspin column for In-Solution Digestion……..……25
2.4.4 Global Profiling of Phosphopeptides from A Crude Mixture of Nuclear snRNP Particles, Individual U snRNP, Spliceosomal Complexes and SR Proteins…….29
2.4.5 Novel and Known Kinase Motifs in Spliceosomal Proteins……...34
2.4.6 Application of TiO_2 Microspin Column……..36
2.5. Conclusion……40

Chapter 3 - Efficient Enrichment of Intact Phosphoproteins prior to Mass Spectrometric Analysis……42
3.1 Summary……..……42
3.2 Introduction……42
3.3 Experiment Sections……44
3.3.1 Materials……44
3.3.2 U1 snRNPs Purification……..44
3.3.3 Ethanol Precipitation……...44
3.3.4 Acetone Precipitation……..44
3.3.5 Optimized Calcium Phosphate Precipitation……...44
3.3.6 In-Solution Digestion……..……45
3.3.7 TiO_2 Enrichment Procedure……...45
3.3.8 Mass Spectrometry Analysis……..45
3.3.9 MASCOT Database Searching.,……46
3.4 Result and Discussion……46
3.4.1 Comparison of Efficiency and Specificity among Different Alkaline Earth Metal ……..……..46
3.4.2 Optimal Condition for Phosphoprotein Precipitation……..48
3.4.3 Effect of Different Denatured Reagents……50
3.4.4 Examine the Power of our Phosphoprotein Isolation Method……...53
3.4.5. Identification of Phosphorylation Sites from U1 Small Nuclear

Ribonucleoproteins……55
3.5. Conclusion……58

Chapter 4 - Pseudo-Neutral-Loss Scan for Selective Detection of Phosphopeptides and *N*-Glycopeptides using Liquid Chromatography Coupled with a Hybrid Linear Ion-Trap / Orbitrap Mass Spectrometer……59
4.1 Summary……59
4.2 Introduction……59
4.3 Experiment Sections……62
4.3.1 Materials……62
4.3.2 In-Solution Digestion……62
4.3.3 Mass Spectrometry……62
4.4 Result and Discussion……64
4.4.1 The Principle of Pseudo-Neutral-Loss……64
4.4.2 Identification of Phosphopeptide by Pseudo-Neutral-Loss Scan……65
4.4.3 Identification of *N*-Glycopeptide by Pseudo-Neutral-Loss Scan……69
4.5 Conclusion……71

Chapter 5 - "ChopNSpice", a Mass Spectrometric Approach That Allows Identification of Endogenous Small Ubiquitin-like Modifier-conjugated Peptides……73
5.1 Summary……73
5.2 Introduction……73
5.2.1 SUMOylation……73
5.2.2 Challenges for Identification of SUMOylation by MS-based Approach……75
5.3 Experiment Sections……76
5.3.1 Materials……76
5.3.2 Software……77
5.3.3 In vitro SUMOylation Assays……77
5.3.4 Cell culture, Immunoprecipitation and Immunoblotting……77
5.3.5 Protein Digestion……77
5.3.6 Liquid Chromatography and Mass Spectrometry……78
5.3.7 Data Analysis……78
5.4 Result and Discussion……79
5.4.1 ChopNspice……79
5.4.2 Identification of SUMO Conjugation Sites in vitro……84
5.4.3 Increasing Sensitivity by using "High Mass" Acquisition……86
5.4.4 Identification of SUMO-Conjugated Sites in vivo……87
5.5 Conclusion……89

References……91

Appendixes……103

Appendix 1. Phosphorylation sites identified from spliceosomal proteins by using in-house TiO_2 microspin column enrichment……103

Appendix 2. MS and MS/MS spectra of phosphopeptides derived from human PRP6 and PRP31……120

Appendix 3. Phosphorylation sites identified from U1 snRNPs by using CPP method in combination with in-house TiO_2 microspin column enrichment……124

Appendix 4A. MASCOT searching result against swissprot bovine database by using regular data-dependent acquisition……127

Appendix 4B. MASCOT searching result against swissprot bovine database by using pseudo-neutral-loss acquisition……131

Curriculum Vitae……132

List of Abbreviations

2DE, two-dimensional gel electrophoresis

3D ion trap, three dimensional ions trap

AC, alternate current

ACN, acetonitrile

BCR, B-cell receptor

CAA, chloroacetamide

CHCA, α-cyano-4-hydroxy-cinnamic acid

CID, collision-induced dissociation

CPP, calcium phosphate precipitation

DC, direct current

DDA, data-dependent acquisition

DeoxyHex, deoxyhexose

DHB, 2,5-dihydroxybenzoic acid

DTT, dithiothreitol

ESI, electrospray ionization

FA, formic acid

FT-ICR, fourier-transform ion-cyclotron resonance

Glu-Fib, [Glu]-Fibrinopeptide B

HEPES, 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

Hex, hexose

HexNAc, *N*-acetylhexosamine

IAA, Iodoacetamide

IMAC, immobilized metal ion affinity chromatography

kDa, kilodalton

LC, liquid chromatography

m/z, mass-to-charge ratio

MALDI, matrix-assisted laser desorption / ionization-time of flight mass spectrometer

µg, microgram

µL, microliter

mg, milligram

min, minute

MOAC, metal oxide affinity chromatography

MS, mass spectrometry

MS/MS, tandem mass spectrometry

MudPIT, multi-dimensional protein identification technology

nL, nanoliter

NeuAc, *N*-acetylneuraminic acid

NeuGc, *N*-glycolylneuraminic acid

NusB, N utilization substance protein B

p53, cellular tumor antigen p53

PA, phthalic acid

PMF, peptide mass fingerprinting

ppm, parts per million

pre-mRNA, precursor messenger ribonucleic acid

PRP31, pre-mRNA-processing factor 31

PRP6, pre-mRNA-processing factor 6

PTMs, post-translational modifications

pSer, phosphoserine

pThr, phosphothreonine

pTyr, phosphotyrosine

Q-TOF, quadrupole-time-of-flight

RanGAP1, Ran GTPase-activating protein 1

RF, radio frequency

RNA, ribonucleic acid

RP, reverse phase

rpm, rounds per minute

S10, 30S ribosomal protein S10

SAC, strong anion exchange

SCX, strong cation exchange

SDS, sodium dodecyl sulphate

SDS-PAGE, sodium dodecyl sulfate polyacrylamide gel electrophoresis

SLP-65, B-cell linker protein

snRNPs, small nuclear ribonucleoproteins

Sp100, nuclear autoantigen Sp-100

SR proteins, serine/arginine-rich proteins

SUMO, small ubiquitin-like modifier

TFA, trifluoroacetic acid

TiO_2, titanium dioxide

TOF, time-of-flight

Uba2, SUMO-activating enzyme subunit 2

USP25, ubiquitin carboxyl-terminal hydrolase 25

UV, ultraviolet

Summary

Protein post-translational modifications (PTMs) possess key functions in the regulation of various cellular processes. In this thesis, several new technologies are developed to map protein PTMs, containing phosphorylation, glycosylation and SUMOylation.

Five major topics are presented in this thesis. **In Chapter 1- General Introduction**, describes what is proteomics, the importance of protein PTMs, the main techniques in proteomic analysis, including the principle of technologies for protein and peptide separation, the theory of mass spectrometry (MS) and concept of proteomic data analysis.

In Chapter 2 - A High-Throughput Method for Phosphopeptide Enrichment of Spliceosomal Proteins and Its Application, a disposable TiO_2 microspin column is fabricated in-house for enrichment of phosphopeptides. The method offers several advantages, including high-throughput, easy to use, low cost, high selectivity and sensitivity. In combination with different proteomic strategies, 1381 unique phosphorylation sites corresponding to 390 distinct proteins were identified in spliceosomal proteins. We further applied this method to explore the phosphorylation sites on PRP6, PRP31 and SLP65 and to study protein-RNA interaction by UV-induced crosslinking reaction.

In Chapter 3 - Efficient Enrichment of Intact Phosphoproteins prior to Mass Spectrometric Analysis, a straightforward and reliable phosphoprotein purification procedure was developed based on calcium phosphate precipitation (CPP). Integration of TiO_2 microspin column, a total of 192 unique phosphorylation sites corresponding to 45 distinct proteins were identified from the U1 small nuclear ribonucleoproteins (snRNPs); of these, 59 phosphorylation sites were not reported previously.

In Chapter 4 - Pseudo-Neutral-Loss Scan for Selective Detection of Phosphopeptides and *N*-Glycopeptides using Liquid Chromatography Coupled with a Hybrid Linear Ion-Trap / Orbitrap Mass Spectrometer, a pseudo-neutral-loss scan on a hybrid LTQ-Orbitrap MS was built up for selectively probing phosphopeptides and glycopeptides. The presence of known characteristic mass pair (phosphoric acid or monosaccharide residues) in the spectrum during in-source collision-induced dissociation (CID) was selected to trigger MS/MS and multi-stage activation MS3 fragmentation. Our method is compatible with nano-liquid chromatography (nano-LC) for separation of complex peptide mixtures without any further enrichment. The consequent spectra provide peptide sequence identification and modified site assignment as well as information of the glycan structure.

In Chapter 5 - "ChopNSpice", a Mass Spectrometric Approach That Allows Identification of Endogenous Small Ubiquitin-like Modifier-conjugated Peptides, a novel, user-friendly and straightforward database search tool was developed, called "ChopNSpice", to unambiguously determine the mammalian SUMO1 and SUMO2/3 conjugation sites in vitro and in vivo by mass spectrometry in combination with MS-based search engines like MASCOT or Sequest. High mass data dependent acquisition (DDA) is highly suitable for the accurate detection and sequencing of larger peptides and additionally facilitates detection of lower abundant SUMO-conjugates. We demonstrated the power of ChopNSpice software in combination with proteomic strategy, resulting in the identification of 10 SUMOylated proteins corresponding to 17 distinct sites in endogenous Hela-S3 cells. 15 SUMOylated sites were identified in this study appeared to be novel, which may provide a valuable resource to the biological research community.

Chapter 1 - General Introduction

1.1 Proteomics

The Human Genome Project and its sister projects for other organisms mark the culmination of twentieth-century biology and have been a tremendous success, rapidly building up a new scope of scientific landscape for the century [1, 2]. Following in the footsteps of genomics, the next step has been the development of proteomics. The term “proteome” was coined by Marc Wilkins in 1995 to describe the total set of proteins encoded by a genome. The word “proteome” was defined as “the PROTEin complement expressed by a genOME” [3, 4]. Proteomics is the study of proteome and was first coined by Peter James in 1997 to make an analogy with genomics [5]. Proteins, the main carriers of biological activity. The function of protein depends on the precise amino acid sequence, the modifications, the protein concentration, the association with other proteins, and the extracellular environment. Accordingly, the proteomics is concerned with determining protein structure, modifications, protein expression levels, protein-protein interactions, localization, and cellular roles of as many proteins as possible.

Proteins are converted to their mature forms through a complicated succession of post-translational processing and decoration events, namely post-translational modifications, PTMs. Far from being merely decoration, PTMs of proteins control many biological processes, and examining their diversity and individual functions are critical for understanding mechanisms of cell regulation. As many as 300 PTMs of proteins are known to occur physiologically (http://www.unimod.org/). PTMs are covalent processing events that change the properties of a protein by proteolytic cleavage or by addition of a modifying group to one or more amino acids. Many of the PTMs are regulatory and reversible, most notably are protein acetylation, methylation, phosphorylation, ubiquitination, SUMOylation, and etc., which control biological function through a multitude of mechanisms. It is known that many enzymes and receptors are switched "on" or "off" by phosphorylation and dephosphorylation, such as the enzyme, glycogen synthase kinase-3 (GSK-3), which is phosphorylated by protein kinase B (PKB) as part of the insulin signaling pathway [6]. Ubiquitin is added to proteins as a tag that predestines them for proteolytic degradation [7]. Small ubiquitin-like modifier (SUMO) is found to be attached on many eukaryotic nuclear proteins which functions in the regulation of nuclear-cytosolic transport, transcriptional regulation, apoptosis, protein stability, response to stress, and progression through the cell cycle [8].

Despite the great importance of PTMs for biological functions, large scale studies have been hampered by lacking suitable methods. Mass spectrometry (MS) is currently the most versatile technology to directly determine PTMs due to its sensitivity and selectivity. However, the identification of PTMs still remains a substantial challenge owing to the low stoichiometry of

modified proteins in combination with large amounts of unmodified proteins in biological sample, which interfere the detection in MS. Hence, direct analysis of PTMs requires isolation of the correctly processed proteins in a significant amount followed by MS-based proteomics, we believe this will lead to a great contribution for the study of protein PTMs.

1.2 Proteomic Analysis Techniques

The extraordinary achievements of current proteomics are based largely on the successful developments in the fields of “separation technology”, “mass spectrometry” and “data analysis”. Once joined, the three disciplines provided a powerful tool to study the proteomics shown in Figure 1.1.

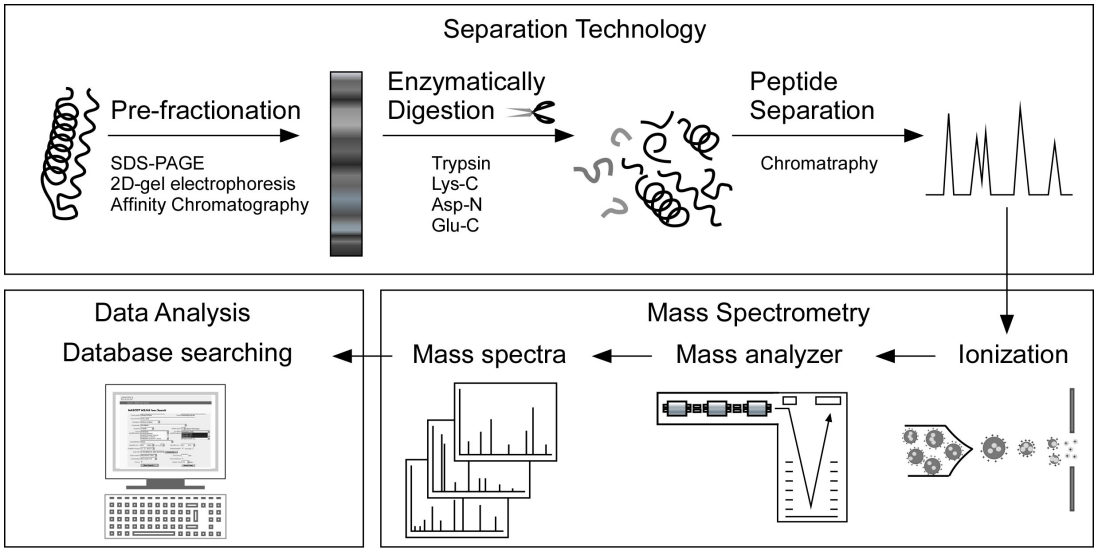


Figure 1.1. Mass spectrometry-based proteomic strategy. The proteins are pre-fractionation, typically separated by SDS-PAGE. The gel lane is cut into several slices, which are then degraded enzymatically into peptides, where C-terminus are protonated amino acids (arginine or lysine) when trypsin is used, providing an advantage in subsequent peptide sequencing. The peptides are separated by one or more steps of chromatographic technologies. The eluted peptides are ionized by either ESI or MALDI. Finally, the resulting mass spectra are searched against protein database to obtain the information of peptide sequences.

1.3 Separation Technology

Mass spectrometry-based proteomics is highly dependent and tightly linked to separation technologies that simplify incredibly complex biological samples prior to MS analysis would be crucial. The detection of low abundance species is required front-end separation due to the overshadowed signal of high abundance species. To reduce the complexity, proteins or

peptides can be resolved into fractions by using various separation methods, including electrophoretic techniques (SDS-PAGE, 2D gel electrophoresis), multi-dimensional chromatography (size exclusion, ion exchange, reverse phase chromatography) and affinity purification.

1.3.1 Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE)

SDS-PAGE is a technique widely used in the separation of proteins according to their electrophoretic mobility (molecular weight) [9, 10]. The proteins have identical charge per unit mass due to the binding of SDS. Since the charge density is constant, the rate of migration depends on the resistive frictional force, thus small proteins migrate faster than big proteins. The distance traveled in a fixed time period is a log function of the molecular weight.

1.3.2 Two-dimensional gel electrophoresis

From the mid 1970s, proteomics was pursued with two-dimensional gel electrophoresis (2DE). The proteins in a sample are separated by isoelectric point and protein mass Each observed protein spot is quantified by its staining intensity. Selected spots are excised, digested and analyzed by mass spectrometer for protein identification shown in Figure 1.2.

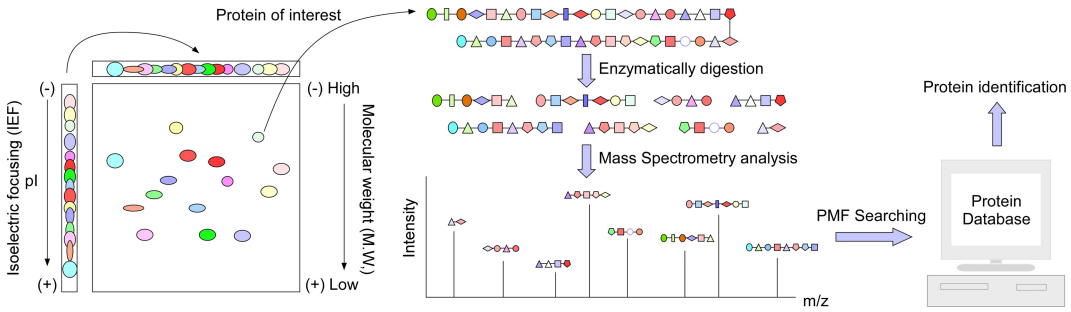


Figure 1.2. Protein identification by two-dimensional gel electrophoresis (2DE) coupling with mass spectrometric peptide mass fingerprinting (PMF). The biological sample is first separated by isoelectric point, called isoelectric focusing (IEF). In the second dimensional separation, an electric potential is again applied, but at a 90 degree angle from the first field, separating the proteins on the basis of their molecular weight. The protein of interest is digested by adequate enzyme, and then analyzed by mass spectrometer. The resulting masses of the peptides of the unknown protein are then compared to the theoretical peptide masses of each protein in the protein database for protein identification, called peptide mass fingerprinting.

2DE has been a mature technique for more than 35 years [11] and was the first technique capable of supporting the concurrent quantitative analysis of large numbers of gene product

[12, 13]. Peptide mass fingerprinting (PMF) coupled with matrix-assisted laser desorption / ionization-time of flight mass spectrometer (MALDI-TOF MS) has become highly efficient at the identification of 2DE separated proteins [14-18]. In this method, the unknown protein of interest is first enzymatically digested into peptides, where masses can be accurately measured with a mass spectrometer. These masses are then compared to the theoretical peptide masses of each protein calculated from a known protein database. The results are statistically analyzed to find the best match. For PMF, the matching of four peptides representing 10 % of the sequence does not constitute a reliable hit and should not be listed as a positive identification. In contrast, six peptides representing 20 % of the sequence may be adequate for a tentative identification. There are several shortcomings for this approach. 2DE has limited dynamic range. The protein sequence has to be present in the database of interest. The presence of a mixture can significantly complicate the analysis and potentially compromise the results. Therefore, protein identification based solely on PMF should no longer be acceptable and must be complemented by tandem mass spectrometry (MS/MS) to achieve sufficient specificity of identification.

1.3.3 Chromatography

Gel-based technologies have been traditionally used with off-line MALDI analysis. In contract, the multi-dimensional chromatography is usually directly coupled to on-line electrospray ionization (ESI) analysis, a continuous separation, due to their buffers compatibility with ESI. Two major chromatographic materials are widely used for separating peptide mixtures, a reverse phase (RP) and a strong cation exchange (SCX). The reverse phase material separates proteins or peptides based on their hydrophobicity, given high resolution, efficiency, reproducibility. However, the single dimension of separation might not provide sufficient peak capacity to separate peptide mixtures as complex as those generated by the proteolysis of protein mixtures, for example, total cell lysates. Another material, SCX, is integrated with reverse phase as the part of a two-dimensional chromatography, improving the resolving power of separation based on peptide charges and hydrophobicity interaction. This technique is known as multi-dimensional protein identification technology (MudPIT) [19, 20]. High complexity sample is first loaded onto an SCX, and is eluted in a series of salt concentration steps. Each eluted fraction is loaded onto an RP column either off-line or directly eluted into an ESI mass spectrometer. The MudPIT analysis is subdivided the sample into several independent MS runs, which increases the confidence of protein identification and the dynamic range of the measurement.

1.3.4 Affinity-Based Enrichment for Sub-Proteome

Another important separation technology is affinity purification which is often used to isolate proteins with PTMs based on a highly specific biological interaction such as that between

antigen and antibody, enzyme and substrate, or receptor and ligand. The low stoichiometry of PTMs such as phosphorylation, glycosylation and acetylation requires enrichment steps prior to MS analysis. Antibody-based affinity enrichment is widely used for the detection of PTMs in proteins. The antibody purification approach has been successfully used for global analysis of protein lysine acetylation [21], arginine methylation [22], tyrosine phosphorylation [23] and so on. In addtion, the N-glycosylated proteins can be isolated by lectins [24] and the protein containing phosphotyrosine can also be purified by protein containing Scr homology 2 (SH2) domains [25]. However, highly specific antibodies, receptor and substrate are not always available for PTMs of interest. The other affinity enrichment method is based on chemical derivation of the modifying group that derives a “tag” for affinity purification. For example, the O-phosphorylated residues [26] and O-GlcNAc residues [27] (serine or threonine) can be modified an affinity tag by a beta-elimination / Michael addition reaction. Special issues should be taken care of due to the possible loss of the low abundant peptides with PTMs and false-positive protein identification of side-products of chemical reactions

1.4 Mass Spectrometry

By fundamental definition, mass spectrometry is designed to measure the mass-to-charge ratio (m/z) of gas phase ions. It consists of an ion source that converts analyte molecules into gas phase ions, a mass analyzer that separates the m/z of the ionized analytes, and a detector that records the number of ions at each m/z value, as shown in Figure 1.3.

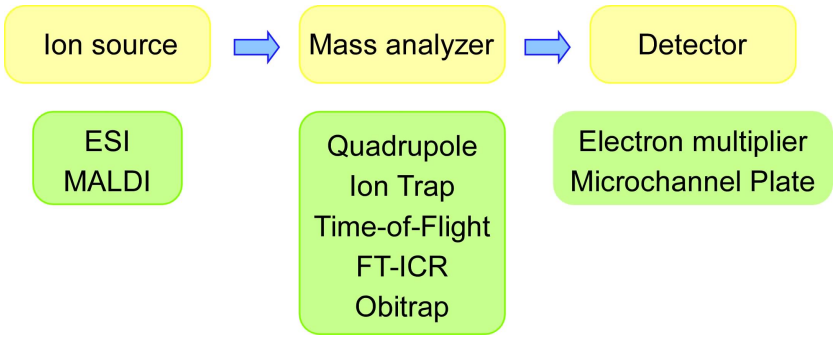


Figure 1.3. Main components of a typical mass spectrometry. Ion source for ion generation, mass analyzer for ion separation and detector to transform analogue signals into digital signals and record a mass spectrum.

Biological mass spectrometry, the technological base of current proteomics studies, was first catapulted to mainstream prominence with the development of ion sources, the electrospray ionization (ESI) [28] and matrix-assisted laser desorption ionization (MALDI) [29-31] techniques. The most notable the discovery and development of protein ionization methods are recognized by the 2002 Nobel Prize in Chemistry (http://nobelprize.org/). Widespread

mass analyzers are quadrupole, ion trap, time-of-flight (TOF), Fourier-transform ion-cyclotron resonance (FT-ICR) and Orbitrap which can be coupled with either ESI or MALDI ion source.

1.4.1 Ion Sources

The development of electrospary ionization (ESI) and matrix-assisted laser desorption ionization (MALDI), the two soft ionization techniques capable of ionizing peptides and proteins, revolutionized protein analysis using MS as shown in Figure 1.4.

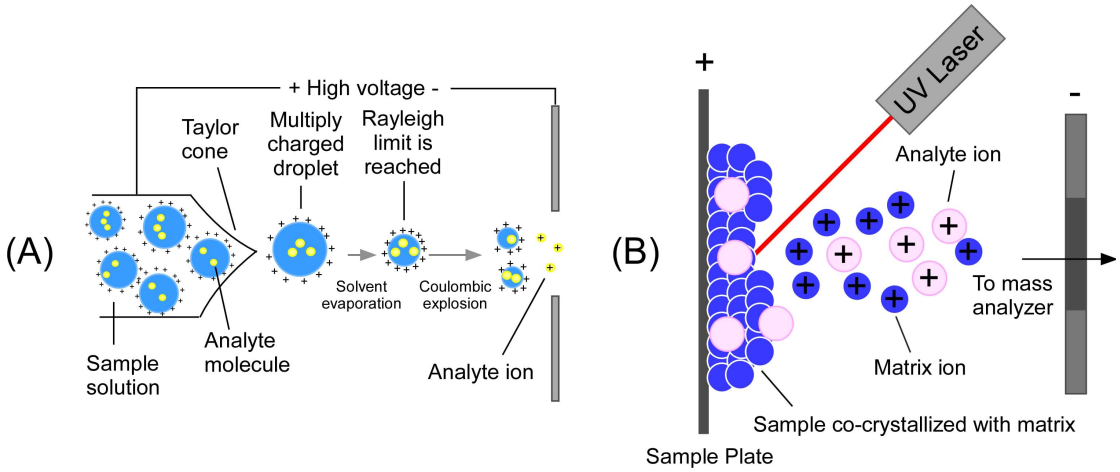


Figure 1.4. The common ionization sources for proteomic research. (A) The scheme of the electrospray ionization (ESI) process. (B) The scheme of matrix-assisted laser desorption ionization (MALDI) process.

The ESI source produces ions from solution at atmospheric pressure shown in Figure 1.4A. ESI is driven by high voltage applied between the end of the capillary column and the inlet of the mass spectrometer. The processes of ESI involve creation of electrically charged spray, Taylor cone [32, 33], followed by formation of an aerosol of charged droplets and desolvation of analyte-solvent droplets. Eventually, ions become free of the solvent that surrounds them, and these ions make their way by voltage-driven into the mass analyzer of the mass spectrometer. The ESI ion formation are the multiply charged species and sensitivity to the analyze concentration and flow rate. Multiply charged ions enable mass spectrometers with limited m/z ranges to analyze higher molecular weight molecule. ESI ionizes the analytes out of a solution and is therefore readily coupled to liquid-based (for example, chromatographic and electrophoretic) separation tools.

Unlike ESI, the MALDI source produces ions from solid phase shown in Figure 1.4B. MALDI relies on the utilization of a matrix compound capable of absorbing ultraviolet (UV) light. The matrix and sample are mixed in the appropriate solvent and deposited onto a sample plate. The solvent is evaporated, forming co-crystallized analyte-matrix molecules. MALDI

sublimates and ionizes the samples out of a dry, crystalline matrix via UV laser pulses. The MALDI matrix absorbs laser energy and transfers the energy to the analyte, whereas the rapid laser heating causes desorption of matrix and analyte ions into gas phase [31, 34]. MALDI is normally used to analyze relatively simple samples, whereas ESI integrated with liquid chromatography (LC) is preferred for the analysis of complex samples. Although MALDI can still be coupled to LC, the effluent from LC run must be deposited on a sample plate and mixed with the MALDI matrix, a process that has thus far proven difficult to automate [35].

1.4.2 Mass Analyzer

The mass analyzer is, literally and figuratively, central to mass spectrometry. For proteomics research, five basic types of mass analyzers are commonly used: quadrupole, ion trap, time-of-flight (TOF), Fourier-transform ion-cyclotron resonance (FT-ICR) and Orbitrap mass analyzer shown in Figure 1.5. They are very different in design and performance, each with its own strength and weakness. These analyzers can be stand alone or, in some cases, put together in tandem to take advantage of strengths of each. In the context of proteomics, key parameters of mass analyzer are sensitivity, resolution, mass accuracy and the ability to generate information-rich ion mass spectra from peptide fragment (tandem mass or MS/MS spectra).

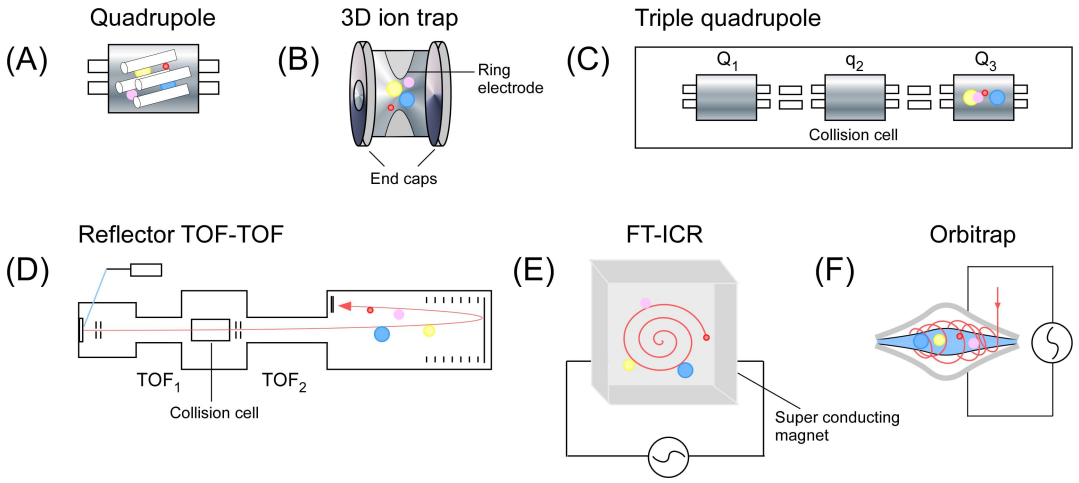


Figure 1.5. Mass spectrometers used in proteomic research. (A) Quadrupole mass spectrometer, the ions are separated by time varying electric fields between four rods, permitting a stable trajectory only for the ions of a particular desired m/z. (B) Three-dimensional ion trap mass spectrometer, the ions maintain a stable trajectories inside the device as a result of the application of a radio frequency voltage to the ring electrode. Mass analysis is achieved by making ion trajectories unstable in a mass-selective manner. (C) Triple quadrupole mass spectrometer, the ions of a particular

m/z are selected in a first quadrupole (Q_1), fragmented in a collision cell (q_2), and the fragment ions are separated in the last quadrupole (Q_3). (D) TOF-TOF mass spectrometer, the ions of different m/z values have different velocities and therefore reach the detector at different times. It incorporates a collision cell between two TOF sections. Ions of one m/z are selected in the first TOF, fragmented in the collision cell, and the fragment ions are separated in the second TOF. (E) Fourier-transform ion-cyclotron resonance (FT-ICR) mass spectrometer, the ions oscillate around the magnetic field at frequencies that are related to their m/z. As ions oscillate near the top and bottom metal plates of the cubic trapping cell, they induce an alternating current that can be measured and then transferred to m/z. (F) Orbitrap mass spectrometer, the ions are trapped in its static electrostatic fields, in which the ions orbit around a central electrode and oscillate in axial direction, converting time domain signal into m/z like FT-ICR.

1.4.2.1 Quadrupole Mass Analyzer

The quadrupole is a mass filter, consisting of four rods to which direct current (DC) and a radio frequency (RF) alternate current (AC) are applied. Only ions carry a certain m/z can pass through, and reach the detector; while others are on unstable trajectories and fail to reach the detector. When these DC and RF voltages are increased, maintaining their ratios constant, ions with increasing value of m/z are recorded at the detector [36, 37]. Up to now, most experiments have been performed on triple quadrupole mass spectrometer that consists of three parts. Two mass separating quadrupoles was divided by a central quadruple whose function is to fragment the selected ion. Due to the presence of two independent quadrupoles, the triple quadrupole can be programmed for a variety of different scan modes, product ion scan, precursor ion scan, neutral loss scan and multiple reaction monitoring [38].

1.4.2.2 Ion Trap Mass Analyzer

The principle of three dimensional ions trap (3D ion trap) is a close relative of the quadrupole mass analyzer. Whereas a quadrupole has electric fields in two dimensions (x and y direction) and the ions move perpendicular to the field (z direction), the 3D ion trap has the electric field in all three dimensions, which can result in ions being trapped in the field. Unlike quadrupole, the spectrum is obtained by increasing the RF voltage that makes ions unstable and ejects for detection. For a MS/MS acquisition, all ions except the selected ion are ejected first, Subsequently, the remaining ion is fragmented and the product ions are analyzed [39, 40]. Ion trap is a robust, sensitive and relatively inexpensive instrument, which has successfully acquired much proteomics data in the literature. A disadvantage of ion trap is their relatively low mass accuracy due to the limited number of ions that can be accumulated, The space charge in ion trap distorts the accuracy of the mass measurement. Owing to the operating

principle of the ion trap, the lower end of the fragment mass range cannot be detected (1/4 low mass cut off) [40].

1.4.2.3 Time-Of-Flight (TOF) Mass Analyzer

A TOF analyzer separates the ions based on velocity. It can be thought of as a race from the same starting point to the detector. Theoretically, all ions are formed at the same time and place in the ion source. Subsequently, the ions are accelerated to a fixed kinetic energy and travel down a flight tube. The low m/z ions achieve higher velocity than the high m/z ions. The spectrum is recorded by the impact of each ion on the detector. In fact, ion velocity is inversely related to the square root of m/z [41].

1.4.2.4 Fourier-Transform Ion-Cyclotron Resonance (FT-ICR) Mass Analyzer

The FT-ICR mass spectrometer is also a trapping mass analyzer. It captures the ions under high vacuum in a high magnetic field. Once trapped, the ions oscillate with a cyclotron frequency that is inversely related to their m/z. The trapped ions are excited by an electric RF with a frequency in resonance with their cyclotron frequency. Although the ion oscillation ratio increases, its frequency is maintained, generating the image current for detection. The frequencies related to m/z can be calculated by a complex mathematical procedure (Fourier-transform, FT). As the frequencies can be measured precisely, the high-resolution and high-precision mass measurement is achieved under high magnetic field [42].

1.4.2.5 Orbitrap Mass Analyzer

The recent development of a novel Orbitrap mass spectrometry has made the exciting new areas of proteomic application possible. The Orbitrap was invented by Alexander Makarov in 1999 [43] and was reported as a tool for proteomics research in 2005 by Hu et al. [44]. In the Orbitrap analyzer, the ions are trapped and the orbit around a central spindle-like electrode and oscillate harmonically along its axis with a frequency characteristic of their m/z values, inducing an image current in the outer electrodes that is processed by Fourier-transform and generates the mass spectrum. The frequency of these harmonic oscillations is independent of the ion velocity and is inversely proportional to the square root of the m/z. The instrument is capable of mass resolution in excess of 100000 and mass accuracy of less than 2 ppm.

To summarize, MALDI is usually coupled to TOF mass analyzer, whereas ESI has mostly been coupled to ion traps, triple quadrupole, FT-ICR and Orbitrap mass analyzer. More recently, other new combinations of mass spectrometry are developed. For example, ESI quadrupole-TOF (Q-TOF) consists of ESI ion source coupled to the TOF analyzer [45]. MALDI TOF-TOF, MALDI ion source has been coupled to two types of TOF instruments. In the first, second TOF sections are separated by a collision cell [46]. These mass spectrometers have

high sensitivity, resolution and mass accuracy, and can be widely used in proteomic research.

1.4.3 Tandem Mass Spectrometry

Tandem mass spectrometry (MS/MS), as the name implies, involves two stages of MS. In the first stage, ions of a desired m/z are isolated (precursor ions). The isolated ions are increased the internal energy, and are induced to collide with an inert gas such as helium, argon or nitrogen, leading to dissociation. The resulting ions (product ions) are analyzed with the second stage of MS. MS/MS is a key technique for protein or peptide sequencing and PTMs analysis. Collision-induced dissociation (CID) has been the most widely used MS/MS technique in proteomics research. In this method, a particular gas phase peptide/protein ions is isolated and subsequently the energy is imparted by collisions with inert gas. The energy causes the peptide to break apart on the peptide backbone. Figure 1.6 shows how the peptide fragment and how the fragment ions are designated.

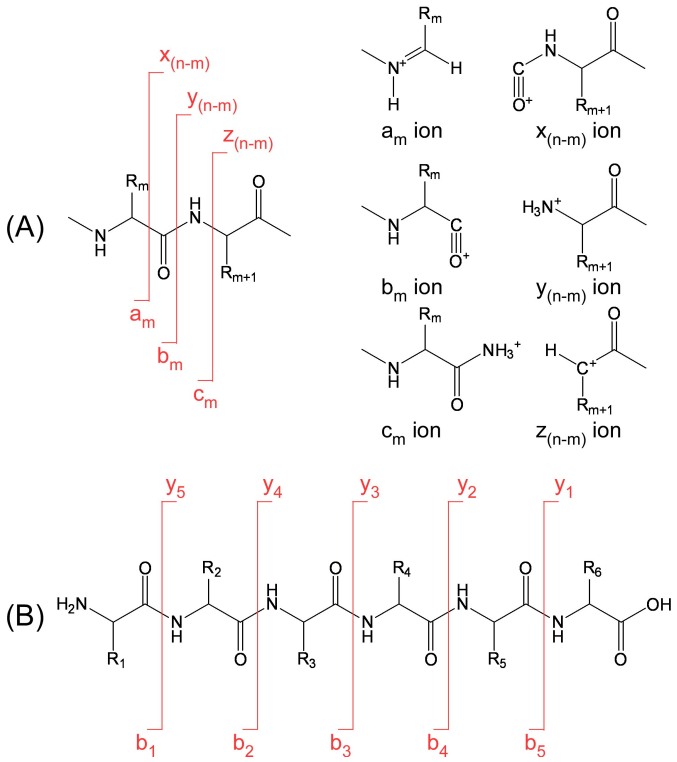


Figure 1.6. The types of peptide fragment ions observed in a MS/MS spectrum. (A) When the charge is localized on the N-terminus, the ion is classed as a_m, b_m or c_m. When the charge is localized on the C-terminus, the type of ion is classed as $x_{(n-m)}$, $y_{(n-m)}$ or $z_{(n-m)}$. (B) The fragment of peptide is induced by collisions with inert gas, and the bond breakage mainly occurs in the lowest energy pathway. That is cleavage of the amide bonds, which leads to form b- and y-ions series.

The most common and informative ions are generated by fragmentation at amide bond between amino acids. The resulting ions are called b-ions if the charge is localized on the N-terminal part of the peptide and y-ions if the charge is localized on the C-terminal part [47, 48]. The common proteomic experiment is performed with trypsin digestion .The resulting peptides have arginyl or lysyl residues at their C-terminus. In this case, the y-ion series are the predominant in the spectra. For an even more in-depth characterization, the fragment ions of peptide can be further fragmented. This is known as MS^3 or more generally, MS^n. It is worth to

note that the internal fragmentation and neutral losses of H_2O, NH_3, and labile PTMs occur frequently during the CID process.

1.5 Data Analysis

As mentioned before, MALDI-TOF MS is still much used to identify protein by what is known as peptide mapping, also referred to as peptide mass fingerprinting (PMF), due to its simplicity, excellent mass accuracy, high resolution and sensitivity. In this method, protein is identified by matching a list of experimental peptide masses with the calculated list of all peptide masses of each entry in a protein database (SwissProt, NCBInr and etc.). Due to the fact that mass mapping requires an essentially purified target protein, the technique is commonly used in combination with protein separation technology using either SDS-PAGE or 2DE, respectively. However, protein identification based solely on PMF should no longer be acceptable and must be complemented by MS/MS spectra. Because the MS/MS spectra contain structural information related to the sequence of the peptide, in addition to precursor ion mass information, these searches are generally more specific and discriminating. Several MS/MS spectra searching engines exist such as Sequest [49] and MASCOT [50]. The MS/MS spectra are collected as many as possible, and the results are searched by an algorithmic comparison via Sequest or MASCOT towards a protein database. These methods do not attempt to extract any sequence information at all from the MS/MS spectrum. Instead, the experimental fragment spectrum is matched against a calculated fragment spectrum for all peptides in the database. A score is given to determine how close between MS/MS spectrum and the calculated peptide sequence. Sequest, a cross-correlation method, peptide sequences in the database are used to construct theoretical MS/MS spectra. The overlap or cross-correlation of these predicted MS/MS spectra with the experimental MS/MS spectra determines the best match. MASCOT, a probability based matching, the calculated fragment ion from peptide sequences in the database are compared with experimental MS/MS spectra. From this comparison, a score is calculated which reflects the statistical significance of the match between the MS/MS spectrum and the sequence contained in the database. In each of these methods, the identified peptides are compiled into a protein hit list, which is the output of a typical proteomic experiment. The protein identification relies on the matches with sequence database, and not all peptides resulting from the enzymatic digestion of a protein can be observed or correctly identified with MS analysis. This would reflect especially on the peptide with unexpected PTMs, and high-throughput proteomics is currently limited largely due to those species for which comprehensive sequence databases are available.

In brief, proteomics is the large-scale study of proteins, particularly their structures and functions. MS-based proteomics has become a formidable tool for the identification of proteins. The limitation in dynamic range of MS analysis only allows for the proteins present at high

relative abundance to be preferentially identified, while information regarding the proteins present at low abundance in the complex mixtures is commonly not detected. Hence, the development of separation technology and the continued improvement of mass spectrometric methodology is crucial for identifying the low abundance proteins.

Chapter 2 - A High-Throughput Method for Phosphopeptide Enrichment of Spliceosomal Proteins and Its Applications

2.1 Summary

Reversible protein phosphorylation is a ubiquitous post-translational modification critical to many cellular processes. In this study, a simple, inexpensive and convenient titanium dioxide (TiO_2) microspin column fabricated in a commercial pipette tip was developed for high-throughput enrichment of phosphopeptides in a crude mixture of spliceosomal proteins, which were digested by trypsin. The spliceosome is a protein-RNA complex which catalyses the excision of introns and ligation of exons of eukarytic pre-mRNAs. Our approach allows the enrichment of twenty-four samples at once. Careful comparison of our novel high-throughput method with the previously described manual TiO_2 enrichment techniques showed similar result in terms of selectivity and sensitivity. Additionally, we evaluated and optimized the titania-based affinity enrichment for global profiling of phosphopeptides in total small nuclear ribonucleoproteins (snRNPs). We found that the use of RapiGest™ SF as detergent during digestion was more efficient than urea. The non-specific binding of non-phosphorylated peptides on TiO_2 materials was reduced, but still maintained the high binding affinity of phosphopeptides without the need for an additional desalting step. Approximately 70 % of the enriched peptides were identified by mass spectrometry as being phosphorylated. Furthermore, a complementary integrated analytical platform involving a combination of in-solution digestion, in-gel digestion from SDS-PAGE, TiO_2 microspin columns, on-line nanoLC ESI-MS and off-line nanoLC MALDI-MS was employed to discover the maximum number of phosphorylation sites in the human spliceosomal proteins present in the crude mixture of nuclear snRNP particles. These strategies allow the complementary measurement of phosphopeptides. When compared with off-line nanoLC MALDI-MS/MS, online LC-ESI MS/MS turned out to be better for determining the exact location of the phosphorylation site. In total, 1381 phosphopeptides were identified in 390 proteins; of these, 640 sites were not previously described. The list of phosphopeptides was used to extract known and novel kinase motifs using the Motif-X algorithm. Finally, we showed three applications of this methodology for identifying phosphorylation sites and for studying protein-RNA crosslinks.

2.2 Introduction

A spliceosome is a complex of specialized ribonucleic acid (RNA) and protein subunits that removes introns from a transcribed precursor messenger ribonucleic acid (pre-mRNA) segment. The process is generally referred as splicing. Each spliceosome is composed of small nuclear RNA proteins, called snRNPs, and a range of non-snRNP associated protein factors. The snRNPs that make up the nuclear spliceosome are named U1, U2, U4, U5, and

U6, and participate in several RNA-RNA and RNA-protein interactions. The snRNPs display a broad variety of molecular sizes and chemical properties such as Arg-/Ser-rich tracts. They are ideally suited to establish methods for analysis of the changes in the phosphorylation during splicing process. Human spliceosome assembly intermediates have been observed include the E, A, B, B*, and C complex shown in Figure 2.1 [51]. The first recognition of pre-mRNA involves U1 snRNP binding to the 5' end splice site of the pre-mRNA and other non-snRNP associated factors to form the E complex. Subsequently, the U2 snRNP tightly associates with the branch point sequence (BPS) with the E complex to form A complex. The U4/U6.U5 tri-snRNP stably interacts to the assembling spliceosome to form complex B. Following several rearrangements in RNA-RNA and RNA-protein interactions, detaching the U1 and U4 snRNPs, give rise to the catalytically activated spliceosome (B* complex) and then converts into C complex, in which the first of the two catalytic steps of splicing has occurred. After the second step, the spliceosome dissociates and the snRNPs are recycled for repeated rounds of spliceosome assembly.

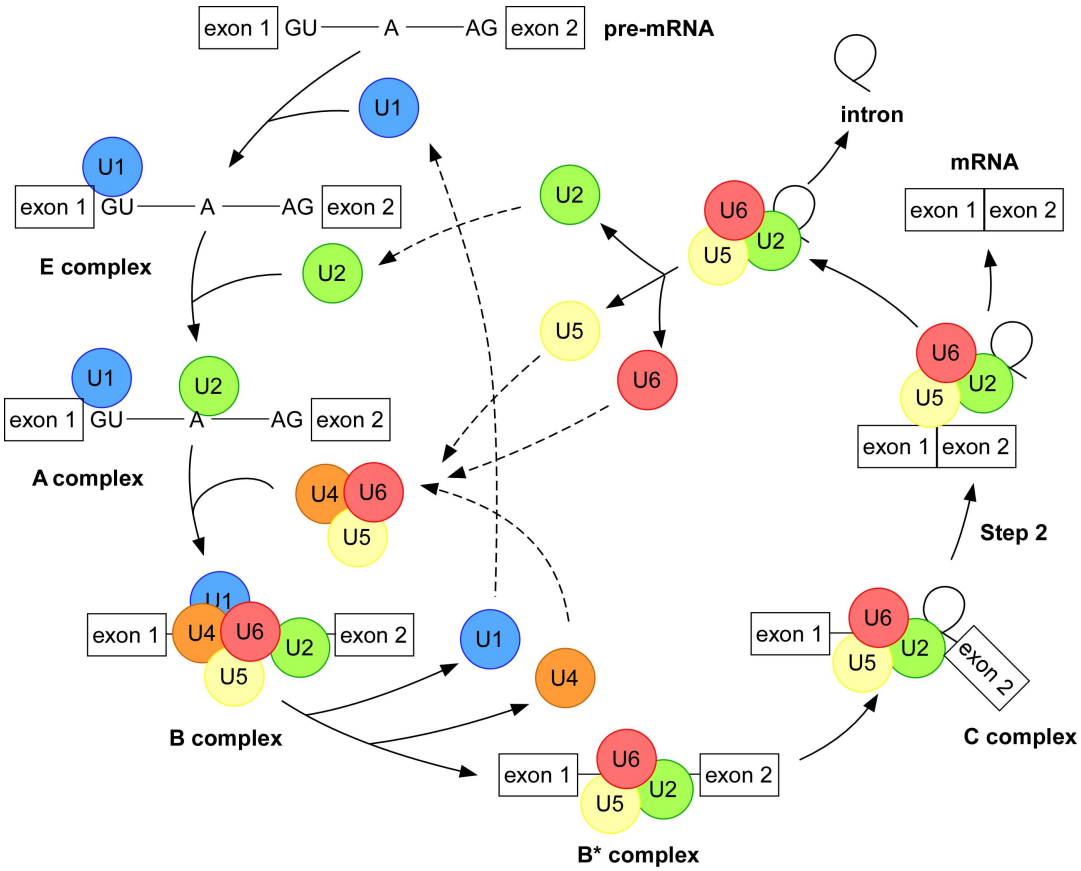


Figure 2.1. The pathway of spliceosome assembly.

PTMs of spliceosomal proteins play a crucial role in triggering conformational changes in protein-protein and RNA-protein interactions during the spliceosome assembly that are

essential for its activity. For example, pre-mRNA splicing can be regulated both positively and negatively by reversible protein phosphorylation [52]. The phosphorylation of SR proteins are mediated by protein kinases / phosphatases and have been shown to be required for the formation of a catalytically active spliceosome [53-55]. Hence, an important focus of future will be to identify PTMs in spliceosomal proteins, and determine whether their modifications regulated during the splicing process.

Protein phosphorylation is one of the most common post-translational modifications and plays an important role in the regulation of a variety of cellular events, including signaling, cell differentiation, metabolism and apoptosis [56, 57]. Hence, the characterization of phosphorylation is a key issue in current proteomic research. Recently, MS-based techniques have been widely applied as powerful tools to characterize protein modifications, including phosphorylation, due to its speed, reliability, high sensitivity and capability for determining phosphorylation sites by MS/MS sequencing. However, large-scale phosphoproteomic analysis still remains a substantial challenge due to low abundance of phosphopeptides combined with large amounts of non-phosphorylated peptides which tend to suppress the ion signal of phosphorylated peptides in MS analysis [58-61]. Therefore, the highly specific separation and enrichment of phosphopeptides from proteolytic digest mixture becomes a critical step prior to MS analysis.

An effective method to resolving these problems is selective enrichment of phosphorylated peptides before MS analysis, like strong cation exchange (SCX) chromatography [62-66], strong anion exchange (SAX) chromatography [62, 64], immobilized metal ion affinity chromatography (IMAC) [67-74] and metal oxide affinity chromatography (MOAC) [75-88]. Among them, IMAC is commonly used and successfully coupled to various mass spectrometries. With this approach the phosphopeptides are captured by chelating interaction with metal ion such as Fe^{3+} or Ga^{3+} shown in Figure 2.2.

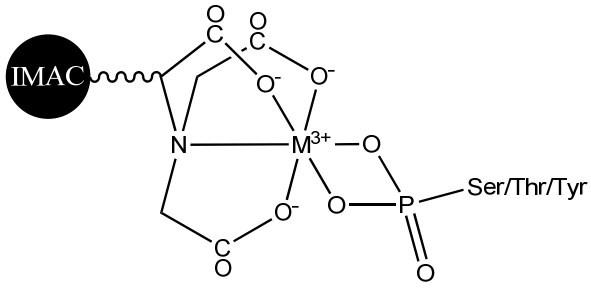


Figure 2.2. Phosphopeptide enrichment with immobilized metal ion affinity chromatography (IMAC). The binding between a phosphopeptide and IMAC resin is shown. The IMAC stationary phase is made by chelation of iminodiacetic acid (IDA) or nitrilotriacetic acid (NTA) with triply charged iron (Fe^{3+}) or gallium (Ga^{3+}) ions.

Enrichment and recovery strongly depend on the type of metal ion, column material, and loading/eluting procedures that are used. However, non-specific enrichment is unavoidable. Acidic peptides tend to bind to the metal resins which complicate the MS detection of enriched phosphopeptides. Such non-specific binding can be reduced by prior methyl esterification of the acidic side chains of amino acid residues [89-92]. Nevertheless, incompleteness and side reactions of methyl esterification process can increase complexity of MS analysis and data interpretation, leading to decrease sensitivity.

Alternative to IMAC, more recently metal oxide affinity chromatography (MOAC) has been demonstrated to be effective material for the selective enrichment of phosphopeptides from proteolytic digests based on chelating interaction between phosphate functional groups and the surface of metal oxide particles. Metal oxides such as TiO_2, ZrO_2, Fe_3O_4 and Al_2O_3 are due to their high chemical stability) [75-88]., tolerance over a broader pH range and physical robustness; hence, some non-volatile acidic additives, including 2,5-dihydroxybenzoic acid (DHB), phthalic acid (PA) and acidic buffers, can be employed to avoid non-specific binding [75, 88]. Among them, TiO_2 has been widely used to selectively capture phosphopeptides shown in Figure 2.3.

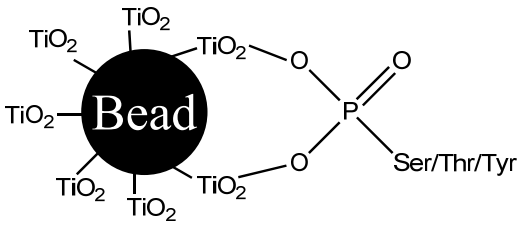


Figure 2.3. Phosphopeptide enrichment with titanium dioxide (TiO_2) bead. The binding between a phosphopeptide and TiO_2 coated resin is shown

Larsen et al. demonstrated that an addition of high quantity DHB in the loading and washing buffer could reduce non-specific bindings on TiO_2 column [75]. However, although a diversity of techniques is available for phosphopeptides enrichment, the mapping of entire phosphoproteome is still a challenging task.

In this study, we invented a simple-to-fabricate, easy-to-use, economic and high efficiency TiO_2 microspin column to enrich phosphopeptides. This design reduces the entire analyzed time in large scare analysis. Three commercially available articles for daily use - coffee filter, pipette tip and eppendorf tube - are utilized to fabricate microspin column. In addition, we integrated our microspin column with different proteomic technique to explore the phosphorylation sites in a crude mixture of nuclear snRNP particles, individual U snRNP, spliceosomal complexes and SR proteins.

2.3 Experiment Sections

2.3.1 Materials

Iodoacetamide (IAA), α-cyano-4-hydroxy-cinnamic acid (CHCA), 2,5-dihydroxybenzoic acid (DHB), phthalic acid (PA), trifluoroacetic acid (TFA), [Glu]-Fibrinopeptide B (Glu-Fib) and ammonium bicarbonate were obtained from Sigma-Aldrich (St. Louis, MO). Sequencing grade, modified trypsin was obtained from Promega (Madison, WI). Dithiothreitol (DTT), formic acid, ammonia solution, acetonitrile (ACN), and ethanol were obtained from Merck (Darmstadt, Germany). RapiGest™ SF was obtained from Waters Corporation (Manchester, UK). Titanium dioxide (TiO_2) resins were obtained from GL Sciences Inc. (Tokyo, Japan).

2.3.2 A Crude Mixture of Nuclear snRNP Particles, Individual U snRNP, Spliceosomal Complexes and SR Proteins Purification

U1, U2, U5 snRNPs and U4/U6.U5 tri-snRNPs were isolated from HeLa nuclear extract by anti-m3G cap-directed immunoaffinity purification with the m3G-specific antibody H-20 to obtain a crude mixture of nuclear snRNP particles [93] and followed glycerol gradient centrifugation to separate the individual snRNPs as described previously [94, 95]. Spliceosomal A, B and C complexes were isolated from in vitro splicing reactions by the MS2 affinity-selection method [96, 97]. SR proteins were isolated by two salt precipitation steps, ammonium sulfate and magnesium chloride precipitation [98]. All protein complexes were purified by Prof. Reinhard Lührmann‘s Laboratory.

2.3.3 Ethanol Precipitation

The purified protein complexes were precipitated by adding 3 volumes of ethanol and 1/10 volume of 3 M sodium acetate, pH 5.3. The mixture was vortexed, incubated at -20 ℃ for 2 hours and then centrifuged 17000 g at 4 ℃ for 30 min. The supernatant was removed and the pellet was washed with 500 µl of 80 % ethanol and centrifuged as above. Discarded supernatant, the pellet was evaporated with a SpeedVac.

2.3.4 In-Solution Digestion

Precipitated proteins were dissolved with 20 µl of 1 % RapiGest™ SF in 25 mM ammonium bicarbonate, pH 8.5, sonicated for 15 min, reduced with 10 µl of 50 mM DTT at 37 ℃ for 1 hour, alkylated with 10 µl of 100mM IAA at 37 ℃ for 1 hour, diluted with 60 µl of 25 mM ammonium bicarbonate, and subsequently digested with trypsin (1:20 enzyme to substrate ratio) at 37 ℃, overnight. Tryptic peptides were acidified with 50 µl of 5 % TFA at 37 ℃ for 2 hours and followed centrifugation at 17000 g for 10 min. The resulting supernatant was transferred to another eppendorf and dried on a SpeedVac for further analysis. Another

procedure employed urea as a denaturation reagent instead of RapiGest™ SF. The ethanol precipitated proteins were dissolved with 20 µl of 6 M urea, sonicated, reduced, alkylated and digested as above and subsequently evaporated with a SpeedVac.

2.3.5 In-Gel Digestion

Precipitated proteins were separated by electrophoresis on a 4-12 % NuPAGE® Novex Bis-Tris Gels system (Invitrogen Corporation, Carlsbad, CA). The gel was stained with Coomassie Blue and cut equally into twenty slices. Each slice was reduced with 60 µl of 50 mM DTT at 37 ℃ for 1 hour, alkylated with 60 µl of 100 mM IAA at 37 ℃ for 1 hour, and subsequently digested with trypsin at 37 ℃, overnight (the enzyme to substrate ratio is 1:20) as described previously [99].

2.3.6 In-House TiO_2 Microspin Column Fabrication

A small plug of coffee filter (1 mm x 1.5mm) was placed at the end of the tip by using a capillary tube shown in Figure. 2.4. The coffee filter serves only as a frit to retain the TiO_2 resins within the commercial pipette tip. A length of approximately 3 mm of TiO_2 materials was packed in the end of a pipette tip. Subsequently, the pipette tip was place into an open hole eppendorf for phosphopeptide enrichment.

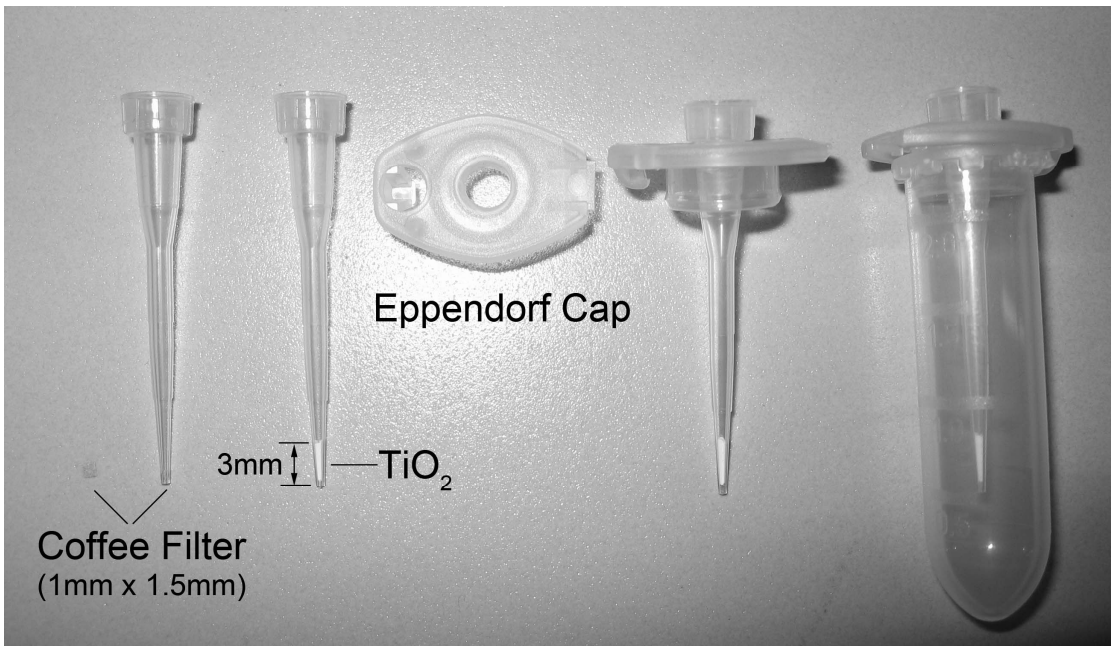


Figure 2.4. Picture of a self-made TiO_2 microspin column.

2.3.7 Comparison of Different TiO_2 Enrichment Procedure

Six different procedures for phosphopeptide enrichment were investigated shown in Figure 2.5. In procedure A, tryptic peptides were dissolved with 60 µl of 200 mg 2,5-dihydroxybenzoic acid (DHB) in 1 ml of 80 % ACN, 5 % TFA and then loaded onto TiO_2 microspin column. The column was washed three times with 60 µl of 200 mg DHB in 1 ml of 80 % ACN, 5 % TFA and five times with 60 µl of 80 % ACN, 5 % TFA. Bound peptides were eluted three times with 40 µl of 0.3 N NH_4OH (pH ≥ 10.5), and subsequently evaporated NH_4OH with a SpeedVac for further analysis. The speed of centrifugation for each step had to be controlled at 3000 rpm for 5 min. Procedure B and C were similar to the procedure A with the exception that an additional desalted step with C18 (Nucleosil 100-5 C18) or NH_2 (Nucleosil 100-5 NH_2) microspin column was employed before TiO_2 enrichment, respectively. Procedure D, E and F were similar to procedure A, B and C, respectively, with the exception that the loading and washing buffer were replaced with saturated phthalic acid (PA) in 80 % ACN, 5 %TFA.

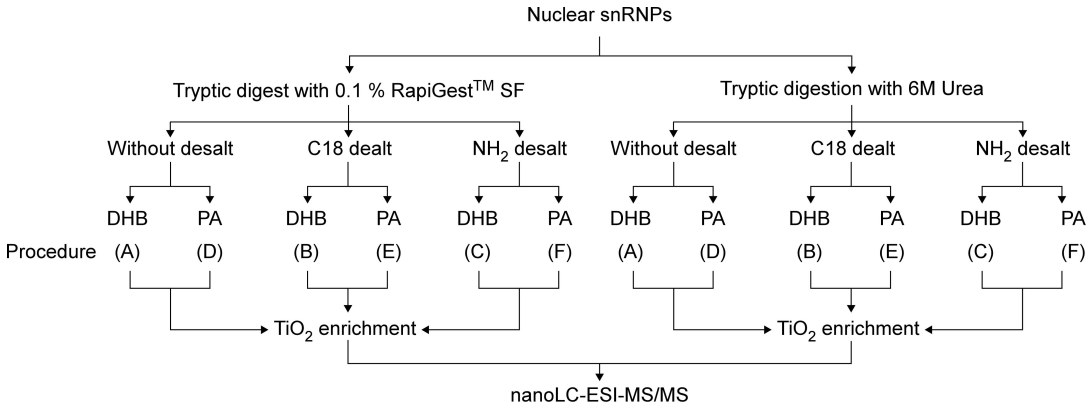


Figure 2.5. Optimization of the purification procedure using microspin TiO_2 column.

2.3.8 NanoLC-ESI and -MALDI Mass Spectrometry Analysis.

The resulting peptides were first loaded at a flow rate of 10 µl/min onto an in-house packed C18 trap column (1.5 cm, 360 µm o.d., 150 µm i.d., Nucleosil 100-5 C18, Macherey-Nagel, GmbH & Co. KG). The retained peptides were then eluted and separated on an analytical C18 capillary column (30 cm, 360 µm o.d., 75 µm i.d., Nucleosil 100-5 C18) at a flow rate of 300 nL/min, with a gradient from 7.5 to 37.5 % ACN in 0.1 % formic acid for 60 min, 120 min or 240 min, using an Agilent 1100 nano-flow LC system (Agilent Technologies, Palo Alto, CA), coupling with Z-spray source or a robotic Probot ™ micro fraction collector (LCPackings/Dionex, Sunnyvale, CA) for ESI-MS and -MS/MS (Waters/Micromass Q-Tof Ultima ™ API mass spectrometer, Milford, MA) or MALDI-MS and -MS/MS (4800 MALDI-ToF/ToF, Applied Biosystems, Framingham, MA) analysis, respectively. For ESI-MS and -MS/MS data dependent acquisition, 1 s survey scans were run over the mass range m/z 350 to 1600. A maximum of three concurrent MS/MS acquisitions were triggered for 2+, 3+,

and 4+ charged precursors detected at intensity above 15 counts, after 3s of acquisition, the system switched back to survey scan mode. For LC-MALDI analysis, The eluent was channelled to a nano-Tee where it was premixed with a matrix solution (10 mg/mL α-cyano-4-hydroxy-cinnamic acid, CHCA and 2.5 fmol/µl Glu-Fib in 80 % ACN, 0.1 % FA) delivered at 1.7 µl/min by a syringe pump and directly spotted onto 384-well MALDI sample plate in every 30 s. Subsequently, MALDI-MS and -MS/MS detected and sequenced peptides. 1000 and 5000 shots were accumulated in positive ion mode MS and MS/MS, respectively. For collision-induced dissociation (CID) MS/MS operation, argon was used as collision gas and the indicated cell pressure was set up 5×10^{-7}, with the potential difference between the source acceleration voltage and the collision cell set at 1 kV. The resolution of time ion selector for precursor ion was set at 200. MS spectra were acquired using Glu-Fib for internal calibration and MS/MS spectra were acquired using instrument default calibration.

2.3.9 Interpretation of Tandem Mass Spectra

All spectra were searched MASCOT server v2.2.0.6 against the NCBInr database limited to human with criteria-peptide mass tolerance, 50 ppm; MS/MS ion mass tolerance, 0.25 Da; allowed up to three missed cleavage; variable modifications were considered phosphorylation of serine, threonine and tyrosine, methionine oxidation and cysteine carboxyamidomethylation. The significant protein hits defined as protein score must be higher than 50 and if the protein score was between 20 and 50, we manually evaluated each MS/MS spectra. The threshold for individual peptide score must be higher than 20 and required bold red. All phosphorylated sited were examined manually by the presence of a 69 Da between fragment ions for phosphoserine and an 83 Da for phosphothreonine. Motif analysis was performed using motif-x [100] (http://motif-x.med.harvard.edu/) with significance set to 0.000001 and occurrence set to 20, using the human IPI as background. All amino acid frequency plots weblogos [101] (http://weblogo.berkeley.edu/) were created as frequency plots.

2.4 Result and Discussion

2.4.1 TiO_2 Microspin Column

Recent reports have shown that TiO_2 can be widely used to enrich phosphopeptides from peptide mixtures; however, the entire enriched procedure requires approximately 15 min for each sample [75, 102, 103]. Hence, we set out to establish a robust and fast (i.e. semi-high-throughput) enrichment procedure for phosphopeptides derived from in-gel digestion and in-solution digestion. TiO_2 materials are packed into a pipette tip that contains ordinary (coffee) filters as a frit to prevent leakage of TiO_2 resins during the enrichment procedure (Figure 2.4). Hundreds of TiO_2 microspin column were fabricated and tested during the period of this study, but we did not find any TiO_2 resins slip out of the column. The system

has the further advantage that the device is small enough to fit into a 1.5 or 2 ml eppendorf tube. It can thus be easily placed into a benchtop laboratory centrifuge with a rotor for 1.5 (2 ml) eppendorf tubes. In this manner 24 samples (using a standard rotor) can be processed in parallel, the cost is reasonable and one-time usage eliminates the risk of carryover. In contrast, phosphopeptide enrichment with TiO_2 microcolumns packed with gel loader tips, which are widely used in many labs, is much more time consuming due to manual handling, i.e. loading, washing and eluting procedure.

2.4.2 Sensitivity of the TiO_2 Microspin Column

In initial experiments the specificity and sensitivity of our semi-automatic device for the enrichment of phosphopeptides was tested and compared with that of TiO_2 microcolumn packed in gel loader tip [75]. We examined the power of TiO_2 enrichment for phosphopeptides that derived from the arginine/serine-rich splicing factor SFRS1, SC35 and 9G8 separated by SDS-PAGE [98] as showed in Figure 2.6.

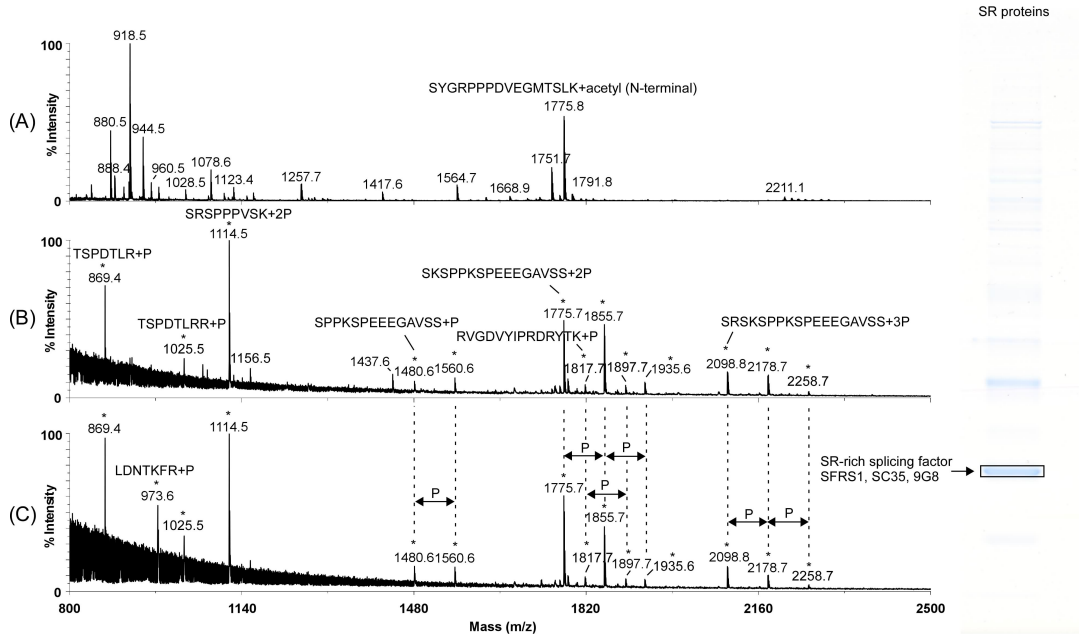


Figure 2.6. MALDI mass spectra of tryptic peptides of arginine/serine-rich splicing factor SFRS1, SC35, 9G8. (A) Before enrichment. (B) Enrichment with the existing manual TiO_2 microcolumn. (C) Enrichment with our TiO_2 microspin column. The inset shows the SDS-PAGE gel image from purified SR proteins. The signals of phosphorylated peptides are marked with asterisk *.

Fig 2.6A showed a direct MALDI peptide mass fingerprint of an aliquot of extracted tryptic peptides derived from arginine/serine-rich splicing factor SFRS1, SC35 and 9G8 without enrichment. The resulting spectrum was dominated by the signals of non-phosphorylated

peptides. After enrichment with TiO_2 microcolumn packed in gel loader tip (Figure 2.6B) or with our TiO_2 microspin column (Figure 2.6C, see Experiment Section 2.3.7, Procedure A in detail), the resulting spectra indicated the majority of non-phosphorylated peptides were removed, and the signals corresponding to the phosphopeptides were enhanced (marked with a star), which were confirmed by ESI-MS/MS (the peptide sequence with the phosphorylation site is listed within the spectrum, see also Appendix 1). In total, fourteen phosphopeptide signals were observed in both enrichment methods. One additional phosphopeptide signal at m/z 973.6 was found during enrichment with our TiO_2 microspin column identified by our microspin column.

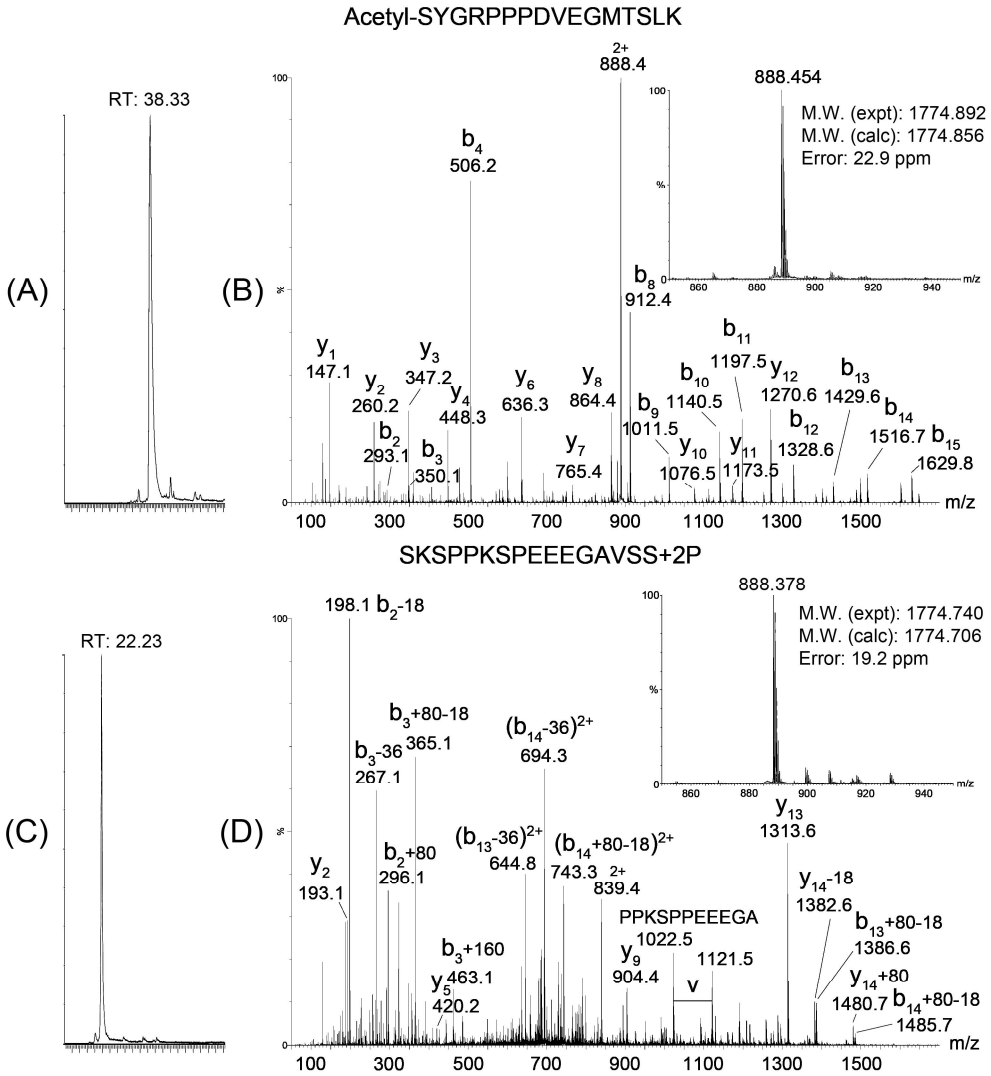


Figure 2.7. (A-B) Extracted ion chromatograms of peptide acetyl-SYGRPPPDVEGMTSLK at m/z 888.454 and phosphopeptide SKSPPKSPEEEGAVSS at m/z 888.378, respectively. (C-D) MS/MS spectra of peptide acetyl-SYGRPPPDVEGMTSLK and phosphopeptide SKSPPKSPEEEGAVSS, respectively. The inset showed the m/z of precursor ion in the mass spectrum.

Strikingly, the abundant peak in the non-enriched spectrum at m/z 1775.8 (Figure 2.6A) was shown by MS/MS spectrum to be the N-terminal acetylated peptide SYGRPPPDVEGMTSLK of 9G8 protein (Figure 2.7B), whereas in both the enriched samples at m/z 1775.7 (Figure 2.6B and 2.6C), it was found to be the phosphopeptide SKSPPKSPEEEGAVSS (Figure 2.7D). The mass difference between these two peaks was 0.152 Da. They were eluted separately from LC column at a retention time of 38.33 min and 22.23 min, respectively (Figure 2.7A and 2.7C). The specificity and sensitivity of our TiO_2 microspin column is equivalent to existing manual microcolumn. Importantly, our TiO_2 microspin column allows a significant higher sample throughput within a given time. In this manner phosphopeptides derived from proteins from in an entire gel lane that was cut into 23-24 slices (e.g. NUPAGE gels) can be enriched in parallel (see also below).

2.4.3 Optimizing Phosphopeptide Enrichment with TiO_2 microspin column for In-Solution Digestion

We next made use of the high-throughput capabilities of the TiO_2 microspin columns and systematically tested for maximum selectivity and sensitivity upon various in-solution digestion procedure, desalting steps and washing conditions that were most favorable to identify the maximum number of phosphopeptides. A crude mixture of total U snRNPs isolated from Hela nuclear extraction by immunoaffinity chromatography [93] was used as a test system.

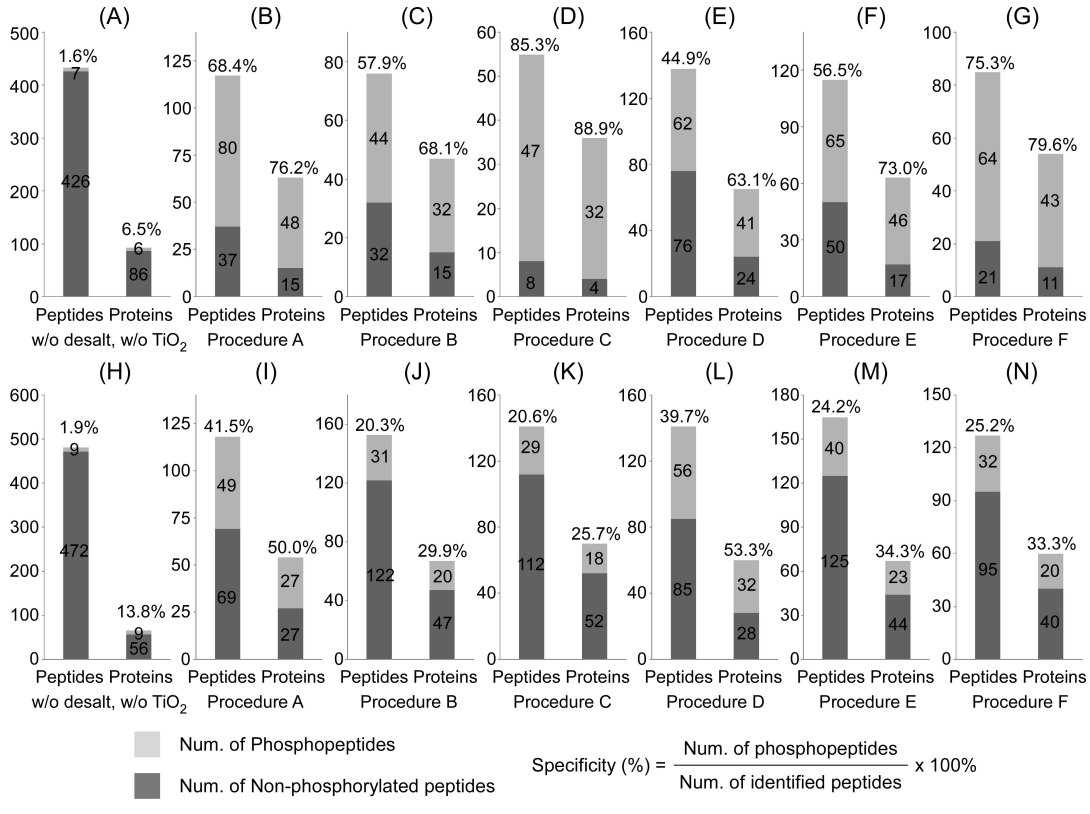


Figure 2.8. Selectivity and sensitivity of the different phosphopeptides enrichment strategies using TiO_2 microspin column. Various surfactants , desalting steps, loading and washing buffers were compared to identify the maximum number of phosphopeptides and phosphoproteins by nanoLC-ESI-MS/MS as shown in A-N, respectively. (A-G) 10 µg total U snRNPs were digested in the presence of RapiGest™ SF. (H-N) instead of RapiGest™ SF with urea. (A) and (H) without enrichment. (B) and (I), procedure A, tryptic peptides were dissolved with 2,5-dihydroxybenzoic acid (DHB) buffer and subsequently loaded onto TiO_2 microspin column; washed with DHB buffer and 5 % TFA in 80 % ACN and then eluted with NH_4OH. (C) and (J), procedure B, an additional C18 desalting step was performed prior to TiO_2 enrichment. (D) and (K), procedure C, instead of C18 with NH_2 material. (E) and (L), procedure D, (F) and (M), procedure E, (G) and (M), procedure F, were the same as procedure A, B and C, respectively, with the exception that loading and washing buffer were replaced with saturated phthalic acid (PA). (see Experiment Section 2.3.7 and Figure 2.5 in detail). Specificity: Num. of phosphopeptides/Num. of identified peptides x 100 %.

Initial optimization of the procedure was evaluated by testing first different surfactant on the impact of TiO_2 enrichment procedure. We first compared urea and RapiGest™ SF as denaturing agent (Figure 2.8.A-G and 2.8H-N, respectively). 10 µg of a crude mixture of total U snRNPs was denatured, reduced, alkylated and digested with trypsin over night. The resulting peptides were enriched as above (see Experiment Section 2.3.7 and Figure 2.5 in detail) and eluted peptides were analyzed by LC-ESI-MS/MS. Without enrichment only seven and nine phosphopeptides were identified in RapiGest™ SF and urea denatured sample, respectively (Figure 2.8A and 2.8H). In general, the total number of peptides derived from urea treated samples was higher as compared to RapiGest™ SF as the total number of peptides included those that are carbamylated at N-terminus, lysine, arginine and cysteine residues as shown in Figure 2.9 [104-106].

We found that digestion with RapiGest™ SF significantly increased the number of sequenced phosphopeptides (Figure 2.8B-G compared to 2.8I-N) and reduced the number of less non-phosphorylated peptides as compared to urea treatment. This can be explained several reasons: First, urea decomposes into ammonium cation and ammonium cyanate during the digestion process (Figure 2.9A). The ammonium cation elutes phosphopeptides from TiO_2 resins and therefore decreases the number of phosphopeptides that can be identified. Secondly, the lone pair electrons of nitrogen on the carbamyl group of carbamaylated peptides also binds to TiO_2 resins. Third, RapiGest™ SF is an acid-labile surfactant. It hydrolyzes in acid solution posterior to in-solution enzymatic digestion to form sulfonic sodium salt in the sample shown in Figure 2.10 [107]. Carboxyl acid peptides in the sample that also bound to

TiO_2 beads are displaced by the sulfonic anion but phosphopeptides were still retained, therefore, the specific for phosphopeptide in TiO_2 enrichment is increased when using RapiGest™ SF [84, 102].

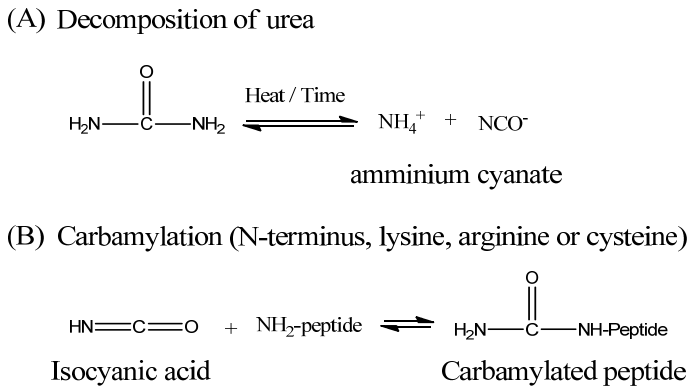


Figure 2.9. Formation of carbamylated proteins derived from urea. (A) A simple decomposition reaction of urea to form ammonia and cyanic acid (or isocyanic acid). (B) Carbamylation of proteins and peptides as isocyanic acid reacts with the N-terminus, lysine, arginine or cysteine residue.

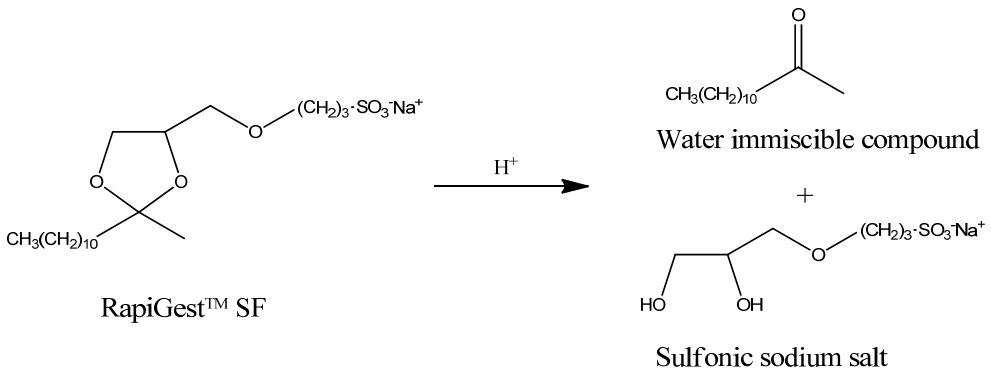


Figure 2.10. Decomposition of RapiGest™ SF in acid solution to form water immiscible compound and sulfonic sodium salt.

Since the surfactant deeply influenced the selectivity and sensitivity of TiO_2 beads, we tested C18 and NH_2 materials for removal of the surfactant prior to TiO_2 to get more specificity and efficiency of enrichment. C18-based extraction method had been widely used in proteomics and other analyses for concentrating and desalting peptides [108, 109]. Only 44 and 31 phosphopeptides were identified by LC-MS/MS after desalting with C18 material and subsequent enrichment with TiO_2 using RapiGest™ SF and urea, respectively (Figure 2.8C and 2.8J). Thus, the specificity of phosphopeptide enrichment was decreased from 68.4 % to 57.9 % in RapiGest™ SF experiment, and from 41.5 % to 20.5 % in the urea experiment. The missed phosphopeptides may be too hydrophilic or stuck in eppendorf and pipette tip during

the desalting step. The removal of sulfonic sodium by C18 material caused the decrease in specificity of TiO_2 enrichment.

Alternatively, we tested for removal of the surfactant with NH_2 material based on hydrophilic interaction and anion exchange [110]. The NH_2 material in acid solution keeps its positive charge and can capture acid peptides including phosphopeptides. Consequently, the specificity of detected phosphopeptides was increased from 76 % to 86 % in this particular experiment (compared Figure 2.8B with 2.8D) and also in most of the following experiments from 44.9 % to 75.3 % (Figure 2.8E and 2.8G) when NH_2 material was used for desalting RapiGest™ SF. Although NH_2 material increased the selectivity of enrichment, an additional desalting step decreased the sensitivity. In urea digested sample, we found that the number of detectable phosphopeptide was decreased when NH_2 material was used for desalting (compare Figure 2.8I to 2.8K). We observed highly carbamylated peptides also bound to NH_2 material due to the formation of hydrogen bond between two amine groups; hence, the combination of urea and NH_2 material caused a serious problems for the specificity of phosphopeptide detection also in the following experiments (Figure 2.8K and 2.8N). In both the experiments we observed a loss of phosphopeptides due to the additional desalting step. Removal of the surfactant with C18 or NH_2 materials prior to TiO_2 enrichment showed no improvement for detection of phosphopeptides. In brief, simplified the entire procedure was required to minimize loss of phosphopeptides.

Furthermore, we also evaluated different loading and washing conditions for enrichment of phosphopeptides. Recent studies demonstrated that both the two variant aromatic carboxylic acids, 2,5-dihydroxybenzoic acid (DHB) [75] and phthalic acid (PA) [88], help to decrease the number of non-phosphorylated peptides bound to TiO_2 and thus increases the number of detectable phosphopeptides in the sample. Indeed, loading digested samples on TiO_2 beads solely in 80 % ACN (v/v), 5 % TFA (v/v) without DHB and PA as additives demonstrates that too many non-phosphorylated peptides bind to TiO_2 that were subsequently eluted in the final enrichment step although no phosphopeptides are detected in the flow through or in the wash [111]. We tested whether PA could increase the number of detectable phosphopeptides in our samples when it was used in the loading buffer instead of DHB (Figure 2.8E-G and 2.8L-N). The results show that irrespective from the applied desalting strategy PA does not significantly increase the number of detectable phosphopeptides (compare Figure 2.8B-D to 2.8E-G and 2.8I-K to 2.8L-N).

In summary, the experiments with the TiO_2 microspin column show that the use of RapiGest™ SF is advantageous for denaturing sample prior to digestion and subsequent TiO_2 enrichment of phosphopeptides as compared to urea. Furthermore, an additional desalting step prior to TiO_2 enrichment does not significantly improve the detection of phosphopeptides. Although the

use of NH_2 for desalting enhances the purification efficiency of phosphopeptides (percentage phosphopeptides to non-phosphorylated peptides) the overall number of phosphopeptides and thus phosphorylated proteins is lower. For the following experiments, all samples were digested in the presence of RapiGest™ SF without any desalting step prior to phosphopeptide enrichment with TiO_2 microspin column to avoid any loss of phosphopeptides (Experiment Section 2.3.7, Procedure A).

2.4.4 Global Profiling of Phosphopeptides from A Crude Mixture of Nuclear snRNP Particles, Individual U snRNP, Spliceosomal Complexes and SR Proteins

The procedure was first applied on a crude mixture of small nuclear ribonucleoprotein (total U snRNPs) immunopurified from nuclear extraction in order to define a robust generally applicable protocol towards the detection of phosphopeptides derived from proteins involved in eukaryotic mRNA splicing. The snRNPs are particles that combine with pre-mRNA and various proteins to form the key components of the spliceosome, which catalyses the excision of introns and ligation of exons of eukaryotic pre-mRNAs

50 µg total U snRNPs were either separated by SDS-PAGE, cut into twenty gel slices and subsequently proteolyzed with trypsin or performed in solution enzymatic digestion with RapiGest™ SF as mentioned above. The resulting peptides were enriched using the TiO_2 microspin columns without any desalting step. Enriched samples were subjected to LC-online coupled ESI-MS/MS or to LC-offline MALDI-MS/MS. The LC-MS/MS analyses were performed in triplicate. The overall workflow is outlined in Figure 2.11.

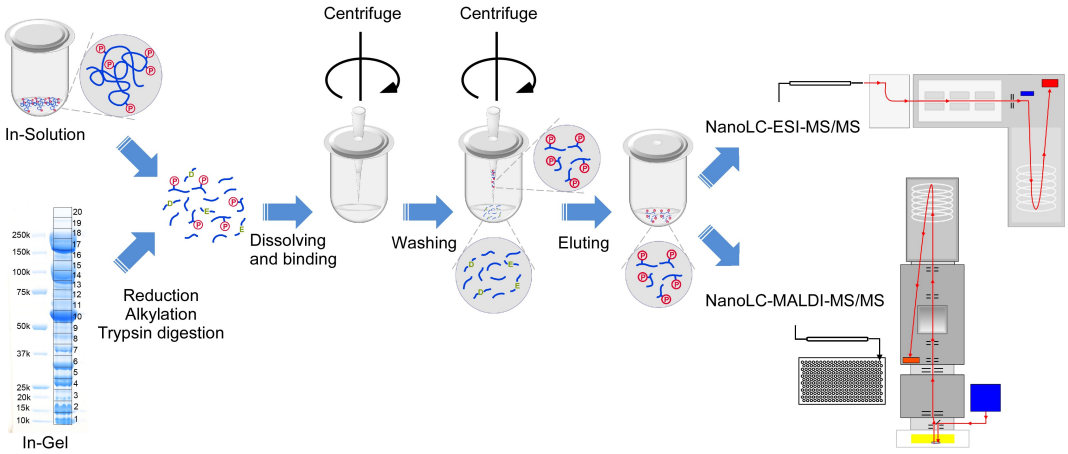


Figure 2.11. Strategy used for the large-scale identification and characterization of phosphorylated sites from total U snRNPs. TiO_2 microspin columns were integrated with nanoLC ESI-MS/MS and off-line nanoLC MALDI-MS/MS to identify phosphopeptides derived from in-gel and in-solution proteolytic digestion.

For the samples derived from in-gel digestion, 60 min LC gradient for separation of (phospho)peptides was applied. In-gel digested samples were not subjected to LC-offline MALDI MS/MS as the overall analysis time of samples (LC off-line and MS and MS/MS on the selected precursor) would exceed a certain threshold that is considered to be reasonable for a routine analysis of phosphopeptides. In-solution digested samples were analyzed by 120 min and 240 min LC gradients (see Experiment Section 2.3.8 in detail). Figure 2.12 summarizes the results.

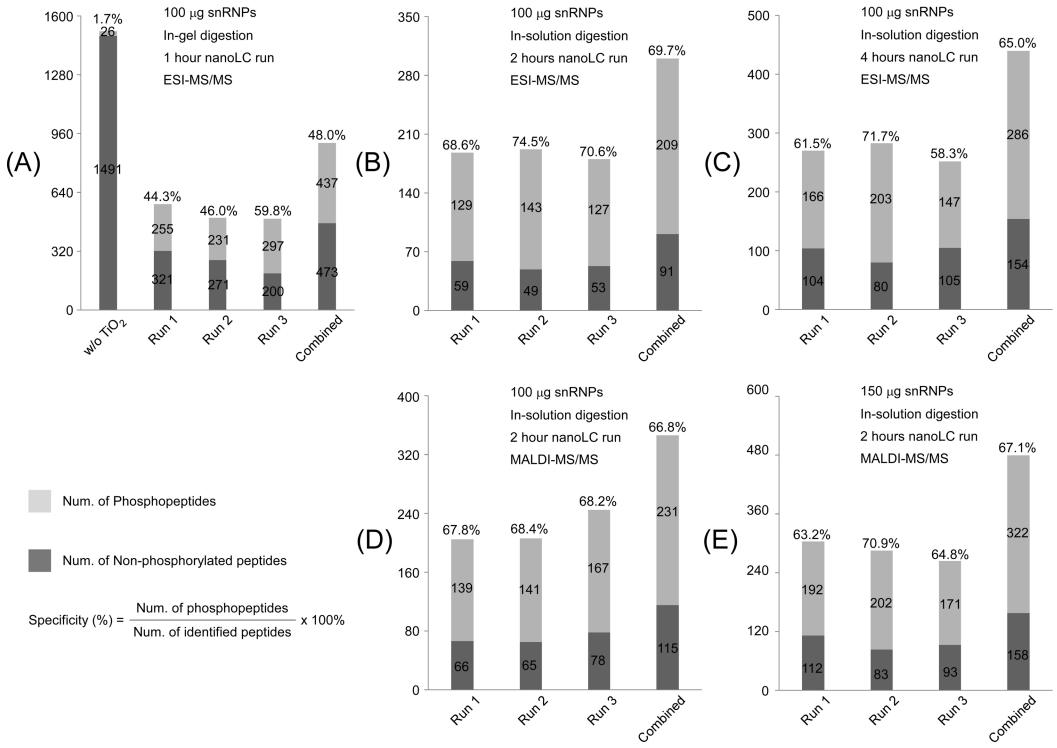


Figure 2.12. The Number of non-phosphopeptides and phosphopeptides identified from tryptic peptides of total U snRNPs by different proteomic approach. (A-D) 100 µg total U snRNPs. (A) In-gel digestion, 1 hour nanoLC gradient, ESI-MS/MS analysis. (B) In-solution digestion, 2 hours nanoLC gradient, ESI-MS/MS analysis. (C) In-solution digestion, 4 hours nanoLC gradient, ESI-MS/MS analysis. (D) In-solution digestion, 2 hours noaoLC gradient, MALDI-MS/MS analysis. (E) 150 µg total U snRNPs, in-solution digestion, 2 hours nanoLC gradient, MALDI-MS/MS analysis.

Without enrichment, only 26 phosphopeptides out of 1491 non-phosphorylated peptides were found in in-gel digested sample (Figure 2.12A). Enrichment of the extracted peptides from each gel slice using the TiO_2 microspin columns and analysis by LC-online-ESI-MS/MS with a 1 hour gradient identified 437 phosphopeptides in 164 different proteins. From three independent LC separations, approximately 48 % of the totally enriched and sequenced peptides were identified as being phosphorylated. This value is lower as compared to the initial

studies with the in-solution digest where approx. 68.4 % of the sequenced peptides were phosphopeptides (Figure 2.8B).

In solution digest of the same sample reduced the sample preparation and analysis time, but on the other hand the number of detected and sequenced phosphopeptides was significantly decreased as compared to in-gel digestion. By applying a 2 or 4 hours LC gradient, 209 and 286 phosphopeptides, respectively were detected and sequenced (as compared to 443 in the in-gel digestion). A longer gradient thus helps to identify more phosphopeptides but also more non-phsophorylated so that the purification efficiency is lower. (65 % compared to approx. 69.7 %, compare Figure 2.12B and 2.12C).

We directly applied the enriched sample to off-line nanoLC MALDI-MS/MS and obtained a similar result as compared with ESI-MS/MS (compare Figures 2.12B to 2.12D). Because the selectivity of our enriched procedure was about 70 %, the amount of enriched phosphopeptides was sufficient to apply α-cyano-4-hydroxy-cinnamic acid (CHCA) as MALDI matrix instead of DHB with phosphoric acid to obtain desirable fragmentation pattern for phosphosite mapping. DHB matrix was shown to enhance the enhance the phosphopeptide signal in MALDI-MS [112], it was nonetheless difficult to generate satisfying MS/MS spectra in MALDI [46]. Increasing the sample amount helps to identify additional phosphopeptides (still with approx. 67.1 % enrichment efficiency) as shown for the LC-offline nanoLC MALDI MS/MS (Figure 2.12E)

Even longer nanoLC gradient (4 hours) applied on in-solution digested samples or larger sample amount did not yield numbers of phosphopeptides comparable to the in-gel digestion with subsequent enrichment on TiO_2 microspin columns. On the other hand the selectivity (i.e. the phosphopeptide enrichment efficiency) is higher when samples are digested in-solution using RapiGest™ SF as surfactant. At least in our hands, it seems that in-gel digestion reduces the complexity of samples more and thus more phosphopeptides are identified as longer in nanoLC gradients. The design of the TiO_2 microspin column makes it possible that phosphopeptides from 24 gel-splices can be enriched in a reasonable time span in a semi high-throughput manner. Importantly, the number of phosphopeptides always varies in the three independent LC separations (extracted peptides from in-gel digestion and peptides derived from in-solution digest). Such phenomenon was described in previous studies with larger proteomes [113] but it was not expected to be similar in our studies with a nuclear sub-proteome, i.e. total snRNP particles.

Figure 2.13A summarizes the number of phosphopeptides that were identified by applying our enrichment approach on large scale purified snRNPs. It is obvious that the combination of in-gel digestion, TiO_2 microspin and LC-ESI-MS/MS in triplicate yield the greatest number of phosphopeptides. Figure 2.13B compares the total number of phosphopeptides and

phosphoproteins identified in the three different approaches (in-gel-digestion, LC-online ESI-MS/MS, and LC-offline MALDI-MS/MS). Notably, each digestion, separation and MS method identified different, partially overlapping segments of the phosphopeptides and phosphoproteins, implying that none of the methods by itself is currently able to comprehensively analyze a phosphoproteome. By combining the results, a total of 717 phosphopeptides were identified from crude mixture of immunoaffinity purified total U snRNPs. In 641 phosphopeptides the phosphorylation site could be unambiguously identified thus confirming 510 distinct phosphorylation sites. 115 phosphopeptides in 61 phosphoproteins were found in all three approaches. 228 phosphopeptides from 69 proteins were identified after in-gel digestion only, 120 phosphopeptides from 41 protein were identified by in-solution digestion and subsequent ESI-MS/MS only, and 78 phosphopeptides from 25 proteins were identified by in-solution digestion and subsequent LC-offline MALDI-MS/MS only. The number of non-confirmed phosphopeptides is relative low in either approach (Figure 2.13A) since the MS/MS spectra were manually evaluated. We also searched the MS/MS spectra against a decoy database to determine the false positive rate less than 1 %.

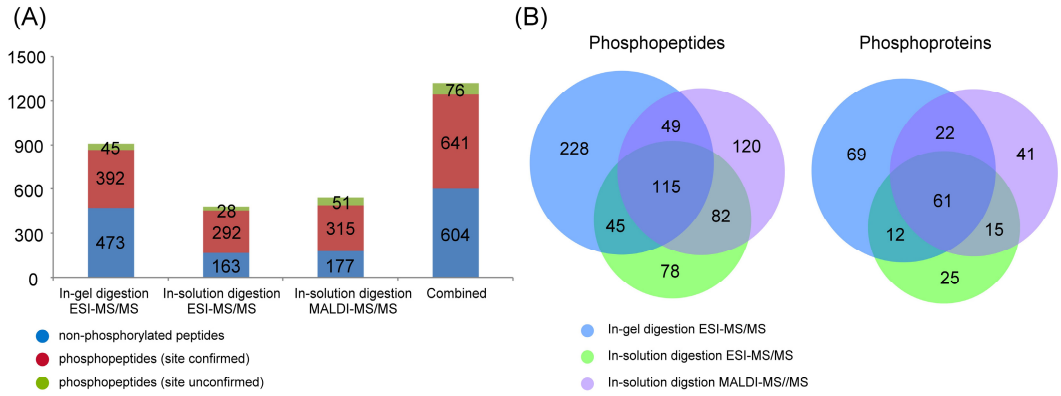


Figure 2.13. (A) The number of non-phosphopeptides and phosphopeptides identified from tryptic peptides of total U snRNPs by different proteomic approach. (B) Overlap of distinct phosphopeptides and phosphoproteins identified with ESI- and MALDI-MS/MS integrated with TiO_2 microspin column enrichment from in-gel and in-solution digested sample.

Taking ESI-MS/MS and MALDI-MS/MS as comparison, the latter seemed laborious to localize the position of phosphorylated site. Approximately 14 % (51/366) of phosphopeptides we could not ascertain the location because the molecular ions corresponding to these phosphopeptides offered high abundance of neutral loss of phosphate moiety (-98 Da) in MALDI-MS/MS (Fig. 2.14B and 2.14D). Compared the MS/MS spectra of phosphopeptide ^{54}CAPSAGpSPAAAVGR^{67}, it was straightforward to confirm the site at S60 relied on y_7, y_8-98, y_9 ions from ESI-MS/MS spectrum. (Figure. 2.14A) However, in MALDI-MS/MS spectrum, y_7 and y_{12} ions indicated two possible sites at S57 or S60 (Figure 2.14B); nevertheless, we still

had large amount of high quality spectra, like phosphopeptide ^{392}FNDpSEGDDTEETEDYR^{407} could distinctly confirm the site at S395 in both systems (Fig. 2.14C and 2.14D). In brief, our TiO_2 enrichment procedure in combination with different proteomic strategy successfully applied to large scale phosphoproteome analysis, resulting in the identification of 245 phosphoproteins corresponding to 510 distinct phosphorylated sites in total U snRNPs.

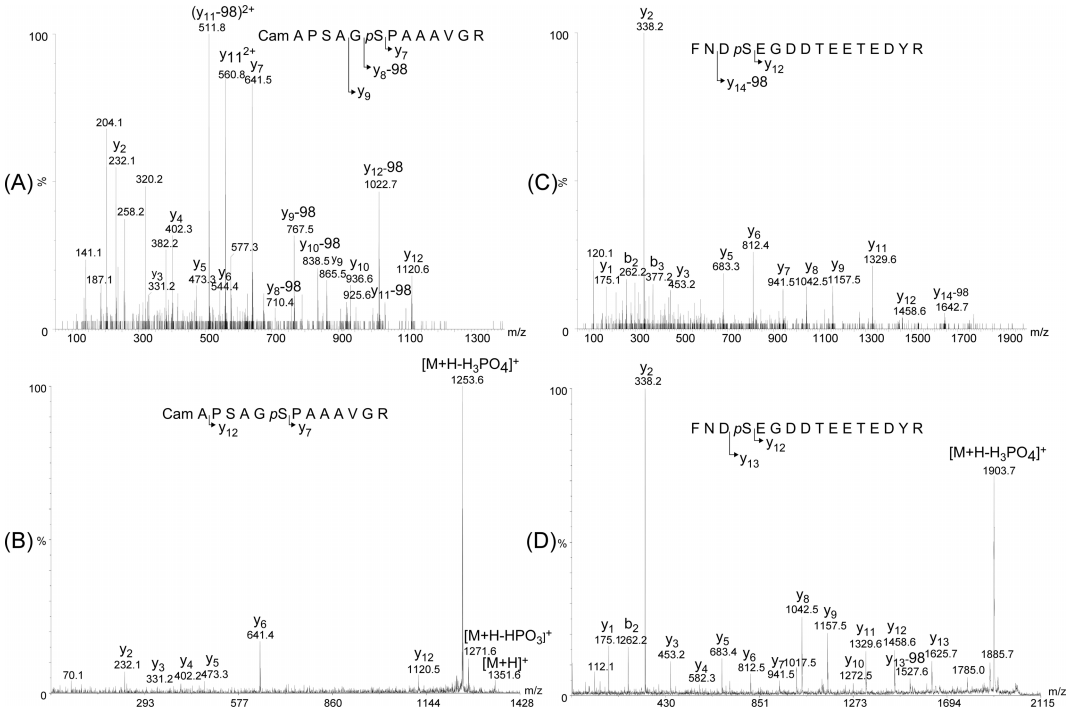


Figure 2.14. MS/MS spectra of phosphopeptides extracted from tryptic peptides of total U snRNPs. (A) and (B) phosphopeptide ^{54}CAPSAGpSPAAAVGR^{67} detected with ESI- and MALDI-MS/MS, respectively. (C) and (D) phosphopeptide ^{392}FNDpSEGDDTEETEDYR^{407} detected with ESI- and MALDI-MS/MS, respectively.

To examine the power of our enrichment for identification of phosphorylation sites, we further applied our strategy on the large scale analysis of individual U snRNPs, SR Proteins and spliceosomal complexes. Direct TiO_2 microspin columns integrated with nanoLC ESI-MS/MS analyses of tryptic peptides from protein complexes in solution as well as gel bands from one-dimensional SDS-PAGE afforded the best combination in efficient construction of phosphorylation map as showed in Figure 2.15. 1381 distinct phosphorylated sites corresponding to 390 phosphoproteins corresponding to were identified in the study (Appendix 1). In the 1381 phosphorylation sites, 84.1 % (1161) and 14.6 % (202) were serine and threonine, respectively, and only 1.3 % (18) were tyrosine. The above values were similar to the results from previous reports [57]. Manually validated phosphorylation sites in the corresponding proteins were searched against Expasy Knowledgebase (http://us.expasy.org/) and PhosphoSite database (http://www.phosphosite.org/) [114]. Both these database have

collected comprehensive information on in vivo protein phosphorylation and their corresponding references. In previous report, 2002 phosphorylation sites corresponding to 967 proteins were identified in large scale nuclear extracted proteins by Beausoleil et al. [65], however, many of the phosphorylated sites identified in this study have not been reported by Beausoleil et al.. Approximately 46.3 % of the sites (640 phosphorylated sites) were considered to be novel.

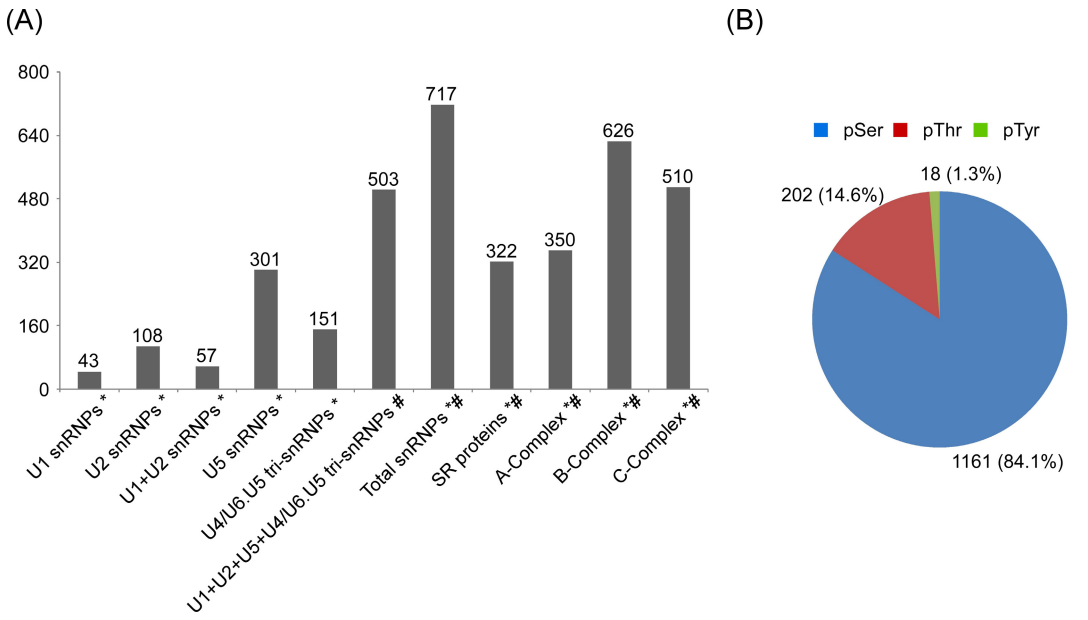


Figure 2.15. (A) Number of phosphopeptides identified from Individual U snRNPs, total snRNPs, SR proteins, spliceosomal complexes. * means in-solution digestion. # means in-gel digestion from SDS-PAGE. (B) The distribution of phosphorylated residues identified from spliceosomal proteins .

2.4.5 Novel and Known Kinase Motifs in Spliceosomal Proteins

After we generally classified all phosphopeptides in our data set into primary sequence categories, we sought to further refine these categories into specific, frequency-corrected phosphorylation motifs. Kinase specificity typically depends on the primary amino acid sequence surrounding the target phosphorylation site [115]. Peptide sequences for phosphorylated sites localized on serine and threonine were all aligned, and their lengths were adjusted to ± 6 amino acid from the central position and submitted to the Motif-X algorithm [100]. Each identified motif, logo-like representations were created to graphically display. These logos included not only the residues strictly discovered to be part of the motif, but also the frequencies of all additional adjacent amino acids as showed in Figure 2.16. The classes included Pro-directed, basophilic and acidiphilic, other motifs generated containing a minimum of 20 pSer or pThr occurrences. The majority of the found pSer- and pThr-containing sites

were Pro-directed which accounted for 12.5 % (145/1161) and 39.1 % (79/202) of all detected phosphorylation, respectively, as showed in Figure 2.16A. Another class are acidic motifs in which acidic residues (Asp or Glu) were abundant on the C-terminus (pSDxE, pSExE, pSxxE, pSxxD, DpSD and pSxExE motifs were shown in Figure 2.16B, respectively). In total 223 sites out of 1161 pSer (19.2 %) were identified with such a consensus motif. These sites are previously described to be a target for casein kinase (like) II [100].

Figure 2.16. Phosphorylation-specific motifs using the Motif-X algorithm. (A-D) Sequence logos of phosphorylation motifs where the phosphorylated residue (Ser or Thr) is centered. Number of occurrences of a particular motif in 1161 unique serine and 202 unique threonine phosphorylation are indicated. (A) Pro-directed motif with a strong favorite for Pro residue next to C-terminal of Ser and Thr residues, these motifs were observed 145 and 79 occurrences, respectively. (B) Representation of acidic motifs of casein kinase II substrate (like) with 57, 51, 41, 31, 23 and 20 occurrences, respectively. (C) Representation of basic motifs of Ser and Thr residues were detected with 354 and 33 occurrences, respectively. ATK referred to as protein kinase B or rac protein kinase. PKA referred to as protein kinase A. (D) Representation of other motifs of Ser residues were detected 246 occurrences. Cdc2/CDK kinase substrate motif pSPxR and pSPxK were observed 109 times.

Interestingly, basic motifs in which basic (Arg or Lys) residues are abundant on the N-terminal site, RSRpSxS, RxRSxpS and RpSRS were found 37, 35 and 31 times, respectively, in SR domains of so-called SR proteins. RxRSxpS had been described to be the consensus motif for ATK as showed in Figure 2.16C [100]. SR proteins facilitate the association of snRNPs with the pre-mRNA during spliceosome assembly, and phosphorylation is essential for SR protein

activity [116, 117] and enhances their interactions with other spliceosomal proteins [118]. A hitherto uncharacterized motif, Rxx*p*SP, was found in 91 sites and appeared to be a combination of both basophilic (RxxpS, PKA/PKC kinase substrate motif [119]) and proline-directed (pSP) motifs. Cdc2/CDK kinase motif pSPxR and pSPxK appeared 78 and 31 sites, respectively, shown in Figure 2.16D [100, 120]. Only eighteen pTyr sites were identified in this study, hence, the data set was significantly less than that for Ser or Thr to find the corresponding motif by alignment.

2.4.6 Application of TiO_2 Microspin Column

We further examine the power of our TiO_2 microspin column. We collaborated with Marc Schneider et al. in Prof. Reinhard Lührmann‘s Laboratory on the project “Human PRP4 kinase is required for stable tri-snRNP association during spliceosomal B complex formation” and published on Nature Structural & Molecular Biology, 2010 [121]. Two previously unidentified phosphoprotein, namely the pre-mRNA processing factors PRP6 and PRP31, were discovered as showed in Figure 2.17. The purified B complex was separated by SDS-PAGE and stained with Coomassie blue. The proteins of interest were excised from gel, digested with trypsin, extracted and dried down with Speedvac. The resulting tryptic peptides were enrichment with our TiO_2 microspin columns and followed by nanoLC ESI-MS/MS analysis. Human PRP6 and PRP31 were phosphorylated at multiple sites after tri-snRNP incorporation in to the spliceosomal B complex. In human PRP6, we identified pSer at positions 263 and 279, and pThr at positions 205, 266 and 275. In human PRP31, we detected pSer at positions 439 and 498, and pThr at positions 440, 448 and 455. At Ser 445, Ser 446, Ser 450 and Ser 451, the exact positions were not assigned due to their close proximity but the MS and MS/MS indicated that two of the four serine residues were phosphorylated, respectively (Appendix 2). Marc Schneider et al. provide the evidence that both proteins are directly phosphorylated by PRP4 kinase, which is required for the assembly of stable, functional B complex. They suggest multiple phosphorylation events on PRP6 and PRP31 assist to stabilize the interaction of the tri-snRNP during the B complex assembly and the splicing can potentially be modulated at multiple regulatory checkpoint.

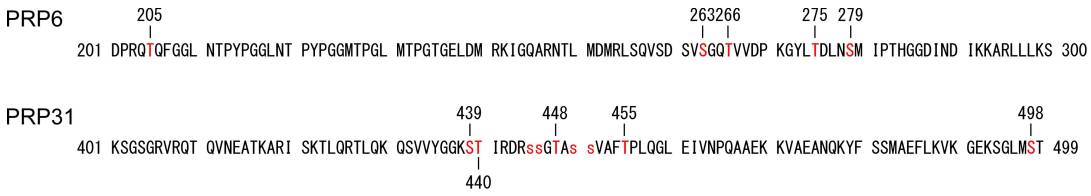


Figure 2.17. Partial amino acid sequence of human PRP6 (aa 201-300) and PRP31 (aa 401-499 and 495-499). Red capital letters indicated phosphorylation sites identified in this study. Red lowercase letters indicated the exact location of phosphorylation sites could not be assigned.

We collaborated with Thomas Oellerich et al. in Prof. Jürgen Wienands's Laboratory on the project "SLP-65 Phosphorylation Dynamics Reveals a Functional Basis for Signal Integration by Receptor-Proximal Adaptor Proteins" and published on Molecular & Cellular Proteomics, 2009 [122]. SLP-65 serves functions as a central linker protein that bridges kinases associated with the B-cell receptor (BCR) with a multitude of signaling pathways, regulating biological outcomes of B-cell function and development. A total of 41 phospho-acceptor sites within SLP-65 were identified under BCR stimulation as showed in Figure 2.18A. The exact location of amino acid were 36 residues; of these, 21 pSer, 6 pThr and 9 pTyr residues were found. For phosphorylation sites located within Ser 397 and Ser 398 and between the amino acids 422 and 429, the exact positions were not assigned due to their close proximity but the MS and MS/MS indicated that one of the two and four of the six serine residues were phosphorylated, respectively. We further integrated SILAC (stable isotope labeling with amino acids in cell culture) approach for profiling phosphorylation dynamics of SLP-65 in resting and activated B cells with BCR stimulation for 2 or 20 min. The most frequently detected phosphopeptides ^{168}KPSLPSPAAKPK^{179} at Ser 170 and Ser 173 were observed (Figure 2.18B-E) The quantity of the singly phosphorylated peptide decreased slightly upon BCR stimulation (Figure 2.18B), whereas the quantity of the doubly phosphorylated peptide increased over time (Figure 2.18C). `The MS/MS spectrum of the singly phosphorylated peptide contained an exclusive phosphorylation at Ser 173 but not at Ser 170 (Figure 2.18D). The MS/MS spectrum of the doubly phosphorylated peptide indicated phosphorylated at Ser 170 and Ser 173. Taking cue together, the amount of the singly phosphorylated peptide at Ser 173 reduced because the formation of phosphorylation at Ser 170 converted the singly phosphorylated peptide into its doubly phosphorylated peptide under BCR stimulation. The result indicated BCR stimulation acted mainly on Ser 170. In brief, 23 distinct phosphopeptides corresponding to 24 phosphorylation sites were quantified. Our data showed that the phosphorylation of SLP-65 acted as early signal initiation and later signal processing during the BCR stimulation.

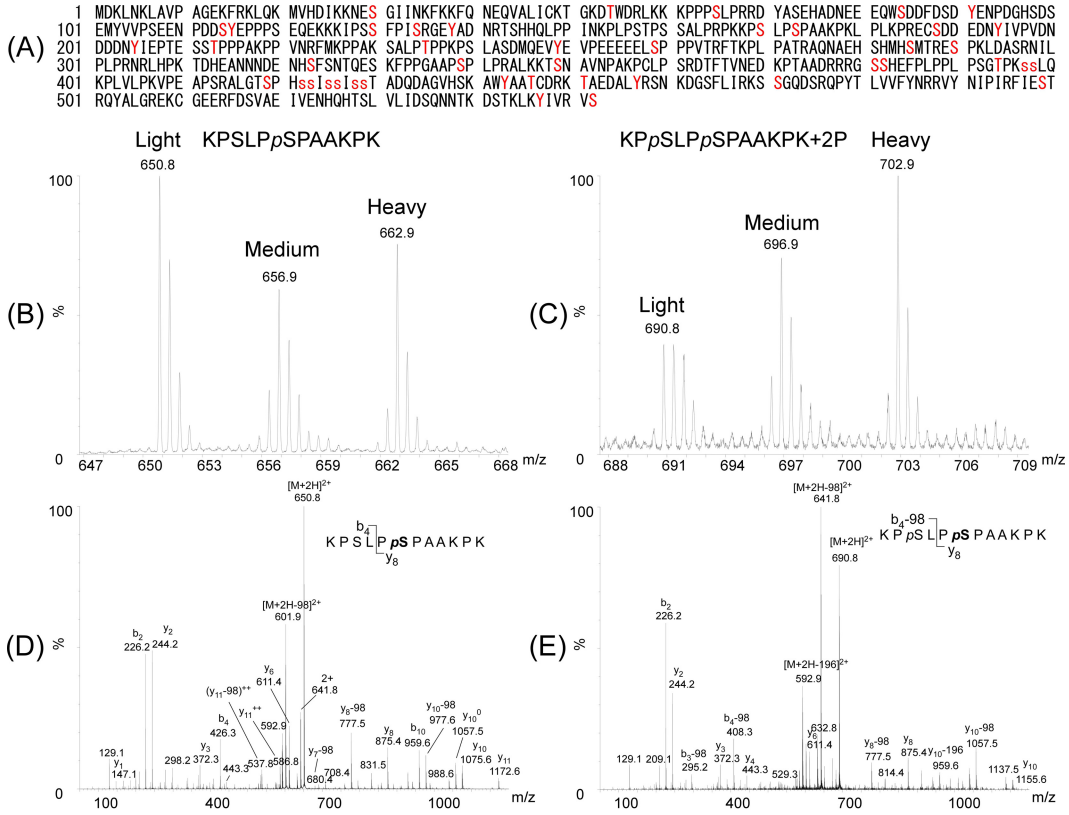


Figure 2.18. (A) The phosphorylation sites in SLP-65 from gallus gallus. Red capital letters indicated phosphorylation sites identified in this study. Red lowercase letters indicated the exact location of phosphorylation sites could not be assigned. (B)-(E) Stable isotope labeling with amino acids in cell culture (SILAC) approach for relative quantification of the most frequently detected phosphopeptide ^{168}KPSLPSPAAKPK^{179} in SLP-65. (B) Mass spectrum of the singly phosphopeptide at Ser 173 in unstimulated cells and in BCR stimulation for 2 or 20 min obtained at m/z 650.8 (light), 656.9 (medium) and 662.9 (heavy), respectively. (C) Mass spectrum of the doubly phosphopeptide at Ser 170 and Ser 173 in unstimulated cells and in BCR stimulation for 2 or 20 min obtained at m/z 690.8 (light), 696.9 (medium) and 702.9 (heavy), respectively. (D) MS/MS spectrum of the singly phosphopeptide ^{168}KPSLPpSPAAKPK^{179} at m/z 650.8. (E) MS/MS spectrum of the doubly phosphopeptide ^{168}KPpSLPpSPAAKPK^{179} at m/z 690.8.

We further collaborated with Xiao Luo et al. in Prof. Markus C. Wahl‘s Laboratory on the project “Structural and Functional Analysis of the E.coli NusB-S10 Transcription Antitermination Complex” and published on Molecular Cell, 2008 [123]. In this study, our TiO_2 microspin column in combination with mass spectrometry was employed to map the hitherto non-characterization of UV-induced protein-RNA contact sites in NusB-S10 complex. The identification of contact sites after UV-induced crosslinks is restricted by the relatively low yield of UV-induced protein-RNA crosslinks. Therefore, the most critical step in such analyses is the effective enrichment of crosslinked species from the excess of non-crosslinked species

prior to MS analysis. A feature shared by protein-RNA crosslinks, free RNA oligonucleotides and phosphopeptides, namely the phosphate groups that these three carry, is used to enrich, as phosphate groups can interact with the TiO_2 material through their free lone pair of electrons. NusB or NusB-S10 protein complex was first incubated with RNA oligomer and exposed to UV light to form NusB-RNA or NusB-S10-RNA complex, respectively. Subsequently, the crosslinked complexes were precipitated with EtOH, hydrolyzed with RNases T1 and A, digested with trypsin. The tryptic peptides were enriched with TiO_2 microspin column for RNA-peptide crosslinks and followed analysis by nanoLC-ESI-MS/MS. Overall, we identified four peptides in NusB and three peptides in S10 that crosslinked to distinct, short RNA elements as showed in Figure 2.19 and Table 2.1. UV-induced crosslinked reaction in the absence of RNA oligomer and MS analysis with complete protein complex but without UV irradiation did not give rise to any peaks corresponding to those of the identified peptide-RNA crosslinks. As previous report, in some cases we saw non-enzymatic breakage or hydrolysis of crosslinked oligonucleotides during enrichment or in the gas phase of the MS [124]. However, we could not identify the actual crosslinked amino acids by mass spectrometry.

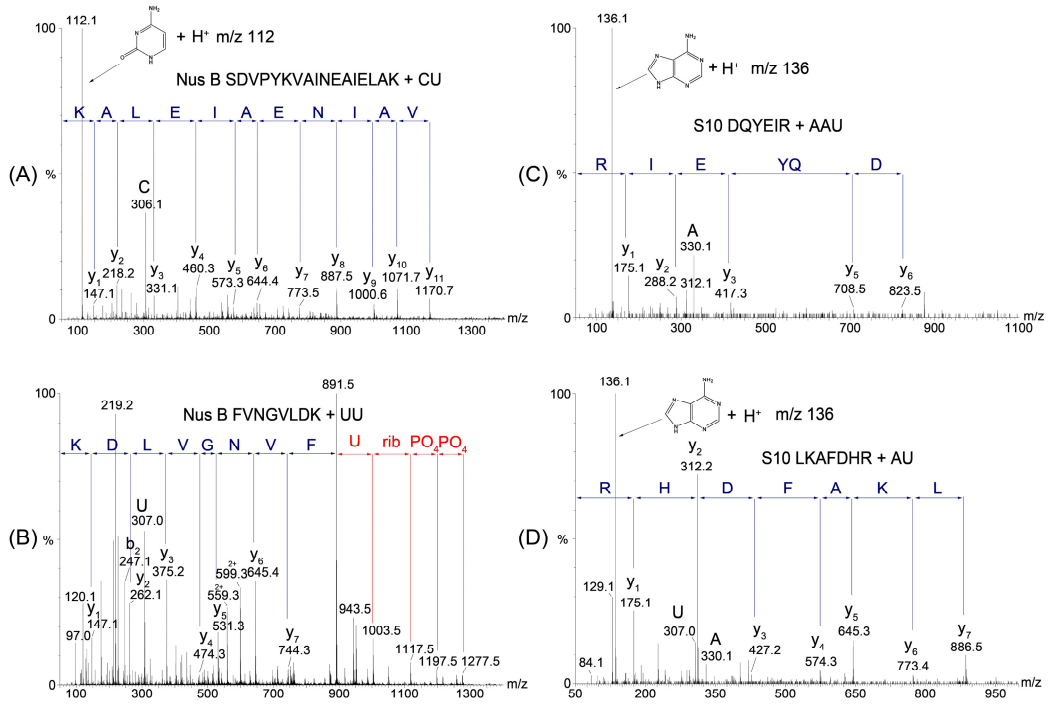


Figure 2.19. MS/MS spectra of protein-RNA crosslinks by TiO_2 enrichment. (A-D) Peptide-RNA crosslinks derived from UV-induced NusB-S10-λBoxA RNA complex (19mer λBoxA oligomer, CACCGCUCUUACACAAUUA) are NusB SDVPYKVAINEAIELAK+CU, NusB FVNGVLDK+UU, S10 DQYEIR+AAU and S10 LKAFDHR+UA, respectively,

Table 2.1. Identification of protein-RNA crosslinks in NusB-S10 complex by TiO_2 enrichment in combination with mass spectrometry.

NusB-S10			NusB					
λ BoxA	rrn BoxA	λ BoxA Core	λ BoxA	rrn BoxA	Exp. M.W.	Crosslinked Peptide+RNA	Peptide M.W.	RNA M.W.
X			X	X	2488.20	NusB SDVPYKVAINEAIELAK+CU	1859.00	629.08
X	X		X	X	2644.25	NusB RSDVPYKVAINEAIELAK+CU	2015.11	629.08
X			X		2807.03	NusB SFGAEDSHKFVNGVLDK+UAC	1848.90	958.13
X	X		X	X	1520.56	NusB FVNGVLDK+UU	890.48	630.06
				X	1826.57	NusB FVNGVLDK+UUU	890.48	936.09
X	X				1520.62	S10 LKAFDHR+UA-18	885.48	653.09
X	X	X			1538.70	S10 LKAFDHR+UA	885.48	653.09
	X				2541.88	S10 FTVLISPHVNK+AAUU	1253.71	1288.17
	X				2235.85	S10 FTVLISPHVNK+AAU	1253.71	982.14
X	X				1804.62	S10 DQYEIR+AAU	822.38	982.14

λ BoxA: CAC CGC UCU UAC ACA AUU A

rrn BoxA: CAC UGC UCU UUA ACA AUU A

λ BoxA core: CGC BrUCU UAC ACA; BrU - 5-bromo uridine

X indicates observation of a particular crosslink in a given mixture

In brief, mass spectrometric mapping of UV-induced protein-RNA crosslinks revealed that the NusB-S10 complex presents an intermolecular, composite, and contiguous binding surface for RNAs containing BoxA antitermination signals. Our TiO_2 enrichment procedure can be applied to any protein-RNA complex for the identification of proteins in direct contact with RNA and for visualization of the site of interaction.

2.5. Conclusion

A disposable TiO_2 microspin column for isolation of phosphorylated peptides of proteolytically digested proteins was fabricated. The advantages of such TiO_2 microspin column include, i) short enrichment and analysis time, ii) ease to use in standard routine applications for phosphopeptide analysis, iii) economy due to low costs, iv) high specificity and sensitivity for the enrichment of phosphorylated peptides from complex sample. It offers further the opportunity, to pack different chromatographic resins (C18, SCX, SAX, IMAC etc.) for different application or mixed resins to integrate several functions in one step prior to mass spectrometric analysis. In comparison with decomposed products of urea, loading sulfonic sodium which hydrolyzed from RepiGest™ SF onto the TiO_2 microspin column significantly increases the efficiency of phosphopeptide enrichment as sulfonic ion competes for binding sites with acidic amino acid residues in the peptides. An additional SDS-PAGE gel separation

prior to enrichment increases the number of detectable phosphopeptides. Our TiO_2 enrichment procedure in combination with different proteomic strategy was applied to phosphoproteome analysis of spliceosomal proteins, resulting in the identification of 1381 unique phosphorylation sites corresponding to 390 distinct proteins. At least 640 phosphorylation sites identified in this study appear to be novel, which may provide a valuable resource to the biological research community. In brief, we designed a microspin column and optimized a enrichment procedure that allowed us not only to carry out higher-throughput analyses assessing phosphorylation sites of multiple samples in a more efficient time scale but also to map UV-induced peptide-RNA crosslinks by mass spectrometer.

Chapter 3 - Efficient Enrichment of Intact Phosphoproteins prior to Mass Spectrometric Analysis

3.1 Summary

The selective enrichment of specific proteins prior to mass spectrometry (MS) analysis facilitates the identification of lower abundance proteins in current proteomic research. Up to now, only few methods have been described that allow for comprehensive purification of intact phosphoproteins from mixtures. In this study, we have developed a straightforward and reliable pre-fractionation procedure that effectively removes most of the highly abundant non-phosphorylated proteins, and allows for the selectively isolation of low abundant phosphoproteins based on calcium phosphate precipitation (CPP). By monitoring Coomassie blue staining of the model phosphoprotein beta-casein on SDS-PAGE to evaluate the efficiency, optimized sets of $CaCl_2$ / Na_2HPO_4 concentration and washing buffer conditions were realized for enrichment. Taking the optimized condition in combined with our in-house titanium dioxide (TiO_2) microspin column provided highly selective enrichment of phosphopeptides. The U1 small nuclear ribonucleoproteins (U1 snRNPs) were used as the complex samples to demonstrate the feasibility of this approach. Many phosphorylation sites identified in this study appear to be novel, including sites from U1-specific proteins (U1-70K, U1-A and U1-C) and Sm proteins (B'/B, D2, D3 and F).

3.2 Introduction

Reversible protein phosphorylation is the most important post-translational modification involved in many regulatory cellular functions such as the regulation of cell cycle, signal transduction, differentiation, proliferation, transformation, and metabolism [56, 57]. Almost 2 % of the human genome encodes protein kinases and an estimated one-third of all proteins in mammalian cells are expected to be phosphorylated [125]. Among the amino acid residues that can be phosphorylated, O-phosphates on serine, threonine, and tyrosine residues are by far the most abundant. The occurrence of phosphorylation on serine and threonine residues is more frequent than on tyrosine residue in a vertebrate cell, with the ratio of pSer/pThr/pTyr in the order of 1800:200:1.8 [57]. The phosphoramidates of arginine, histidine, and lysine also occur as do acyl derivatives of aspartic and glutamic acid, although they are less abundant.

Recently, MS has been widely applied as a powerful tool for the identification of proteins and the site assignment of protein modifications, including phosphorylation, taking the advantage of its speed, high sensitivity, reliability, and capability for determining phosphorylation site by tandem mass spectrometric techniques. However, large scale phosphoproteomic analysis still remains a substantial challenge due to the lower abundance of phosphoprotein, the substoichiometric nature of phosphorylation, and some technical limitations. The presence of

large amounts of non-phosphorylated peptides in the digestion product also suppresses the ion signal of phosphopeptides in MS detection [58-61]. Therefore, the development of efficient methods for highly specific enrichment of phosphoproteins and phosphopeptides prior to MS analysis becomes a critical step for studying protein phosphorylation.

Notwithstanding the complete separation of the phosphoproteins and phosphopeptides is still a challenging task, the analysis of phosphoproteome is one of the most exciting field in the current proteomic research. Although phosphoproteins can be separated by antibody immunoprecipitation [89, 126-130], the enrichment with anti-pSer, anti-pThr, or anti-pTyr antibodies depends on the affinity and specificity of antibodies, which limits comprehensiveness of the protein phosphorylation. Over the past decade, a variety of methodologies have been developed for phosphopeptide and phosphopeptide enrichment. One of these methods, immobilized metal ion affinity chromatography (IMAC), has been successful applied to enrich not only phosphopeptides but also phosphoproteins in large scale proteomic analysis [67-74]. However, multiple acidic residues of peptides are frequently retained by this procedure, interfering the further MS analysis [89-92]. The denaturing conditions or low pH was used that does not maintain enzyme activity of phosphoprotein. In addition, a promising phosphopeptide enrichment strategy was introduced by Sano et al., in which titanium TiO_2 had been used as an alternative to IMAC [75-79]. More recently, CPP had been proven to be very efficient for enriching phosphopeptides from a complex tryptic peptides by Xumin Zhang et al. , in which the phosphopeptides are pulled down by the formation of an insoluble calcium phosphate [69, 131]. Nevertheless, phosphoproteome studies are still hindered by lacking efficient methods for comprehensive purification of intact phosphoproteins until now.

Here, a straightforward and reliable phosphoprotein purification method was developed based on CPP. It opens up a new approach facilitating large scale phosphoproteomic analysis. The majority of non-phosphorylated proteins were removed by this procedure which then led to an enhanced detection of the lower abundant phosphoproteins. In initial studies, a protein mixture consisting of beta-casein, bovine serum albumin and myoglobin was used as model to optimize and test the procedure. We found that our method is highly tolerable with detergents. In addition, by coupling CPP with our in-house microspin column TiO_2 enrichment procedure, we were able to minimize co-enrichment of non-phosphorylated peptides. Finally, we have evaluated our approach on native U1 snRNPs complexes isolated from HeLa nuclear extract. A total of 192 unique phosphorylation sites corresponding to 45 distinct proteins were identified in the glycerol gradient purified U1 snRNPs. In summary, our method paves the way for pre-fractionation of phosphoproteins prior to enrichment of phosphopeptides which will be particularly useful in large scale phosphoproteome studies.

3.3 Experiment Sections

3.3.1 Materials

Chloroacetamide (CAA), ammonium bicarbonate, 2,5-dihydroxybenzoic acid (DHB) and trifluoroacetic acid (TFA) were obtained from Sigma-Aldrich (St. Louis, MO). Sequencing grade, modified trypsin was obtained from Promega (Madison, WI). Sequencing grade endoproteinase, Asp-N and Lys-C were obtained from Roche (Mannheim, Germany). RNase T1 and A were obtained from Ambion (Austin, TX). Calcium chloride ($CaCl_2$), disodium hydrogen phosphate (Na_2HPO_4), magnesium chloride ($MgCl_2$), benzonase nuclease, dithiothreitol (DTT), formic acid, ammonia solution, acetonitrile (ACN), and ethanol were obtained from Merck (Darmstadt, Germany). RapiGest™ SF was obtained from Waters Corporation (Manchester, UK). Titanium dioxide (TiO_2) resins were obtained from GL Sciences Inc. (Tokyo, Japan).

3.3.2 U1 snRNPs Purification

U1 snRNPs were purified from HeLa nuclear extract with mAb H20 by Prof. Reinhard Lührmann‘s Laboratory [93] and fractionated by glycerol gradient centrifugation by Monika Raabe [94]. The resulting complex was treated RNase (T1, A, Benzonase) at 37 ℃ for 2 hours, and subsequently at 52 ℃ for 2 hours. After, the proteins were isolated either by ethanol or acetone precipitation. The resulting pellet was resuspensed by 0.5 % RapiGest™ SF in 50 mM HEPES at pH 7.5, and diluted to 0.1 % RapiGest™ SF with 50 mM HEPES at pH 7.5 for CPP.

3.3.3 Ethanol Precipitation

U1 snRNPs were precipitated by adding 3 volumes of ethanol and 1/10 volume of 3 M sodium acetate, pH 5.3. The mixture was vortexed, incubated at -20 ℃ for 2 hours and then centrifuged 17000 g at 4 ℃ for 30 min. The supernatant was removed and the pellet was washed with 500 µl of 80 % ethanol and centrifuged as above. Discarded supernatant, the pellet was evaporated with a SpeedVac.

3.3.4 Acetone Precipitation

U1 snRNPs were precipitated by adding 4 volumes of acetone at -20 ℃ for 2 hours and then centrifuged 17000 g at 4 ℃ for 30 min. The supernatant was removed and the pellet was washed with 500 µl of 80 % acetone and centrifuged as above. Discarded supernatant, the pellet was evaporated with a SpeedVac.

3.3.5 Optimized Calcium Phosphate Precipitation

Phosphoproteins were precipitated by adding one-tenth volumes of 100 mM $CaCl_2$ in 50 mM HEPES, pH 7.5 for 10 min and then adding another one-tenth volumes of 50 mM $NaHPO_4$ in 50 mM HEPES, pH 7.5 for 10 min to form co-precipitation. The solution was centrifuged at 17000 x g for 10 min, removed the supernatant to eppendorf, and 10 µl of 50 mM $CaCl_2$ in 50 mM HEPES, pH 7.5 was applied to wash the pellet. After centrifugation as described above, the washing solution was collected to the same eppendorf, and the resulting pellet was dissolved in 5 µl of 1 % RapiGest™ SF in 25mM ammonium bicarbonate, pH 8.5, sonicated for 15 min, and subsequently analyzed by 1-D SDS-PAGE or performed in solution enzymatic digestion.

3.3.6 In-Solution Digestion

The protein solution was reduced with 5 µl of 10 mM DTT at 37 ℃ for 1 hour, alkylated with 10 µl of 20 mM CAA at 37 ℃ for 1 hour, diluted to 0.1 % RapiGest™ SF with 30 µl of 25 mM ammonium bicarbonate, and subsequently digested with either trypsin or Asp-N/Lys-C at 37 ℃, overnight (the enzyme to substrate ratio is 1:20). The peptides were acidified with 50 µl of 5 % TFA at 37 ℃ for 2 hours and centrifuged at 17000 g for 10 min. Afterward, the supernatant was transferred to another eppendorf and dried on a SpeedVac for further analysis.

3.3.7 TiO_2 Enrichment Procedure

Phosphopeptides were enriched by TiO_2 as described (chapter 2). In brief, aliquots of peptides were dissolved with 20 µl of 200 mg DHB in 1 ml of 80 % ACN, 5 % TFA, and then loaded onto TiO_2 column. The column was washed 3 times with 20 µl of 200 mg DHB in 1 ml of 80 % ACN, 5 % TFA and 5 times with 20 µl 80 % ACN, 5 % TFA. Bound peptides were eluted 3 times with 20 µl of 0.3 N NH_4OH, pH ≥ 10.5 and subsequently evaporated with a SpeedVac for further MS analysis.

3.3.8 Mass Spectrometry Analysis

For MALDI-MS analysis, the resulting peptides were mixed with a 10 mg/ml matrix solution of DHB in 70 % ACN with 0.1 % trifluoroacetic acid, and then spotted onto MALDI target plate. Subsequently, MALDI-MS detection was performed on a 4800 MALDI ToF/ToF mass spectrometry (Applied Biosystems/MDS Sciex) equipped with an Nd:YAG laser (355 nm wavelength, and 200 Hz repetition rate). 1000 shots were accumulated in positive ion mode MS. For the nanoLC-ESI MS/MS analysis, the resulting peptides were first loaded at a flow rate of 10 µl/min onto an in-house packed C18 trap column (1.5 cm, 360 µm o.d., 150 µm i.d., Reprosil-Pur 120 Å, 5 µm, C18-AQ, Dr. Maisch GmbH, Germany). The retained peptides were then eluted and separated on an analytical C18 capillary column (15 cm, 360 µm o.d., 75 µm i.d., Reprosil-Pur 120 Å, 5 µm, C18-AQ, Dr. Maisch GmbH, Germany) at a flow rate of 300

nL/min, with a gradient from 7.5 to 37.5 % ACN in 0.1 % formic acid for 60 min using an Agilent 1100 nano-flow LC system (Agilent Technologies, Palo Alto, CA), coupling with LTQ-Orbitrap XL hybrid mass spectrometer (Thermo Electron, Bremen, Germany). The LTQ-Orbitrap was operated in the data-dependant mode. Briefly, survey full scan MS spectra were acquired in the Orbitrap (m/z 350–1600) with the resolution set to 30,000 at m/z 400 and automatic gain control (AGC) target at 10^6. The five most intense ions were sequentially isolated for CID MS/MS fragmentation and detection in the linear ion trap with previously selected ions dynamically excluded for 90 second. Ions with singly and unrecognized charge state were also excluded. To improve the fragmentation spectra of the phosphopeptides, “multistage activation” corresponding to a neutral loss of phosphoric acid from doubly- and triply charged precursor ion was enabled in all MS/MS events [132].

3.3.9 MASCOT Database Searching

All spectra were searched MASCOT server v2.2.06 with the decoy database searching option (for assessing false positive rate) against the International Protein Index (IPI) human database (v3.62, 83947 protein sequences) with criteria-peptide mass tolerance, 7 ppm; MS/MS ion mass tolerance, 0.5 Da; allow up to three missed cleavage; variable modifications considered were phosphorylation of serine, threonine and tyrosine, methionine oxidation and cysteine carboxyamidomethylation. Filtering criteria were established based on decoy database search results to provide a <1 % false positive rate (FPR) for overall peptide identifications. All phosphorylated sited were examined manually by the presence of a 69 Da between fragment ions for phosphoserine and an 83 Da for phosphothreonine.

3.4 Result and Discussion

3.4.1 Comparison of Efficiency and Specificity among Different Alkaline Earth Metal

The identification of low abundant phosphoproteins and the subsequent identification of the actual phosphorylation site(s) typically necessitate their enrichment prior to MS analysis. Our goal was to develop a straightforward, sensitive, and MS compatible method for the enrichment of intact phosphoproteins. The basic concept originated from the lowered solubility product of phosphate group when formed a complex with divalent alkaline earth metal cations in aqueous solution.

To demonstrate our speculation, we decided to first carry out preliminary experiment with the same concentration ratio of $MgCl_2$ / Na_2HPO_4 and $CaCl_2$ / Na_2HPO_4 (2:1) for segregation of phosphoproteins. In this study, beta-casein was used as a model phosphoprotein due to its multiple phosphorylation sites [71, 75, 133]. 10 µl of standard protein solution, consisting of a

total amount of 5 μg each of beta-casein, BSA and myoglobin in 50 mM HEPES buffer solution (pH 7.5), was utilized for examine the efficiency of phosphoprotein enrichment. One-tenth volume of 50, 100, 250 and 500 mM $MgCl_2$ or $CaCl_2$ solution was added to sample solution for 10 min, and followed additional one-tenth volume of 25, 50, 125 and 250 mM $NaHPO_4$ was added for another 10 min to form co-precipitation of phosphoprotein, respectively. The resulting protein pellet could be observed after high speed centrifugation (17000 x g for 10 min). After, the supernatant and pellet were analyzed by 1-D SDS-PAGE to visualize their effects shown in Figure 3.1.

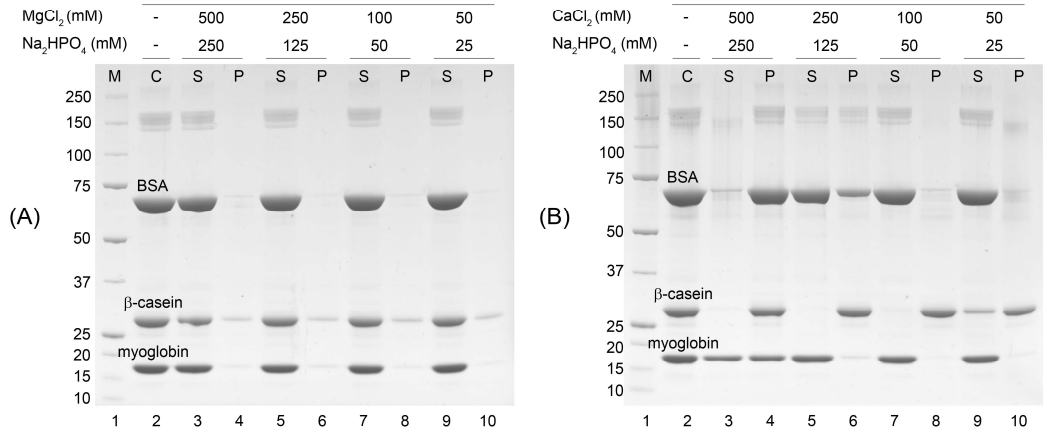


Figure 3.1. SDS-PAGE (4-12 %, Coomassie blue staining) analysis of standard proteins precipitated by alkaline earth metal. (A) Magnesium phosphate precipitation, 5 μg each of beta-casein, BSA and myoglobin were used to precipitate model phosphoprotein, beta-casein, by adding $MgCl_2$ / Na_2HPO_4 in concentration order from 500 mM / 250 mM (Lane 3 and 4), 250 mM / 125 mM (Lane 5 and 6), 100 mM / 50 mM (Lane 7 and 8) to 50 mM / 25 mM (Lane 9 and 10), respectively. (B) Calcium phosphate precipitation (CPP), instead of $MgCl_2$ / Na_2HPO_4, $CaCl_2$ / Na_2HPO_4 was used to precipitate beta-casein with the same concentration. M: Protein marker. C: Control, standard proteins. S: Supernatant. P: Pellet.

By using $MgCl_2$ / Na_2HPO_4 precipitation of phosphorylation (Figure 3.1A), most of the proteins, beta-casein, BSA and myoglobin, were present in the supernatant. Only a few amount of beta-casein could be effectively separated into pellet fraction and the loss of beta-casein took place. Even the concentration of $MgCl_2$ / Na_2HPO_4 increased up to 500 mM / 250 mM, the majority of beta-casein still retained in supernatant (Lane 3 and 4). With respect to $MgCl_2$ / Na_2HPO_4 precipitation, $CaCl_2$ / Na_2HPO_4 seemed to be more efficiency to isolate beta-casein from BSA and myoglobin (Figure 3.1B). As 50 mM / 25 mM of $CaCl_2$ / Na_2HPO_4 was used to precipitate, more than half of beta-casein was precipitated into pellet fraction (Lane 10). But, it

still remained a smaller part of beta-casein in the supernatant (Lane 9). Until the concentration of $CaCl_2$ / Na_2HPO_4 raised to 100 mM / 50 mM, the majority of beta-casein was segregated into pellet fraction (Lane 8), and non-phosphorylated proteins still retained in the supernatant (Lane 7). However, while the concentration of $CaCl_2$ / Na_2HPO_4 increased up to 250 mM / 125 mM, BSA started to form precipitation (Lane 6). Under high concentration of $CaCl_2$ / Na_2HPO_4 (500 mM / 250 mM), almost all BSA and half of myoglobin were present in the pellet fraction (Lane 4). Hence, we found out optimized concentration of $CaCl_2$ / Na_2HPO_4 was 100 mM / 50 mM. According to these results, we can draw a short conclusion that $CaCl_2$ / Na_2HPO_4 has a strong ability to efficiently isolate phosphoproteins.

3.4.2 Optimal Condition for Phosphoprotein Precipitation

We further investigated whether the results could be improved by altering Na_2HPO_4 concentration. The standard proteins were performed CCP by using 100 mM $CaCl_2$ in combined with 250, 125, 50 and 25 mM Na_2HPO_4, respectively. The resulting pellet and supernatant were analyzed by SDS-PAGE shown in Figure 3.2. Under 250 mM Na_2HPO_4 condition, non-phosphorylated proteins could be formed precipitation (Lane 2 and 3). No significant difference was observed by using 50 or 125 mM Na_2HPO_4 (Lane 4 to 7). All beta-casein could be fractionated into pellet fraction and non-phosphorylated proteins still remained in the supernatant. Under 25 mM Na_2HPO_4 condition, the result indicated this concentration did not provided sufficient phosphate anion to co-precipitate beta-casein with 100 mM $CaCl_2$, leading to a smaller part of beta-casein in the supernatant (Lane 8 and 9). The result reveals that beta-casein only be isolated within 50 to 125 mM Na_2HPO_4.

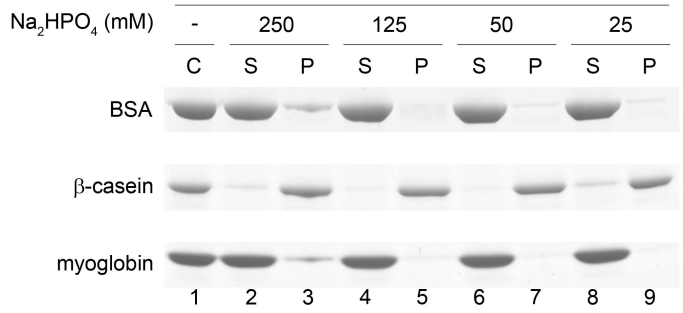


Figure 3.2. SDS-PAGE analysis of standard protein under different concentration of Na_2HPO_4 to optimize CPP procedure. One-tenth volume of 100 mM $CaCl_2$ was added to 5 μg each of beta-casein, BSA and myoglobin solution, and followed additional one-tenth volume of 250, 125, 50 and 25 mM Na_2HPO_4 to form co-precipitation, respectively. C: Control, standard proteins. S: Supernatant. P: Pellet.

In this context, the selective isolation of phosphoprotein would necessitate the incorporation of not only $CaCl_2$ / Na_2HPO_4 concentration but also a wash buffer solution that provides sufficient strength to compete off all other non-phosphorylated proteins. This would leave almost all phosphoproteins in pellet, which could then be directly applied to digest for MS and MS/MS analysis. The CPP resulting pellet was washed 10 µl of 50 mM $CaCl_2$ or 25 mM Na_2HPO_4 solution either one time or two times shown in Figure 3.3A. Comparative analyses showed that 50 mM $CaCl_2$ performed significantly better than 25 mM Na_2HPO_4 for retaining beta-casein and washing off the non-phosphorylated proteins (Lane 2 to 5). The obtained pellet of beta-casein could be displaced with phosphate anion, leading to low yields (Lane 6 to 9). In addition, one time wash of $CaCl_2$ was sufficient. The signal of beta-casein was absent in the supernatant which is indicative of its minimal loss from at this stage (Lane 2 and 3). We also tested the effect of pH value in wash buffer solution, as shown in Figure 3.3B. The CPP resulting pellet was washed with 25 mM $CaCl_2$ in 50 mM HEPES, pH 7.5; 0.1 % formic acid (FA), pH 3; 0.1 % trifluoroacetic acid (TFA), pH 2 or 0.3 N ammonium hydroxide (NH_4OH), pH 10.5, respectively. A low abundance signal of beta-casein could be detected in the supernatant with respect to FA, TFA and NH_4OH washing buffer (Lane 4, 6 and 8). Our result suggests that it is better to use 50 mM HEPES pH 7.5 to wash off non-phosphorylated proteins, which is also compatible for further enzymatic digestion and MS analysis.

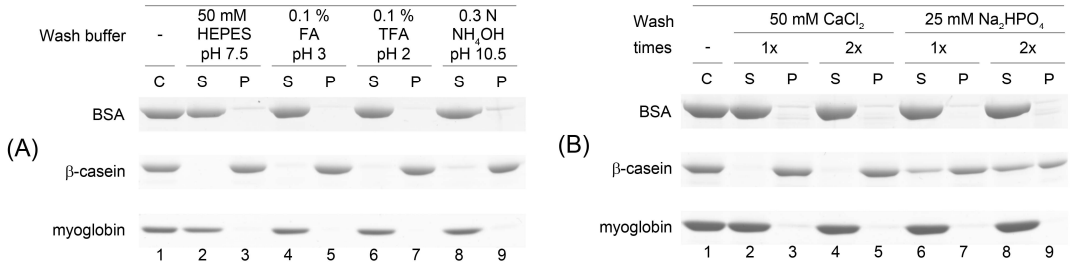


Figure 3.3. SDS-PAGE analysis of standard protein by using different wash conditions to optimize CPP procedure. (A) After CPP ($CaCl_2$ / Na_2HPO_4, 100 mM / 50 mM), the resulting protein pellet was washed with 10 µl of 50 mM $CaCl_2$ or 25 mM Na_2HPO_4 either one time or two times. (B) After CPP ($CaCl_2$ / Na_2HPO_4, 100 mM / 50 mM), the protein pellet was washed with 50 mM $CaCl_2$ in 50 mM HEPES buffer, 0.1 % formic acid (FA) , 0.1 % trifluoroacetic acid (TFA), or 0.3 N ammonium hydroxide (NH_4OH), respectively. C: Control, standard proteins. S: Supernatant. P: Pellet.

Here, we propose two possibilities that allow the formation of co-precipitation of beta-casein. One is due to the phosphate groups on beta-casein that can interact with calcium chloride and disodium hydrogen phosphate to form insoluble calcium-phosphate-protein complex. The other is owing to the structure property of beta-casein. To clarify the two possibilities, the

standard protein mixture was first dephosphorylated by alkaline phosphatase and performed the optimized CPP procedure shown in Figure 3.4. The result indicated beta-casein was no longer precipitating while phosphate groups on beta-casein were removed by alkaline phosphatase (Lane 5 and 6). This result indeed confirms our speculation that phosphoprotein can be isolated by our CPP based on their solubility.

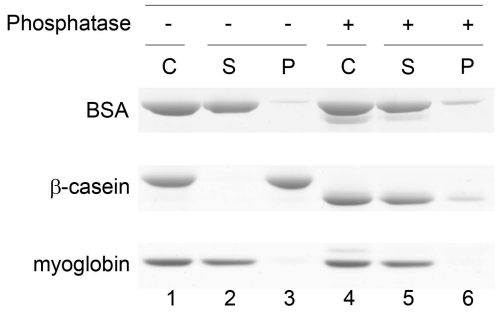


Figure 3.4. Proof of the CCP procedure by virtue of the formation of insoluble calcium-phosphate-protein complex. The standard proteins, 5 µg of each beta-casein, BSA and myoglobin, were performed with optimized CPP procedure to segregate model phosphoprotein, beta-casein, shown in Lane 1, 2 and 3. Before CPP, the standard proteins were treated with alkaline phosphatase, shown in Lane 4, 5 and 6. C: Control, standard proteins. S: Supernatant. P: Pellet.

3.4.3 Effect of Different Denatured Reagents

For shotgun proteomic analysis, proteins must be soluble and denature in an aqueous solution for enzymatic digestion. The most frequent method of dissolving proteins is the use of ionic detergents such as sodium dodecyl sulphate (SDS). To test whether our CPP method is compatible to SDS, 5 µg each of beta-casein, BSA and myoglobin was dissolved with 0.5, 0.1, 0.05 or 0.01 % SDS in 50 mM HEPES (pH 7.5) buffer solution, respectively, and isolation of beta-casein with optimized CPP procedure was followed, as shown in Figure 3.5.

The majority of beta-casein was preferentially present in the supernatant under 0.5 % SDS condition (Lane 3). Although beta-casein could form precipitation under 0.1 % and 0.05 % SDS conditions, a smaller proportion of BSA and myoglobin was also observed in the pellet fraction, which might due to the formation of protein micelles with SDS reagent (Lane 6 and 8). In the presence of 0.01 % SDS, beta-casein was segregated again, inferring the concentration of SDS had to reduce less than 0.01 % SDS in sample solution prior to CPP enrichment (Lane 9 and 10). The result indicated that the maximum concentration of SDS in sample solution had to be diluted less than 0.01 % before the execution of CPP process.

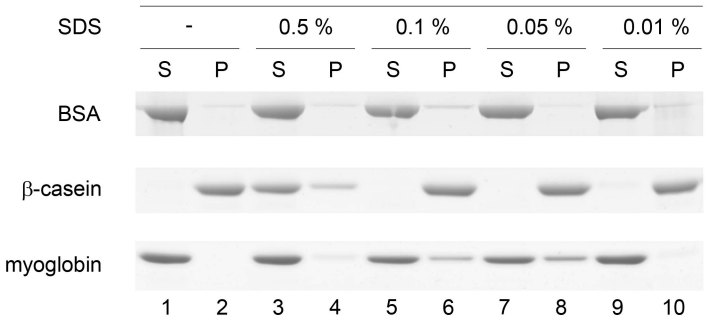


Figure 3.5. SDS-PAGE analysis of the influence of SDS detergent on our CPP method. 5 μg each of beta-casein, BSA and myoglobin were dissolved in 0.5, 0.1, 0.05, and 0.01 % SDS, respectively, and phosphoprotein was subsequently isolated by optimized CPP procedure.

We also tested another strong protein denaturant, urea, which unfold the protein and is widely used to dissolve sample in proteomic research [99, 134-136]. The same amount of standard proteins was dissolved with 0.6, 0.4, 0.2 and 0.1 N urea in 50 mM HEPES (pH 7.5) buffer solution, respectively, and then precipitated by optimized CPP procedure shown in Figure 3.6. Although the majority of non-phosphorylated proteins were remained in the supernatant, a trace proportion of base-casein was observed in the pellet fraction under 0.6, 0.4 and 0.2 N urea condition (Lane 3, 5 and 7). Until the concentration of urea was decreased to 0.1 N, the isolation of beta-casein by CPP was reconstructed (Lane 9 and 10). The result indicated that the maximum concentration of urea in sample solution should be diluted less than 0.1 N before the execution of CPP.

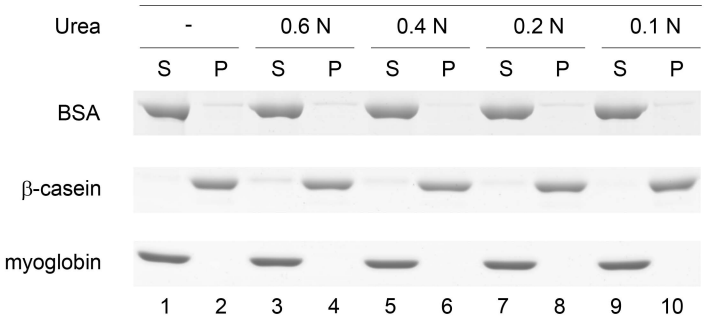


Figure 3.6. SDS-PAGE analysis of the influence of denature reagent, urea, on our CPP method. 5 μg each of beta-casein, BSA and myoglobin were dissolved with 0.6, 0.4, 0.2 and 0.1 N urea, respectively, and subsequently isolated phosphoprotein by optimized CPP procedure.

Our final goal is to develop a straightforward and reliable method that can be applied to global

proteome study. Not only pellet but also supernatant had to be analyzed. Notwithstanding SDS and urea under 0.01 % and 0.1 N, respectively, have a good efficiency for phosphoprotein enrichment, both reagents have shortcomings in shotgun proteomic method based on in-solution enzymatic digestion of proteins prior to MS analysis. The SDS is trapped on the reverse phase column and forms a cluster of ions during ESI ion source inferring peptide detection in MS. The urea caused carbamylation at N-terminus of peptide residue or lysine residue [104-106], increasing the complexity of sample. Thus, we have to remove SDS and urea by running SDS-PAGE or other chromatographic technology prior to MS detection.

For this reason, we further tested a MS compatible denatured surfactant, RapiGest™ SF, an acid-cleavable anionic detergent that is used to enhance enzymatic digestion of proteins under 0.1 % condition, and can be removed by high speed centrifugation [107]. 5 μg each of standard proteins were dissolved with 0.5, 0.1, 0.05 and 0.01 % RapiGest™ SF in 50 mM HEPES buffer solution, respectively, and subsequently precipitated by optimized CCP procedure in Figure 3.7.

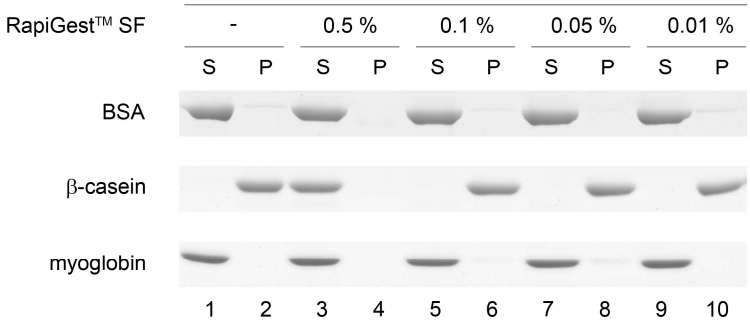


Figure 3.7. SDS-PAGE analysis of the influence of RapiGest™ SF on our CPP method. 5 μg each of beta-casein, BSA and myoglobin were dissolved with 0.6, 0.4, 0.2 and 0.1 N urea, respectively, and subsequently isolated phosphoprotein by optimized CPP procedure.

Under 0.5 % RapiGest™ SF condition, all these three standard proteins were remained in the supernatant during the CPP process (Lane 3 and 4). Under 0.1, 0.05 and 0.01 % RapiGest™ SF conditions, the majority of beta-casein was present only in the pellet fraction (Lane 6, 8 and 10), and BSA and myoglobin were still retained in the supernatant (Lane 5, 7 and 9). The result indicated our CPP method could tolerate upto 0.1 % RapiGest™ SF that was susceptible to enzymatic cleavage without inhibiting enzyme activity. It is worth to note that Xumin et al. used calcium phosphate to precipitate phosphopeptides, and the resulting pellet was dissolved in 5 % formic acid prior to MS analysis [69]. However, 5 % formic acid inhibits most of in-solution

enzymatic digestion such as trypsin, Asn-N, Glu-C and so on. Taken together, the advantages of using RapiGest ™ SF include its ability to dissolve the protein complex pellet under 0.5 % condition, compatible for CPP under 0.1 % condition, improving enzymatic digestion, and acid-labile that can be easily removed by high speed centrifugation without the interference of the followed MS analysis.

3.4.4 Examine the Power of our Phosphoprotein Isolation Method

To assess the efficiency and specificity of our CPP procedure for enrichment of phosphoprotein, different amounts of beta-casein (5, 3.75, 2.5, 1.25, 0.5 and 0.25 μg, respectively) containing 5 μg BSA and 5 μg myoglobin were prepared for examining the power of our CPP procedure shown in Figure 3.8.

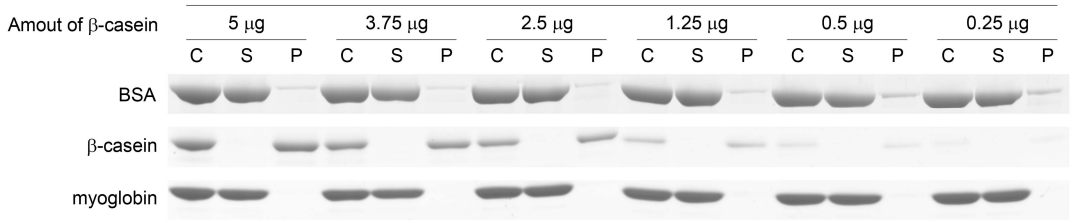


Figure 3.8. The efficiency and specificity of our CPP procedure. Different amount of beta-casein from 5, 3.75, 2.5, 1.25, 0.5 to 0.25 μg containing of 5 μg BSA and 5 μg myoglobin in solution was performed optimized CCP procedure, and subsequently analyzed by SDS-PAGE, respectively. C: Control, standard proteins. S: Supernatant. P: Pellet.

The result indicated that most of BSA and myoglobin were present in the supernatant. Beta-casein only remained in the pellet fraction, even in low amount of beta-casein (0.25 μg, approximately 10 pmol, close to the limit of detection of Coomassie blue staining). However, a smaller proportion of BSA still retained in the pellet fraction, indicating that our CPP method is still not sufficient for isolating 20 fold substoichiometric levels of phosphoprotein for subsequently unseparated enzymatic digestion of protein analysis by MALDI-MS. An additional enrichment for phosphopeptides is necessitated to increase the specificity by reducing the amount of non-phosphopeptides. So, we coupled our CPP method with in-house TiO_2 microspin column enrichment. The CPP resulting proteins, either pellet or supernatant, were digested with trypsin, enriched with in-house TiO_2 microspin column and then spotted on MALDI target plate for MALDI-MS analysis. Figure 3.9.A showed MALDI-MS spectrum of tryptic peptides derived from 0.25 μg beta-casein containing 5 μg BSA and 5 μg myoglobin. However, signals of phosphopeptides derived from beta-casein at m/z 2061.9 and 3122.3 were not detected in the mass spectrum. That is because the signal corresponding to

phosphopeptides which is readily detectable in MS is suppressed in the company of other more readily ionized non-phosphorylated peptides of BSA and myoglobin. Even these peptides were enriched with TiO_2, only one phosphopeptide of beta-casein at m/z 2061.9 was observed in the mass spectrum shown in Figure 3.9D.

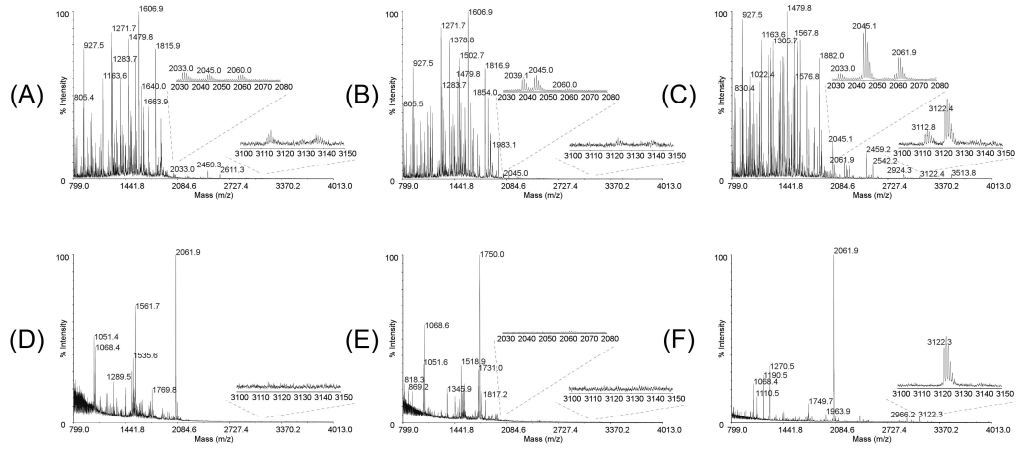


Figure 3.9. The identification of phosphopeptide using CPP method coupling with in-house TiO_2 microspin column. MALDI mass spectrum of tryptic peptides derived from 0.25 µg beta-casein containing of 5 µg BSA and 5 µg myoglobin in solution, one two-hundredth tryptic peptides were spotted on the MALDI target plate. (A) and (D) without CPP; (B) and (E) the CPP resulting supernatant; (C) and (F) the CPP resulting pellet. (D), (E) and (F) with an additional phosphopeptide enrichment with in-house TiO_2 microspin column. Two signals of phosphopeptides derived from beta-casein at m/z 2061.9 and 3122.3 are FQpSEEQQQTEDELQDK and RELEELNVPGEIVEpSLpSpSpSEESITR, respectively.

Possible explanation of not detecting tetraphosphorylated peptide at m/z 3122.3 in the mass spectrum may due to large amount of non-phosphorylated peptides in solution which could inference TiO_2 enrichment. After CPP, the spectrum obtained from tryptic peptides of supernatant (Figure 3.9B) was almost identical to that of the original sample (Figure 3.9A). All signals in the spectrum belonged to tryptic peptides of BSA and myoglobin. Even an additional TiO_2 enrichment was performed, the signals corresponding to phosphopeptides of beta-casein at m/z 2061.9 and 3122.3 were still not detectable in the spectrum shown in Figure 3.9E, indicating that the loss of phosphoprotein is minimal during the entire CPP procedure. The CPP resulting pellet was redissolved in 0.5 % RapiGest™ SF buffer solution and then was diluted to 0.1 % for trypsin digestion. The respective phosphopeptides for beta-casein, including the signals at m/z 2061.9 and 3122.3 which were not detected before were now observed shown in Figure 3.9C; however, the spectrum still accompanied with other signals of tryptic peptides

derived from BSA and myoglobin. In combination with TiO_2 enrichment, the majority portion of non-phosphorylated peptides were removed by TiO_2 and phosphopeptide signals derived from beta-casein at m/z 2061.9 and 3122.3 dominated the spectrum shown in Figure 3.9F. This result indeed indicated that our CPP and TiO_2 were complementary methods for trace level phosphopeptides enrichment. Taking cues from previous findings, the CPP procedure can specifically and effectively concentrate phosphoproteins from the complex protein mixture and further TiO_2 enrichment of phosphopeptides from CPP resulting pellet enable efficient recovery of phosphopeptides, even when substoichiometric levels of the non-phosphorylated protein is 20 folds higher than that of the phosphoprotein.

3.4.5. Identification of Phosphorylation Sites from U1 Small Nuclear Ribonucleoproteins

To investigate whether our CPP method in coupled with TiO_2 enrichment was applicable to a very complicated biological sample for identification of phosphorylation site, the method was tested in one of our ongoing projects, the U1 small nuclear ribonucleoproteins (snRNPs). The snRNPs are the protein complexes combined with pre-mRNA to form a spliceosome. The pre-mRNA splicing can be regulated both positively and negatively by reversible protein phosphorylation. Figure 3.10 showed the SDS-PAGE analysis of 10 μg U1 snRNPs fractioned by CPP method.

However, the pre-mRNA contains phosphate groups, and the strong protein-protein and protein-RNA interaction occur in U1 snRNPs, resulting in the co-precipitation of proteins which does not contain any phosphorylation shown in Lane 3 and 4. To minimize the affection, 0.5 % RapiGest ™ SF was used to denature U1 snRNPs, and then diluted to 0.1 % RapiGest ™ SF for the CPP of phosphoproteins (Lane 5 and 6). Nevertheless, the existing of pre-mRNA in solution also binds to TiO_2 strongly, and interferes further phosphopeptides detection in MS analysis. Hence, prior to CPP enrichment, the U1 snRNPs was treated with ribonuclease (RNase T, T1 and benzonase) to cleave pre-RNA into smaller components, precipitated proteins with either acetone or ethanol to remove small RNA, and subsequently redissolved by RapiGest™ SF shown in Lane 7 to 10, respectively. The result indicated that acetone precipitation of U1 snRNPs had better proteins recovery than ethanol.

After RNase treatment, acetone precipitation, and CPP method, we kept the resulting supernatant and pellet in 0.1 % RapiGest™ SF, that was compatible for further in-solution enzymatic digestion in coupled with in-house TiO_2 microspin column enrichment linked to MS analysis for identification of phosphorylated sites in U1 snRNPs.

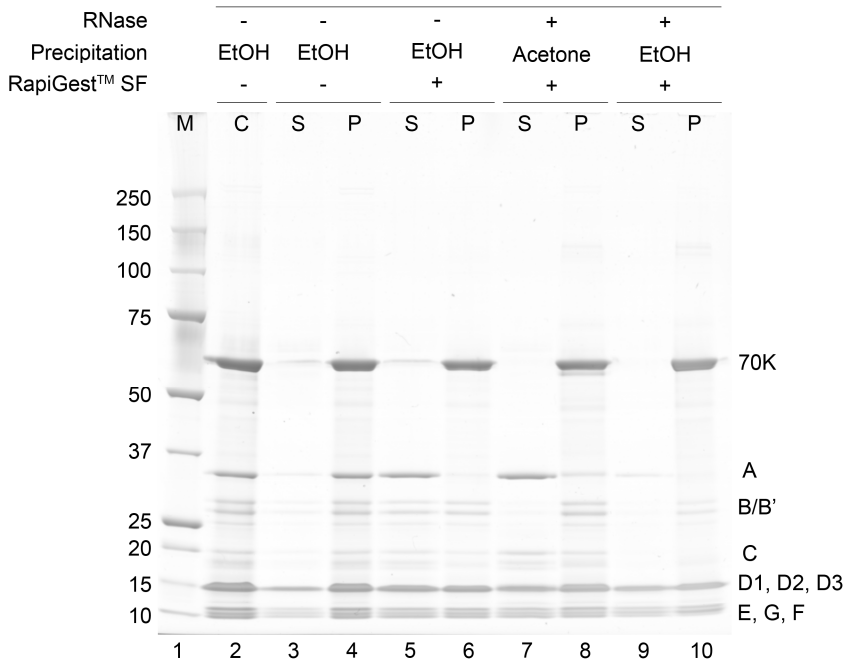


Figure 3.10. SDS-PAGE analysis of 10 µg of U1 snRNPs extracted from HeLa cell. The extracted U1 snRNPs were purified by EtOH precipitation (Lane 2), dissolved in 50 mM HEPES buffer, and subsequently performed with optimized CPP (Lane 3 and 4). Instead of HEPES buffer, RapiGest™ SF was used to dissolve U1 snRNPs (Lane 5 and 6). Before EtOH precipitation, the RNase was treated to cleavage pre-mRNA (Lane 9 and 10). Instead of EtOH precipitation, acetone was used to purify U1 snRNPs (Lane 7 and 8). M: Protein marker. C: Control. S: Supernatant. P: Pellet.

Table 1 provided a partial list of proteins found in the U1 snRNPs. Three U1-specific proteins (U1-70K, U1-A and U1-C) and Sm proteins (B'/B, D1, D2, D3, E, F and G) were identified with high MASCOT score on both fraction. However, the phosphopeptide was not detectable in the supernatant without TiO_2 enrichment. Even these peptides were enriched with TiO_2, only 4 phosphorylation sites were found. In comparison, a total of 192 unique phosphorylation sites corresponding to 45 distinct proteins had been identified from pellet fraction (Appendix 3). All phosphorylated sites were confirmed by manual validation. 59 phosphorylation sites identified in this study have not been reported previously, searching against Expasy Knowledgebase (http://us.expasy.org/) and PhosphoSite database (http://www.phosphosite.org/) [114]; these two databases that have collected comprehensive information on in vivo protein phosphorylation and references.

Table 3.1. Partial list of phosphorylation sites identified from the leading proteins of U1 snRNPs by using CPP method in combination with in-house TiO_2 microspin column enrichment. For all phosphorylation sites, see Appendix 3. The underline indicates that the phosphorylation site was not identified previously.

		Supernatant			Pellet		
		w/o TiO_2		TiO_2	w/o TiO_2		TiO_2
		Mascot score	p-site	p-site	Mascot score	p-site	p-site
U1 specific	70K	618		S226; S266; S268	825	S117; S410	Y38; Y112; S117; Y126; Y206; S211; S226; S245; S257; S259; S266; S268; S281; S293; S295; S385; S410
	A	624			390		S115
	C	273			232		Y5; S17
Sm proteins	B/B'	352			516		T30
	D1	108			177		
	D2	393			209		T12; T40
	D3	346			339		S2; S44
	E	174			241		
	F	266			233	S2	S2
	G	98			89		

The U1-70K protein is a heavily phosphorylated protein and its reversible phosphorylation is required for the first step of splicing [137, 138]. We found 17 phosphorylation sites on U1-70K out of which 11 phosphorylation sites were not previously described. The C-terminal half of U1-70k, which contains the RS1 and RS2 regions and has been implicated in protein-protein interaction with SR proteins that mediate 5' splice site recognition [139], is identified to be phosphorylated on serine 245, 257, 259, 266, 268, 281, 293, 295 and 385. The U1-specific proteins (U1-A and U1-C) and the common Sm proteins (B'/B, D2, D3 and F), which are important to splicing activity [140], are also found to be phosphorylated. A total of 38 phosphorylation sites corresponding to 9 SR proteins have been identified in this study. SR proteins facilitate the association of snRNPs with the pre-mRNA during spliceosome assembly, and phosphorylation is essential for SR protein activity [116, 117] and enhances their interactions with other spliceosomal proteins [118].

3.5. Conclusion

Our method opens up a new approach for large scale phosphoproteome study using calcium chloride and disodium hydrogen phosphate to precipitate low abundant phosphoproteins coupling with TiO_2 to enrich phosphopeptides. The phosphoproteins precipitation by calcium chloride is more efficiently than magnesium chloride owing to the lower solubility product constants of calcium phosphate ($Ca_3(PO_4)_2$, Ksp = $1x10^{-32}$; $Mg_3(PO_4)_2$, Ksp = $1x10^{-24}$). Furthermore, we have demonstrated that several commercialyl available denatured reagents, SDS, urea and RapiGest ™ SF, are compatible with our method. Coupling with in-house TiO_2 microspin column increases the specificity toward phosphopeptides by reducing the amount of non-phosphorylated peptides. Even in complex biological samples such as U1 snRNPs, trace phosphopeptides derived from U1 snRNPs can be effectively identified by the CPP in combination with TiO_2 enrichment. In summary, our CPP method for phosphoprotein isolation has several advantages - easy to use, sensitivity, selectivity and inexpensive, which can be strategically adapted for concerted extraction and pre-concentration of both non-phosphorylated and phosphorylated proteins prior to MS analysis in proteomic applications.

Chapter 4 - Pseudo-Neutral-Loss Scan for Selective Detection of Phosphopeptides and *N*-Glycopeptides using Liquid Chromatography Coupled with a Hybrid Linear Ion-Trap / Orbitrap Mass Spectrometer

4.1 Summary

In this study, we describe a novel method, termed “pseudo-neutral-loss scan”, for selectively probing phosphopeptides and glycopeptides on a hybrid LTQ-Orbitrap mass spectrometer (MS). The instrument has been programmed such that phosphopeptides or glycopeptides eluted from reverse phase liquid chromatography (LC) can automatically be discovered and identified in a way similar to that of the use of neutral-loss scanning. An in-source collision-induced dissociation (CID) energy is applied on all species ionized in electrospray ion source, resulting in a neutral loss(es) of phosphoric acid or monosaccharide residues. Subsequently, the characteristic mass pair that differ in the mass of a neutral loss of phosphoric acid or monosaccharide residues are automatically selected for CID MS/MS to obtain modified peptide sequence information. However, the CID MS/MS for glycopeptide predominantly generates fragment ions from cleavages of glycosidic bonds without breaking the peptide bonds. Therefore, a targeted MS3 acquisition for peptide+HexNAc ion is directly following the MS/MS scan to collate identification and characterization of the glycopeptides in one experimental scan cycle. Performing the experiment in the LTQ-Orbitrap enables subsequent high resolution / high mass accuracy full scan in-source CID mass spectra. As the accurate mass pair of neutral loss is pre-determined (error within 5 ppm), the enormous number of MS/MS scans on non-paired species can be greatly reduced. Our method is compatible with nano-LC for separation of complex peptide mixtures without any further enrichment, resulting in a highly selective and sensitive approach to identify phosphopeptides and glycopeptides from a purified protein digestion, and could be potentially be of use in more protein mixture. The consequent spectra provide peptide sequence identification and modified site assignment as well as information of the glycan structure.

4.2 Introduction

Post-translational modifications (PTMs) of proteins modulate the activity of the proteins, protein-protein and protein-ligand interaction within cells. Analysis of PTMs presents immense analytical challenges, but their identification has led to essential knowledge and understanding of biological function. Because of its sensitivity and selectively, mass spectrometry (MS) is currently the most versatile technology to determine PTMs directly. However, the identification of PTMs still remains a substantial challenge owing to the low abundance of modified peptides and thus the relatively large amounts of non-modified peptides in samples; the excess of non-modified peptides results in suppression of the ion signal of the modified peptides in MS [58]. Protein phosphorylation and glycosylation is of particular interest since it is the most

common protein modification shown in Figure 4.1.

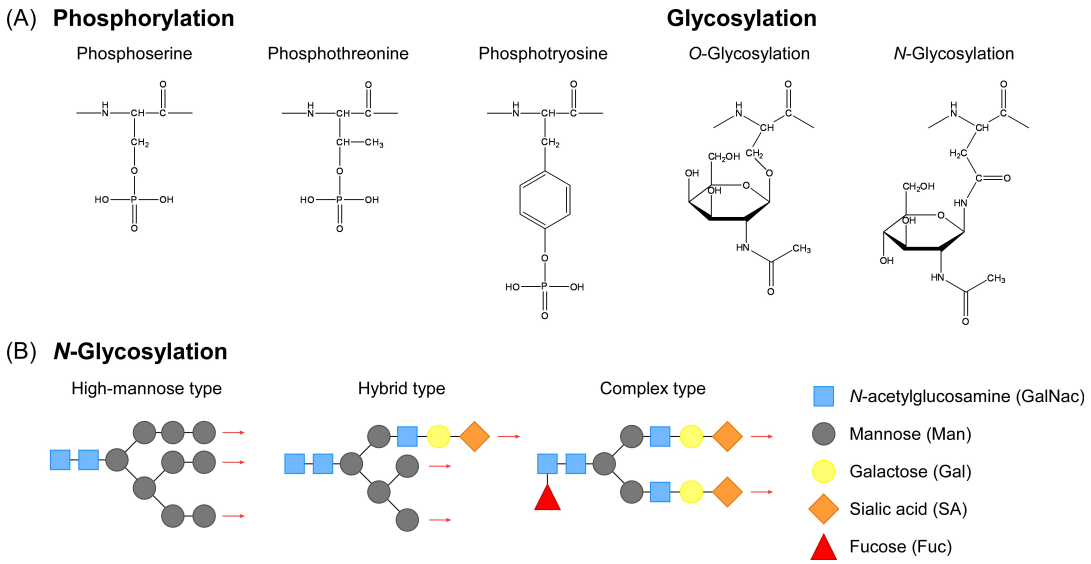


Figure 4.1. The common covalent post-translational modifications - phosphorylation and glycosylation. (A) Phosphorylation, the phosphate group covalently links to serine, threonine or tyrosine residue. *O*-glycosylation, the *N*-acetylgalactosamine covalent links to serine or threonine residue. *N*-glycosylation, the *N*-acetylgalactosamine covalent links to asparagine residue. (B) There are three major classes of *N*-glycosylation protein in mammalian cell, high-mannose type, complex type and hybrid type.

Reversible phosphorylation of proteins is an important regulatory mechanism in a wide range of cellular processes (see chapter 2 and 3). Glycosylation plays a crucial role not only in protein structure but also in cell-cell recognition, and perhaps even signaling mechanisms [141]. Hence, the analysis of protein glycosylation is essential if its biological roles are to be understood and evaluated. To completely describe protein glycosylation, three sections of information are required: (1) identification of the glycosylated proteins and peptides, (2) the position of the glycosylation, and (3) the structure of the glycan. *O*-glycosylation occurs at the side chain of serine or threonine residues during the later stage of protein processing, probably in the Golgi apparatus [142, 143]. *N*-glycosylation generally occurs at the side chain of asparagine where Asn is a part of the triplet Asn-X-Ser/Thr, where X can be any amino acid residue except proline [144, 145]. The process of oligosaccharide attached to asparagine of protein migrates through the endoplasmic reticulum (ER) and the Golgi apparatus, leading to the formation of high-mannose, complex, or hybrid type glycoprotein in mammalian cells, The high-mannose type contains two *N*-acetylglucosamine and many of mannose residues, The complex type contains two *N*-acetylglucosamines, three mannoses, and additional *N*-acetylglucosamine, galactose, sialic acid, and fucose residues. The hybrid type has both high-mannose and complex type characteristics [146, 147]. The glycosylation modification is

highly heterogeneous, making them challenging for characterization in MS analysis. The most common approach for identification of phosphorylation and glycosylation is the separation technology such as affinity-based enrichment of phosphopeptides [23] or glycopeptides [24] before MS analysis. An alternative isolation strategy, demonstrated as being highly successful, is designated to the precursor ion and the neutral loss scanning techniques utilising electrospray ionization (ESI) in a tandem mass spectrometer (MS/MS) shown in Figure 4.2.

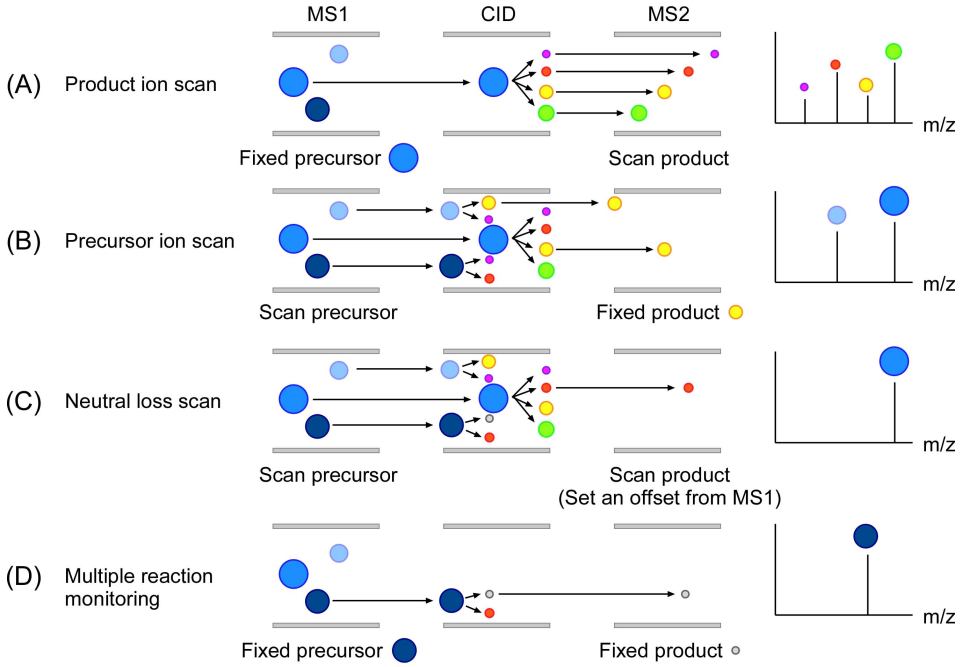


Figure 4.2. The scheme of various types of tandem mass spectrometry (MS/MS) experiments. (A) Product ion scan is the most common MS/MS experiment in proteomic research. The first MS selects one specific precursor ion. The selected ion undergoes collision-induced dissociation (CID) in the collision cell, and the resulting fragment ions are analyzed by the second MS. The spectra can be used for identification of peptide sequences. (B) Precursor ion scan is used to analyze a subset of peptides containing a specific function group such as a phosphate or a saccharide residue. The first MS is making a regular full scan. All ions that pass through the first MS are fragmented in the collision cell. The MS2 is fixed, allowing to transmit only one specific fragment ion. (C) Neutral loss scan, all precursor ions that undergo the neutral loss of a specific functional group are monitored. The neutral fragment is lost in the collision cell. The first MS scans all masses and the second MS also scans, but at a set offset from the first MS. This offset corresponds to a neutral loss that is commonly observed for the class of compounds, such as a neutral loss of phosphoric acid via beta-elimination in CID for identification of protein phosphorylation at serine and threonine residue. (D) Multiple reaction monitoring (MRM) is used for quantification of targeted analytes with the known fragmentation properties. The first MS allows multiple masses through and the second MS monitors for specific user defined fragment ions. Use of MRM increases selectivity, improves S/N ratio.

Modified peptides can be recognised from a common diagnostic fragment ion (m/z at 216 for phosphotyrosine immonium ion) [148, 149] or neutral loss mass (loss of 98 Da in the case of phosphoserine or phosphothreonine) [150, 151], as well as characteristic neutral losses of carbohydrates on glycosylated peptides (loss of 162 Da or 203 Da in the case of Hex or HexNAc, respectively) [152]. In the case of a neutral-loss experiment, the m/z ratios of ion pairs are selectively determined that exhibits the mass difference of the neutral component (e.g. H_3PO_4 or Deoxyhexose (DeoxyHex), Hexose (Hex), *N*-acetylhexosamine (HexNAc), *N*-acetylneuraminic acid (NeuAc) or *N*-glycolylneuraminic acid (NeuGc)). The most common MS set-up for recording precursor-ion and neutral-loss scans is a triple quadrupole MS, in which the single quadrupoles can selectively and rapidly scan for m/z ion pairs. However, “real” neutral-loss scan experiments coupled with on-line LC on quadrupole time-of-flight (Q-TOF), ion-trap or Fourier-transform ion cyclotron resonance (FT-ICR) MS instruments are not possible, because the experiments require two continuous scanning mass analyzers with an offset of neutral loss. In this study, we developed a novel scan mode in the LTQ-Orbitrap. This we call a pseudo-neutral-loss scan, as it mimics the neutral-loss scan in triple-quadrupole MS. It shows potential value in the identification of phosphorylated and *N*-glycosylated sites on proteins.

4.3 Experiment Sections

4.3.1 Materials

Ammonium bicarbonate, beta-casien, Chloroacetamide (CAA) and ribonuclease B (RNase B) were obtained from Sigma-Aldrich (St. Louis, MO). Sequencing grade, modified trypsin was obtained from Promega (Madison, WI). RapiGest™ SF was obtained from Waters Corporation (Manchester, UK). Dithiothreitol (DTT), formic acid and acetonitrile (ACN) were obtained from Merck (Darmstadt, Germany).

4.3.2 In-Solution Digestion

The standard protein was dissolved in 10 μl 1 % RapiGest™ SF. The protein solution was reduced with 10 μl of 10 mM DTT at 37 ℃ for 1 hour, alkylated with 10 μl of 20 mM CAA at 37 ℃ for 1 hour, diluted to 0.1 % RapiGest™ SF with 70 μl of 25 mM ammonium bicarbonate, and subsequently digested with either trypsin at 37 ℃, overnight (the enzyme to substrate ratio is 1:20). The resulting peptides were acidified with 50 μl of 5 % TFA at 37 ℃ for 2 hours and centrifuged at 17000 g for 15 min. Subsequently, the supernatant was transferred to another eppendorf and dried down by SpeedVac for further MS analysis.

4.3.3 Mass Spectrometry

MS analysis was performed with a LTQ-Orbitrap mass spectrometer (Thermo Fisher Scientific),

operated under Xcalibur Software 2.0.7, equipped with a nanoelectrospray ion source and coupled to an Agilent 1100 HPLC system (Agilent Technologies), fitted with a C18 column made in-house. Tryptic peptides were first loaded at a flow rate of 10 µL/min onto a C18 trap column (1.5 cm, 360 µm o.d., 150 µm i.d., Reprosil-Pur 120 Å, 5 µm, C18-AQ, Dr. Maisch GmbH, Germany). Retained peptides were eluted and separated on an analytical C18 capillary column (15 cm, 360 µm o.d., 75 µm i.d., Reprosil-Pur 120 Å, 5µm, C18-AQ, Dr. Maisch GmbH, Germany) at a flow rate of 300 nL/min, with a gradient from 7.5 to 37.5 % ACN in 0.1 % formic acid for 60 min. Typical MS conditions were: spray voltage, 1.7 kV; heated capillary temperature, 150 °C; normalised collision-induced dissociation (CID) collision energy 37.5 % for MS/MS in LTQ. An activation q = 0.25 and activation time of 30 ms were used. The Xcalibur Software 2.0.7 implements all the functions we performed in this study, called “source fragmentation”, “use m/z values as masses”, “mass tags” and “multistage activation”. The function of source fragmentation allows to turn on and off ion source collision-induced dissociation (CID) of a specified scan event. The function of use m/z values as masses determines the charge state of the ionized species from the full MS spectrum and converts the mass-to-charge ratios into masses. The function of mass tags allows a pair of mass tagged ions is to trigger a MS/MS scan. The function of multistage activation is the process of collapsing one or more neutral loss MS3 experiments down to a single MS/MS experiment. For phosphopeptide analysis, the mass spectrometer was automatically switched between MS and MS/MS acquisition when the neutral loss of phosphoric acid was detected. Survey full scan MS spectra (from m/z 350–1600) were acquired in the Orbitrap with resolution R=30,000 at m/z 400 (after accumulation to a ‘target value’ of 1,000,000 in the Orbitrap), with an in-source CID energy of 70 V, and the option of m/z value as masses was enabled. The option of mass tags was enabled and the value was set at 97.9769. The mass-pair ions were isolated sequentially and fragmented in the linear ion trap by using multi-stage activation CID MS/MS at a target value of 100,000 [132]. For glycopeptide analysis, the mass spectrometer was automatically switched between MS and MS/MS acquisition when the neutral loss of monosaccharide residues was detected. Survey full-scan MS spectra (m/z from 150 to 1600) were acquired in the Orbitrap with resolution R=30,000 at m/z 400 (after accumulation to a ‘target value’ of 1,000,000 in the Orbitrap), and an in-source CID energy of 30 V. The mass range for selecting MS-data-dependent masses was set 1500–1,000,000, and the option of m/z value as masses was enabled. The option of mass tags was enabled and the value was set at 146.0579, 162.0528, 203.0794, 291.0954, 307.09033 for DeoxyHex, Hex, HexNAc, NeuAc and NeuGc, respectively. The mass-pair ions were isolated sequentially and fragmented in the linear ion trap using CID MS/MS at a target value of 100,000. The theoretical m/z of peptide+HexNAc was entered in the product mass list for triggering multi-stage activation CID MS3, limiting the selection of ions for data-dependent acquisition to the top 20 ion intensities.

4.4 Result and Discussion

4.4.1 The Principle of Pseudo-Neutral-Loss

Our method is based on the presence of an ion pair of known characteristic mass in the mass spectrum that is generated during in-source CID fragmentation. If the signals of the mass pair exhibited in the MS spectrum correspond to the representative neutral loss (phosphoric acid or monosaccharide residues, see below), these ions can be selected for MS/MS shown in Figure 4.3.

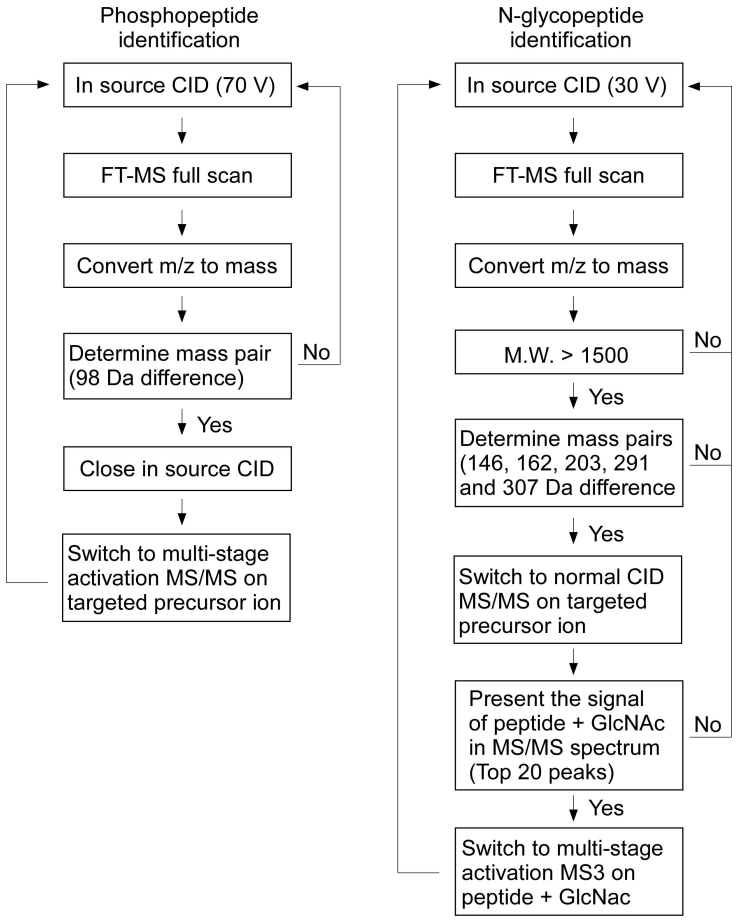


Figure 4.3. Workflow of pseudo-neutral-loss scan in LTQ-Orbitrap for the detection of phosphorylated peptides and *N*-glycosylated peptides.

For phosphopeptide identification on peptides that elute from the LC into the orifice of the Orbitrap MS, an additional potential difference of 70 V in the ion transfer region is applied for in-source CID fragmentation; this removes phosphoric acid on phosphoserine and phosphothreonine residues by charge directed of neutral loss [153, 154]. The LTQ-Orbitrap instrument determines the charge state of the ionized species from the full MS spectrum and converts the mass-to-charge ratios into masses. If a mass pair corresponding to the loss of

neutral phosphoric acid (98 Da) is present in the full scan mass spectrum and its intensity is above the user-defined data-dependent threshold, the collision energy in the ion source is turned off, and multi-stage activation MS/MS is triggered subsequently for the corresponding selected mass. Importantly, the high mass accuracy of the LTQ-Orbitrap mass spectrometer allows the neutral loss of phosphoric acid to be specified within an error of 5 ppm. Operating under this strict criterion dramatically reduces the number of false positives as encountered earlier in triple- quadrupole instruments.

For the identification of *N*-glycopeptides, an additional potential difference in the ion transfer region is set to 30 V, which is sufficient to produce marker ions at m/z = 204 ($HexNAc^+$), 292 ($NeuAc^+$), and 366 ($Hex\text{-}HexNAc^+$) [155]. Monitored mass-to-charge ratios are then converted into masses. The setting of precursor mass selection for MS/MS sequencing is applied only to deconvoluted masses that exceed 1500 Da, which is the minimum molecular weight of mammalian tryptic *N*-glycopeptides. Under reverse-phase (RP) LC conditions, chromatograms of heterogeneous glycopeptides harbouring the same peptide sequence typically show multiple co-eluting glycosylated forms in different mass and/or charge states [156]. This is because the oligosaccharide moieties are hydrophilic and the peptide - depending on its sequence - is hydrophobic, and hydrophobicity of the peptides can be regarded as main reason for separation under RP conditions. Therefore, the appearance of mass pairs in the MS full-scan spectrum corresponds either to the co-elution of glycopeptides that contain different carbohydrate moieties or to the neutral loss of monosaccharide residues by in-source CID (DeoxyHex, 146 Da; Hex, 162 Da; HexNAc, 203 Da; NeuAc, 291 Da; NeuGc, 307 Da). Accordingly, all species that reveal the given mass difference in the deconvoluted full-scan mass spectrum are selected for CID MS/MS. However, the MS-based analysis of glycopeptides using CID fragmentation is difficult, as it typically reveals predominantly information on the composition and sequence of the glycan moiety, without revealing any information about the peptide sequence. Therefore, an additional MS3 scan is required to obtain information about the peptide. To select for fragments within the MS/MS that are suitable for MS3, we made use of the fact that *N*-glycosylated peptides with the known consensus motif Asn-X-Ser/Thr, (where X can be any amino acid except Pro) always carry HexNAc as the first sugar at Asn-residue. This allows one to calculate the theoretical mass of a [peptide-HexNAc] in order to then trigger the multi-stage activation MS3 and finally to obtain a peptide fragmentation pattern.

4.4.2 Identification of Phosphopeptide by Pseudo-Neutral-Loss Scan

To demonstrate our working scheme for the various neutral losses, we first carried out a preliminary pseudo-neutral-loss scan experiment for the identification of phosphorylation sites by using 10 fmol tryptic peptides on column derived from tryptic digestion of beta-casein.

Beta-casein is frequently used as a model phosphoprotein in the development of MS methods for analysis of phosphorylation sites [71, 75, 133]. At first, we optimised the in-source CID energy in order to generate fragment ions showing neutral loss of phosphoric acid (H_3PO_4). We examined the in-source CID energy from 0 V to 80 V as shown in Figure 4.4.

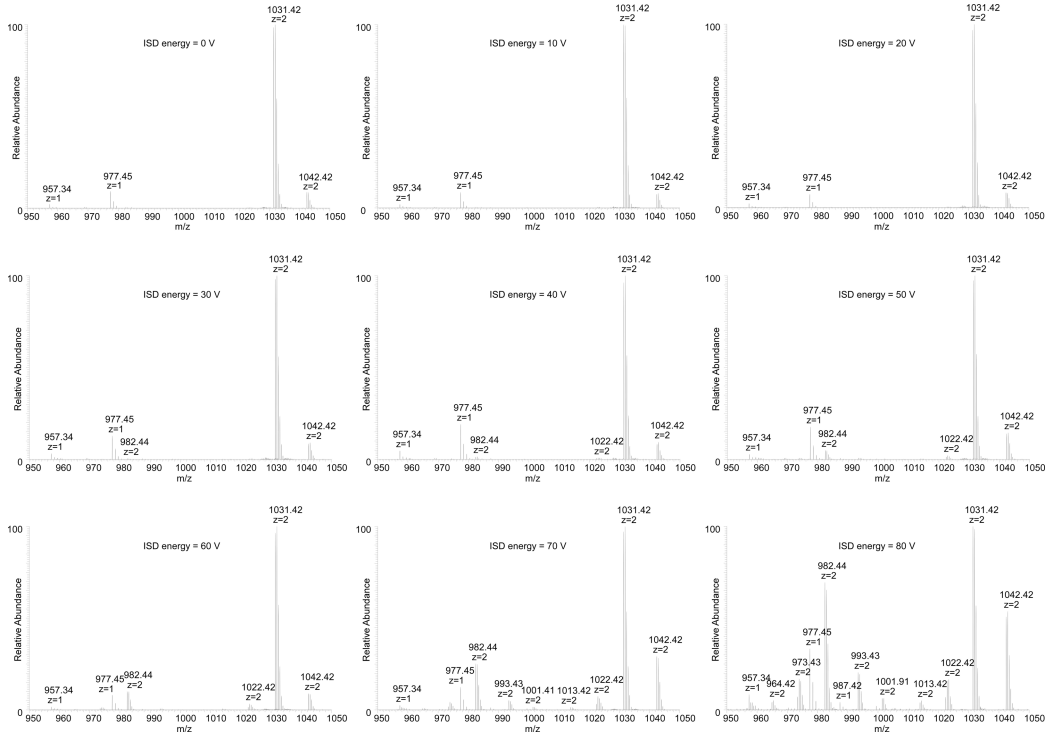


Figure 4.4. Optimisation of in-source CID energy for phosphopeptide detection with pseudo-neutral-loss scan in LTQ-Orbitrap. Mass spectra were acquired with in-source CID energies from 0 V to 80 V. The signal at *m/z* 982.44 (2+) corresponds to the neutral loss of phosphoric acid on the phosphopeptide ^{48}FQpSEEQQQTEDELQDK^{63} of beta-casein at *m/z* 1031.42 (2+).

Below 20 V, the signal at m/z 982.44 [^{48}FQpSEEQQQTEDELQDK^{63}–H_3PO_4], corresponding to neutral loss of phosphoric acid on the phosphopeptide of beta-casein at m/z 1031.42 [^{48}FQpSEEQQQTEDELQDK^{63}] was not present in the spectra. When the energy was raised to 30 V, a weak signal of the corresponding neutral loss ion started to be detectable. Note that the relative intensity ratio of neutral-loss ion to precursor ion was in direct proportion to the in-source CID energy. The maximum signal of neutral-loss ion was observed at 80 V; however, many peptide backbone cleavages were already generated under these conditions, resulting in the significant decrease in the signal intensity of phosphopeptide as compared with 70 V, as shown in Figure 4.5. The signal of the phosphopeptide dropped approximately 15 % and 70 % at 70 V and 80 V, respectively. We therefore set the in-source CID energy to 70 V for further analysis.

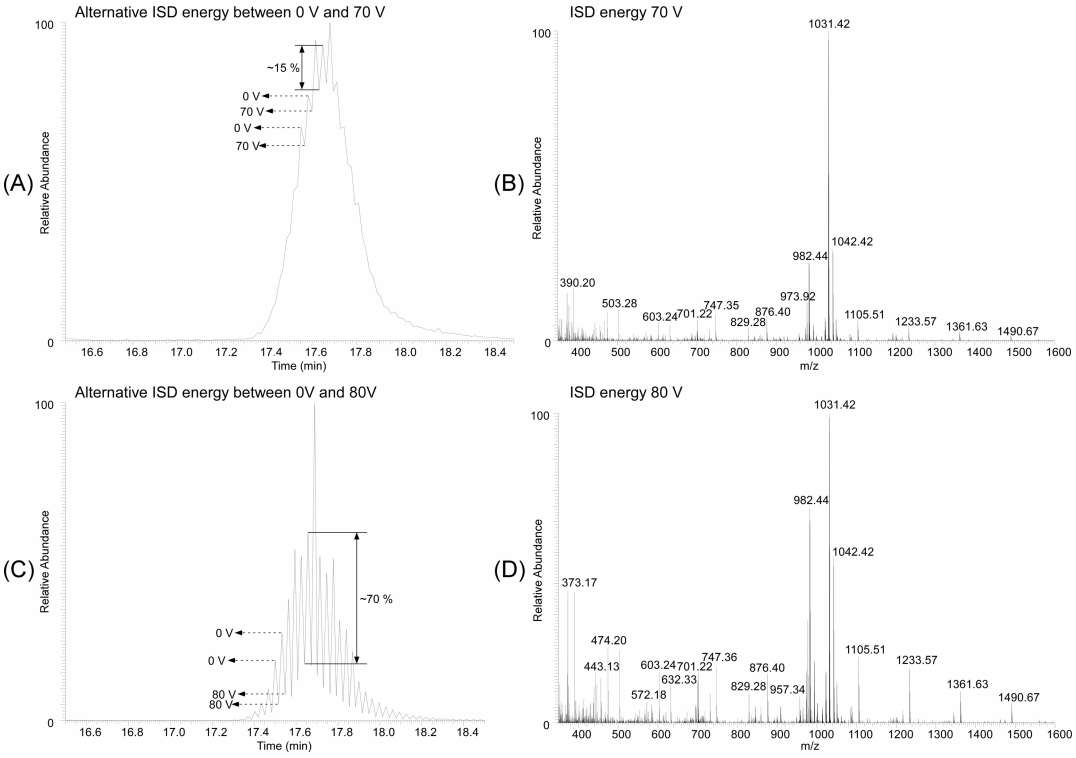


Figure 4.5. (A) Extracted ion chromatography of the phosphopeptide ^{48}FQpSEEQQQTEDELQDK^{63} of beta-casein at *m/z* 1031.42 acquired with an alternative in-source CID energy between 0 V and 70 V. (B) Mass spectrum under in-source CID energy 70 V. (C) and (D) An alternative in-source CID energy was set to 0 V and 80 V instead of 0 V and 70 V. The signal of phosphopeptide dropped by approximately 15 % and 70 % to below 70 V and 80 V, respectively, owing to peptide backbone cleavage.

As illustrated in the workflow (Figure 4.3), the appearance of the neutral loss of phosphoric acid in the case of precursors containing phosphoserine and phosphothreonine under high in-source CID energy caused the instrument to switch automatically into the product-ion mode (MS/MS) to acquire peptide-sequence information. This step involved measuring the masses of the precursor ions, determining their charge states, identifying the neutral loss of phosphoric acid and generating a list of precursor masses for MS/MS sequencing. Figure 4.6 shows the workflow for the identification of the phosphopeptide ^{48}FQpSEEQQQTEDELQDK^{63} of beta-casein using nanoLC coupled with a pseudo-neutral-loss scan in the LTQ-Orbitrap. When an in-source CID energy was employed at 70 V, a neutral loss of phosphoric acid (48.99) at m/z 982.43 derived from a doubly charged phosphopeptide precursor at m/z 1031.42 was observed (Figure 4.6A). The mass-to-charge ratios were converted into masses (Figure 4.6B). The presence of a mass pair equivalent to a neutral loss of phosphoric acid (97.98 Da) was selected for CID MS/MS to obtain peptide sequence information. To generate fragment ions of the phosphopeptides that were structurally more informative, “multi-stage activation” [132] corresponding to a neutral loss of phosphoric acid from doubly and triply charged precursor

ions was enabled in all MS/MS events and identified the phosphorylation as being on serine 50 (Figure 4.6C). The pseudo-neutral-loss scan of phosphoric acid greatly reduced the large number of false positive candidate precursor ions based on the highly accurate mass pair (Figure 4.6D), as compared with regular data-dependent acquisition (DDA) (Figure 4.6E).

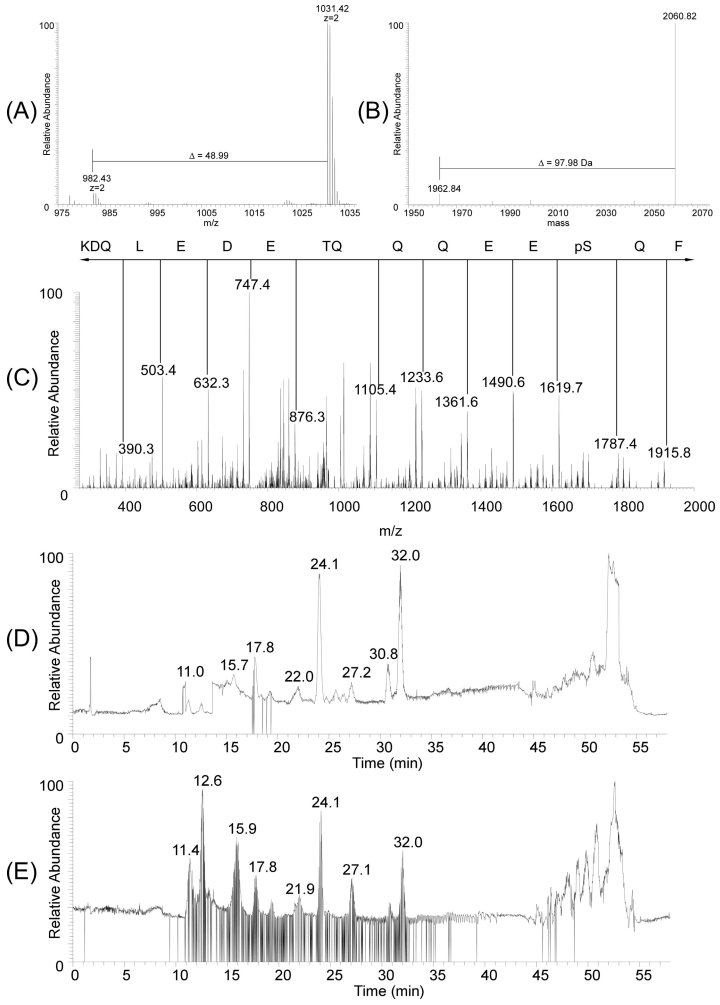


Figure 4.6. Identification of tryptic phosphopeptide ^{48}FQ*p*SEEQQQTEDELQDK^{63} from beta-casein using a pseudo-neutral-loss scan. (A) A full mass spectrum obtained from the LTQ-Orbitrap mass spectrometer at an in-source CID energy of 70 V, the difference in *m/z* value of 48.99 corresponding to the neutral loss of doubly charged phosphoric acid. (B) A deconvoluted and deisotoped mass spectrum based on direct assignment of charge to the measured signal from Figure 4.6A. (C) Pseudo-neutral-loss scan of phosphoric acid (97.98 Da) triggered multi-stage activation MS/MS acquisition for phosphopeptide ^{48}FQ*p*SEEQQQTEDELQDK^{63} at *m/z* 1031.42. (D) The total ion chromatogram obtained from pseudo neutral loss scan, in which MS/MS was triggered only the mass pair with pre-defined neutral loss of phosphoric acid identified in the MS scan. The phosphopeptide ^{48}FQ*p*SEEQQQTEDELQDK^{63} was eluted and was identified at retention time 17.51 min. (E) The total ion chromatogram generated from regular data-dependent acquisition.

We also examined the power of our method by analyzing tryptic peptides derived from a mixture of 10 fmol beta-casein and 50 fmol BSA (Figure 4.7 and Appendix 4). Despite of the fact that the phosphopeptide ^{48}FQpSEEQQQTEDELQDK^{63} at m/z 1031.42 was present in the regular DDA (Supplementary Figure 4.7C), however, the signal was not intensive to be triggered MS/MS. Compared with regular DDA, pseudo-neutral-loss acquisition could identify the phosphopeptide ^{48}FQpSEEQQQTEDELQDK^{63} by using MASCOT database search, and the result indicated large amount of non-phosphorylated peptides derived from BSA and beta-casein were removed (Appendix 4).

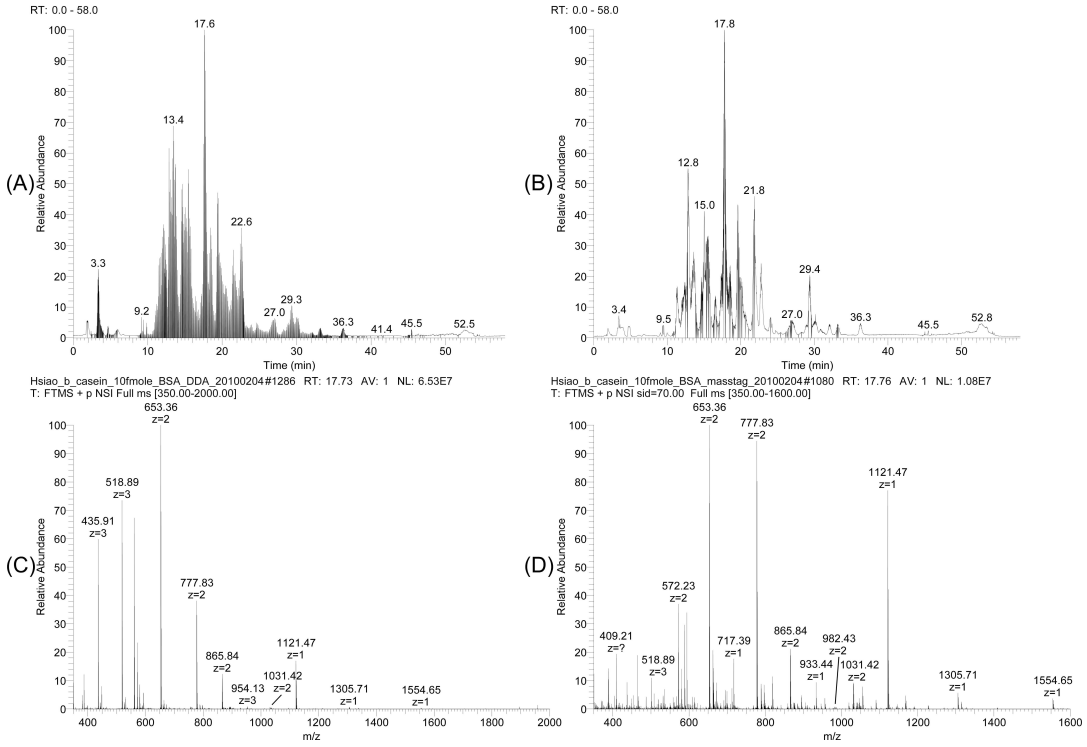


Figure 4.7. Identification of tryptic phosphopeptide ^{48}FQpSEEQQQTEDELQDK^{63} at m/z 1031.42 derived from 10 fmol beta-casein containing 50 fmol BSA tryptic peptides. (A) Regular data-dependent acquisition. (B) Pseudo-neutral-loss acquisition. (C) A full scan mass spectrum obtained from (A) at retention time 17.73 min. (D) A full scan mass spectrum obtained from (B) at retention time 17.76 min. The mascot searching result showed the phosphopeptide ^{48}FQpSEEQQQTEDELQDK^{63} only could be identified by using pseudo-neutral-loss acquisition and large amount of non-phosphorylated peptides derived from BSA and beta-casein were removed (Appendix 4).

4.4.3 Identification of *N*-Glycopeptide by Pseudo-Neutral-Loss Scan

The performance of pseudo-neutral-loss scans for the detection of glycopeptides is particularly attractive because mass pairs of monosaccharide residues are readily obtained during the

in-source CID fragmentation. However, since in-source CID causes the neutral loss of different monosaccharides attached to a peptide, the full MS spectrum will necessarily reveal several mass pairs. On the other hand, front-end separation of glycopeptides under reversed-phase conditions cannot separate peptide populations that carry a heterogeneous carbohydrate moiety; consequently, the mass pairs observed after in-source CID are due either to the stepwise neutral loss of a single modified peptide or to the presence of a naturally occurring heterogeneous peptide population eluting into the instrument. Despite this ambiguity under our liquid-chromatography conditions, it is nonetheless worthwhile to investigate the potential of pseudo-neutral-loss scanning of the monosaccharide residues following CID MS/MS in the ion trap combined with further multi-stage activation MS3 experiments as it makes our in-source CID essentially universal for glycopeptide identification and glycan structure profiling.

Figure 4.8 shows the results of a glycopeptide analysis from a 25 fmol tryptic digestion of ribonuclease B (RNase B) loaded onto a nanoLC ESI-Orbitrap with an in-source CID energy of 30 V. In Figure 4.8A, the mass spectrum shows a peptide that contains a high-mannose type *N*-linked oligosaccharides (HexNAc2Man4 to HexNAc2Man9), and thus further demonstrates that different glycan structures on the same peptide backbone either co-eluted from the RP column or were generated by in-source CID (see above). Consequently, the presence of a mass pair due to the loss of mannose (162 and/or 203 Da) switched the LTQ into the MS/MS mode to record its product ion spectrum (Figure 4.8B). When the ion with m/z 927.4 was taken as a precursor for CID MS/MS, the monosaccharide ladder served to outline oligosaccharide structures (HexNAc2Man6) with an interval of 162 or 203 for singly charged ions or 81 and 101.5 for doubly charged ions. The ions for the peptide moiety (m/z 475.4) and the [peptide-HexNAc] conjugate (m/z 658.4) were also present in the spectrum. Note that the collision-induced dissociation in LTQ often led to preferential cleavage of the glycosidic bonds rather than of the polypeptide bonds and, as a result, it usually provides information primarily about the glycan structure and not about the peptide sequence. Subsequently, a targeted MS3 of the [peptide-HexNAc] ion at m/z 678.4 was carried out (Figure 4.8C). The b2 and b3 ions at m/z 431.4 and 532.3 both retained the *N*-linked GlcNAc, and in combination with other product ions revealed the amino-acid sequence NLTK, where N was the glycosylation site of RNase B. The neutral loss of HexNac at m/z 475.4 was the dominant signal of the [peptide-HexNAc] MS/MS spectrum (Figure 4.8C). To improve the peptide sequence coverage, a multi-stage activation MS3 corresponding to a neutral loss of HexNAc was employed to fragment the [peptide-HexNAc] (Figure 4.8D). Instead of regular CID MS3, we observed clearly an increase in the abundance of nearly every b/y-type ion, along with y2, y3, b2 and b3 ions at m/z 248.2, 361.3, 228.1 and 329.2, respectively. Of note, as the MS3 fragment spectrum was used to search against protein database using MASCOT as search engine, the neutral loss of HexNAc in all the MS3 fragment ions had to be defined as a variable modification so that the confidence in the peptide identification and the MASCOT score is increased.

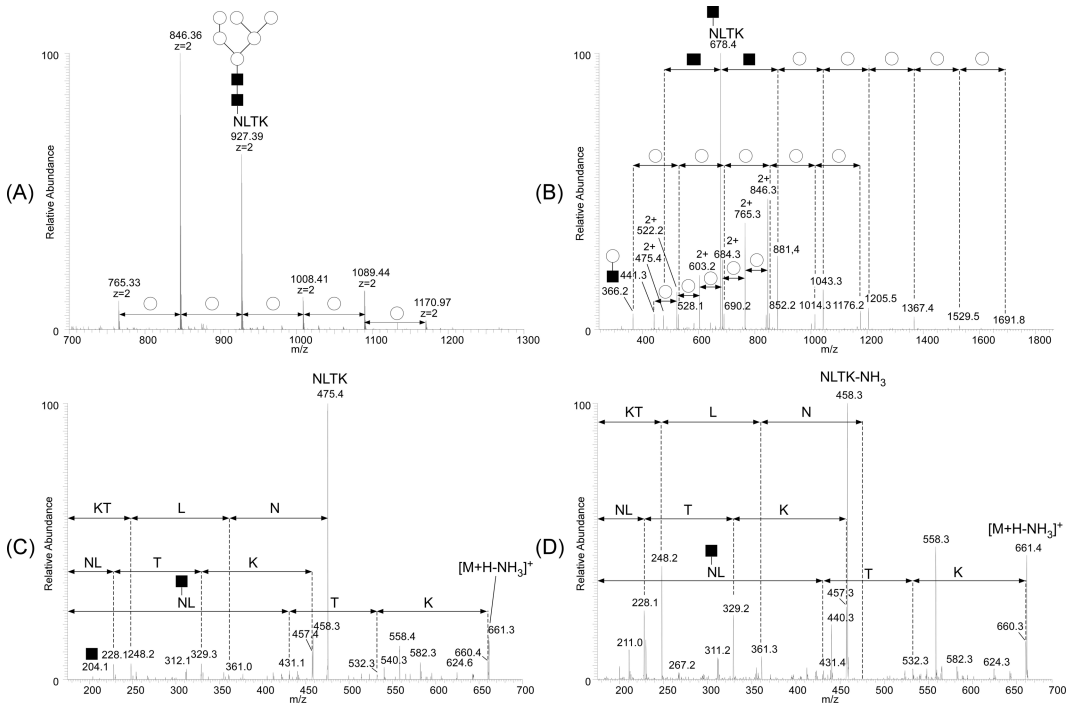


Figure 4.8. Online-nanoLC ESI-Orbitrap coupling with pseudo-neutral-loss scan to identify the tryptic glycopeptide from RNase B. (A) Mass spectrum of tryptic glycosylation peptide ^{34}NLTK^{37} attached to asparagine-34 containing a high-mannose type oligosaccharide. (B) MS/MS spectrum of the tryptic glycopeptide at *m/z* 927.4 triggered by a pseudo-neutral-loss scan for monosaccharide residues. (C) The fragment ion at *m/z* 678.4 (peptide + HexNac) in an MS/MS spectrum was subjected to a second ion isolation/fragmentation cycle to obtain peptide sequence information (targeted normal CID MS3 experiment). (D) Multi-stage activation MS3 corresponding to a neutral loss of HexNAc applied to fragment peptide+HexNAc instead of regular CID MS3. Square, *N*-acetylglucosamine (HexNAc) and circle, mannose (Man).

4.5 Conclusion

We have developed a novel scan function in LTQ-Orbitrap that is highly suitable for the specific detection of phosphorylation at serine and threonine residues and of glycosylation at asparagine residues, and for obtaining sequence information about the peptide moiety. This approach relies on the highly accurate mass spectrometry used here to detect the corresponding mass pairs in the mass scan. The appearance of mass pairs triggers MS/MS and multi-stage activation MS3 to obtain modified peptide sequence information. Nevertheless, there are certain limitations, as in the case of phosphotyrosine the neutral loss of the H_3PO_4 moiety is not commonly observed. The list of targeted MS3 for glycopeptide analysis is limited to 10 precursors by Xcalibur software restricting large scale glycoproteomic anlalysis. However,

we anticipate that this issue could be solved by the ThermoFisherScientific company. We expect this method to be extensively applied for moderately labile PTMs in future, not only for phosphorylation and glycosylation but also for sulphonation, nitrosylation, nitration etc., taking the advantage of CID or even electron-transfer dissociation (ETD) to identify peptide sequence and locate the modification site.

Chapter 5 - “ChopNSpice”, a Mass Spectrometric Approach That Allows Identification of Endogenous Small Ubiquitin-like Modifier-conjugated Peptides

5.1 Summary

Post-translational modification by conjugation of small ubiquitin-like modifier (SUMO) to its substrate is emerging as an important protein regulatory mechanism involved in diverse range of cellular processes from yeast to human. Typically, SUMO is covalently conjugated to lysine residues within a SUMO consensus motif (Ψ-K-x-D/E, where Ψ is a hydrophobic amino acid residue L, I, V, or F, x is any amino acid residue, and D or E is an acidic residue); however, an increasing number of SUMOylated proteins are on non-consensus sites. To appreciate the functional consequences of SUMOylation, the identification of SUMO attachment sites is of critical importance. Discovery of SUMO acceptor sites is usually performed by a laborious mutagenesis approach or using MS. In MS, identification of SUMO acceptor sites in higher eukaryotes is hampered by the large tryptic fragments of SUMO1 and SUMO2/3. Current mass spectrometric (MS)-based protein modification search engines in combination with known databases lack the possibility to search MS/MS spectra for larger modifications, such as SUMOylation. We therefore developed a novel, user-friendly and straightforward database search tool called “ChopNSpice” (http://chopnspice.gwdg.de/). In combination with current proteomic search engines like MASCOT or Sequest, ChopNSpice successfully allows unambiguous identification of mammalian SUMO acceptor sites from proteins SUMOylated in vivo and in vitro. To increase the sensitivity for the experimental detection of SUMOylated peptides, we used high mass MS/MS acquisition conditions in an Orbitrap mass spectrometer. Under these conditions only peptides with masses exceeding 2154 Da (for SUMO 1) or 3568 Da (for SUMO 2/3) are selected. Our approach is highly suitable for the accurate detection and sequencing of the large SUMOylated peptides and additionally facilitates their detection although they are low abundance. By applying this approach we demonstrated the power of this technique by the identification of the SUMO acceptor sites in, among others, endogenous SUMO1, SUMO2, RanBP2, and Ubc9.

5.2 Introduction

5.2.1 SUMOylation

Post-translational modification with ubiquitin and ubiquitin-like modifiers (Ubl) such as SUMO play an important role in most - if not all - cellular processes shown in Figure 5.1 [8, 157-161]. Conjugation of Ubls to their targets involves an isopeptide bond between the carboxyl group of the modifier and the ε-amino group of a lysine residue within the targets shown in Figure 5.1.

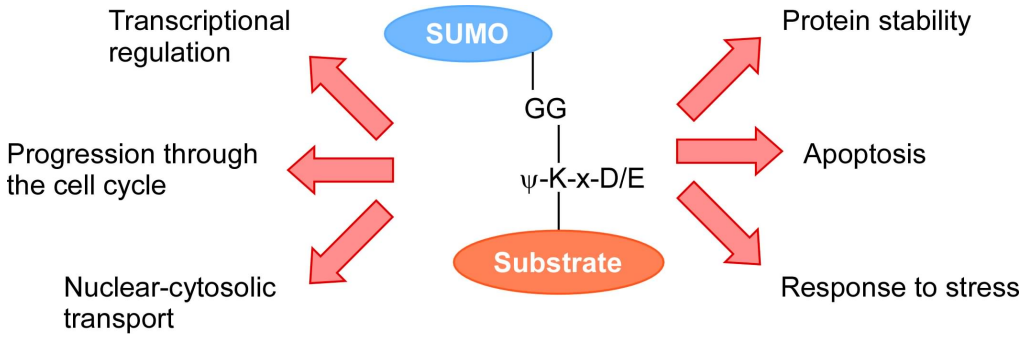


Figure 5.1. Small ubiquitin-like modifier (SUMO). SUMOylation is a covalent post-translational modification located within the consensus motif Ψ-K-x-D/E, where Ψ represents a large hydrophobic amino acid, and x represents any amino acid. The SUMOylation involves in various cellular processes such as transcriptional regulation, progression through the cell cycle, nuclear-cytosolic transport, protein stability, apoptosis and response to stress.

Attachment of SUMO to specific substrates involves four-step enzymatic pathway analogous to ubiquitylation shown in Figure 5.2 [162].

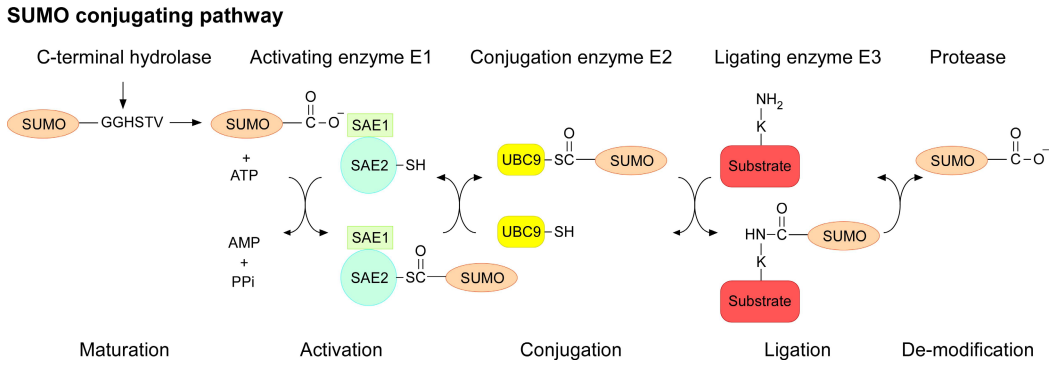


Figure 5.2. SUMO conjugating pathway. The SUMO precursor is processed by a SUMO specific protease to reveal the C-terminal di-glycine and then is activated by the E1 enzyme. After transesterification onto the E2 conjugating enzyme (Ubc9), the SUMO conjugated substrate is selected. The SUMO is ligated to the substrate with the help of an E3 ligating enzyme. SUMOylation is reversible and can be deconjugated from the targeted protein by the action of specific SUMO proteases.

First the SUMO is processed by protease, known as SENP for human and Ulp1 for yeast, to expose their C-terminal di-glycine. Subsequently, the SUMO is conjugated to E1 by ATP-dependent formation of a thioester bond between the C-terminal glycine of SUMO and the SUMO-specific E1 activating enzyme, known as a heterodimeric activating enzyme SAE1/SAE2 and Uba2p/Aos1p for human and yeast, respectively. The activated SUMO

protein is then transferred thioester bond from the E1 activating enzyme to the E2 conjugating enzyme, Ubc9. The Ubc9 is reported to recognize and bind directly to a consensus SUMOylation motif Ψ-K-x-D/E [163-166]. The E3 ligating enzyme, such as RanBP2 or members of the PIAS family [161, 167], in turn efficiently catalyzes the formation of an isopeptide bond between the C-terminal glycine of SUMO and the ε-amino group of lysine residue in the substrate proteins. The conjugation status undergoes perpetual change, and is governed by a small family of SUMO proteases that hydrolyze the isopeptide bond between SUMO and its target [168, 169]. Unlike yeast and other invertebrates, which is present only one SUMO gene, vertebrates express at least three different SUMO isomers: SUMO1, SUMO2 and SUMO3. Human SUMO2 and SUMO3 (referred to as SUMO2/3) share 86 % sequence identity, but SUMO1 exhibits 44 % sequence identify with SUMO2/3 [161]. Although all SUMO proteins share the conserved ubiquitin domain and utilize the same E1 and E2 enzymes for cleavage/attachment of the C-terminal diglycine of SUMO, the molecular basis for their substrates favorite and the additional recruitment of E3 ligases remain unclear [164, 170]. While the list of known SUMO substrates is growing rapidly, our understanding of the functional consequences for many of these targets is lagging behind. At a molecular level, the functional consequences of SUMO conjugation can be explained by a gain or loss of interaction with other macromolecules [157, 159]. The SUMO-dependent intramolecular conformational changes have also been described [171, 172]. Thus, to appreciate the role that SUMO plays in the regulation of specific substrates, identification of the acceptor site(s) for SUMO conjugation is of key importance. So far, identification of SUMO acceptor sites has relied largely on mutation of the consensus SUMOylation motif Ψ-K-x-D/E. This motif is recognized by Ubc9, if presented in an extended conformation [163-166]. However, an increasing number of protein substrates, such as PCNA, E2-25K, Daxx and USP25, turn out to be SUMOylated on lysine residues that do not consist with the SUMO consensus site [173-176]. For this category of proteins, as well as for proteins that contain a large number of SUMO consensus sites, the identification of acceptor lysines is a burdensome task that often involves mutagenesis of each lysine residue within the substrate in turn.

5.2.2 Challenges for Identification of SUMOylation by MS-based Approach

Mass spectrometry (MS) is currently one of the state-of-the-art technologies to identify protein factors and their post-translational modifications in an unbiased and sensitive manner. Several groups have shown that, using over-expressed tagged SUMO, MS can be efficiently exploited to identify endogenous substrates for SUMO conjugation [177-179]. However, the identification of SUMO-acceptor lysines using MS has remained a more challenging task [177, 180-182]. To date, using affinity tagged SUMO technology, unbiased identification of acceptor lysines for endogenous substrates has only been observed in S. cerevisiae [179]. The identification of substrates in higher eukaryotes has been hampered by the large conjugated SUMO peptide

that arises upon tryptic digestion (> 2154 Da with human SUMO1, >3568 Da with human SUMO2/3; compared with 484 Da for Smt3 in S. cerevisiae). Such large fragments, in addition to the mass of the conjugated peptide, can impede their in-gel digestion, extraction, detection and sequencing in MS. Consequently, MS and MS/MS result in fragment-ion spectra that are too complex to interpret manually. To circumvent these problems, several different strategies have been developed: (1) identification of targets using an in vitro to in vivo approach [180], (2) a mutational SUMO approach yielding a smaller tryptic fragment of SUMO that simplifies the identification of SUMO acceptor sites by MS [181], and (3) development of an automated recognition pattern tool (SUMmOn) [182]. Although these approaches have been applied successfully for the identification of SUMO conjugations in vitro and in vivo, unbiased identification of SUMO conjugations in vivo has not been achieved in higher eukaryotes. Another hurdle to such identification of SUMO conjugations is the variety of masses that can theoretically arise for just one SUMO-conjugated lysine in a given protein due to tryptic mis-cleavages. Thus, the unambiguous identification of SUMO acceptor sites requires the mass of the modified peptide carrying the conjugated SUMO (fragment) to be measured with high accuracy and - most importantly - it requires sequence analysis of the modified peptides. Since available proteomic search engines lack the possibility to search MS/MS spectra for larger modifications, e.g. those that occur upon SUMOylation, we developed a novel, simple and straightforward database search tool (ChopNSpice) that, in combination with current proteomic search engines (such as Sequest [49] or MASCOT [50]), allows one to identify SUMO1 and SUMO2/3 acceptor sites unasmbiguously. We confirmed this strategy in vitro on various substrates, and demonstrate the power of this technique by the identification of acceptor lysines within several endogenous targets from HeLa cells.

5.3 Experiment Sections

5.3.1 Materials

Ammonium bicarbonate, iodoacetamide (IAA) were obtained from Sigma-Aldrich (St. Louis, MO). Sequencing grade, modified trypsin was obtained from Promega (Madison, WI). Dithiothreitol (DTT), formic acid and acetonitrile (ACN) were obtained from Merck (Darmstadt, Germany). Mouse monoclonal α-SUMO1 antibodies were kindly provided by M. Matunis and goat anti-SUMO1 antibodies were produced by Prof. Dr. Frauke Melchior's laboratory [183, 184]. Secondary antibodies were obtained from Jackson Laboratories. Plasmids for bacterial expression of Aos1/Uba2, Ubc9, SUMO1, SUMO2 (NM_006937), GST-SP100, RanGAP1, were produced by Prof. Dr. Frauke Melchior's laboratory [176, 185]. A plasmid for GST-p53 was kindly provided by Dr. Moshe Oren. Recombinant protein purification for SUMO1, SUMO2, Aos1/Uba2, Ubc9, RanGAP1, GST-SP100, GST-p53 and PIAS1 were produced by Prof. Dr. Frauke Melchior's laboratory [176, 185-187].

5.3.2 Software

ChopNSpice was written in PHP by Benedikt Frank. The software tool is freely available online (http://chopnspice.gwdg.de/), and also releases as open source under the terms of the General Public License v3 (GPLv3).

5.3.3 In vitro SUMOylation Assays

The in vitro SUMOylation assays were performed by Dr. Erik Meulmeester. SUMO conjugation reactions were performed at 30 ℃ for 1 hour in the presence or absence of 5 mM ATP in 20 µL transport buffer TB (20 mM Hepes/KOH pH 7.3, 110 mM potassium acetate, 2 mM magnesium acetate, 0.5 mM EGTA, 1 mM DTT supplemented with protease inhibitors). Reactions contained 100 ng Aos1/Uba2, 200 ng Ubc9, 2.5 µg SUMO1 or SUMO2 and 1 µg target protein (GST-p53, mouse RanGAP1, GST-SP100 or Aos1/Uba2) in a volume of 20 µL.

5.3.4 Cell culture, Immunoprecipitation and Immunoblotting

Cell culture, immunoprecipitation and immunoblotting experiments were performed by Dr. Erik Meulmeester. Hela-S3 cells were maintained in Jokliks medium supplemented with 10 % fetal bovine serum and antibiotics. To immunoprecipitate SUMO1 conjugates, $1x10^8$ HeLa cells were washed twice with PBS containing 10 mM NEM and lysed in two pellet volumes RIPA buffer (20 mM NaP_i (pH 7.4), 150 mM NaCl, 1 % Triton, 0.5 % sodium deoxycholate, 0.1 % SDS) supplemented with protease inhibitors and 10 mM NEM. Lysates were centrifuged (16.000g for 15 min. at 4 ℃) and filtered (0.45 µm) prior to addition of 25 µg monoclonal α-SUMO1 antibodies. After 2 hours incubation at 4 ℃, the lysates were centrifuged (16.000g for 15 min. at 4 ℃) and the supernatant was incubated for another 2 hours at 4 ℃ with protein G agarose. After collection and extensive washing of bound proteins, samples were eluted with 2x sample buffer and separated on SDS-PAGE followed by Coomassie staining or western blotting. In a second larger experiment using $1x10^9$ cells were lysed in TB (0,1 % triton, 10 mM ATP) and were treated with 10 mM NEM after lysis. Immuno-precipitation was similar as described above, using 100 µg GMP1 antibodies. The SUMO acceptor site in RanGAP1 was observed in both purification methods while the other targets were identified in the second up scaled experiment.

5.3.5 Protein Digestion

SUMO-conjugated proteins were excised from gel, reduced with 50 mM DTT for 1 hour, alkylated with 100 mM IAA for 1 hour and in-gel digested with modified trypsin (Promega) overnight, all at 37 ℃. SUMO-conjugated proteins from solution were reduced with 50 mM DTT for 1 hour, alkylated with 100 mM IAA for 1 hour and subsequently digested with modified trypsin overnight, all at 37 ℃.

5.3.6 Liquid Chromatography and Mass Spectrometry

The resulting tryptic peptides were dissolved in 2 µL 50 % ACN with 0.1 % formic acid and added to 18 µL 0.1 % formic acid for further MS analysis. MS analysis was performed by nanoscale liquid chromatography-tandem mass spectrometry (LC-MS/MS) using a LTQ-Orbitrap mass spectrometer (Thermo Fisher Scientific) equipped with a nanoelectrospray ion source and coupled to an Agilent 1100 HPLC system (Agilent Technologies), fitted with an self made C18 column. Tryptic peptides were first loaded at a flow rate of 10 µL/min onto a C18 trap column (1.5 cm, 360 µm o.d., 150 µm i.d., Reprosil-Pur 120 Å, 5 µm, C18-AQ, Dr. Maisch GmbH, Germany). Retained peptides were eluted and separated on an analytical C18 capillary column (15 cm, 360 µm o.d., 75 µm i.d., Reprosil-Pur 120 Å, 5 µm, C18-AQ, Dr. Maisch GmbH, Germany) at a flow rate of 300 nL/min, with a gradient from 7.5 to 37.5 % ACN in 0.1 % formic acid for 60 min. Typical MS conditions were: spray voltage, 1.8 kV; heated capillary temperature, 150 ℃; normalized collision-induced dissociation (CID) collision energy 37.5 % for MS/MS in LTQ. An activation q = 0.25 and activation time of 30 ms were used. The mass spectrometer was operated in the data dependent mode to automatically switch between MS and MS/MS acquisition. Survey full scan MS spectra (from m/z 350-2000) were acquired in the Orbitrap with resolution R=30,000 at m/z 400 (after accumulation to a ‘target value’ of 1,000,000 in the Orbitrap). The five most intense ions were isolated sequentially and fragmented in the linear ion trap using CID at a target value of 100,000. For all measurements with the Orbitrap detector a lock-mass ion from ambient air (m/z 445.120025) was used for internal calibration. For high-mass data-dependent mode, the mass range for selecting MS data dependent masses was 2154-1000000 and 3568-1000000 for SUMO1 and SUMO2/3, respectively, and used m/z value as masses.

5.3.7 Data Analysis

For protein identification, all MS/MS spectra were searched against a Swiss-Prot database using MASCOT with the following parameters: mass tolerance of 10 ppm in MS mode and 0.8 Da in MS/MS mode; allow up to two missed cleavage; variable modifications considered were methionine oxidation and cysteine carboxyamidomethylation. The sequence of the protein of interest was manually saved to a FASTA file and ChopNSpice was used to create a new FASTA file with the following parameters: spice species was H. sapiens; spice sequences were SUMO1 and SUMO2, respectively; spice site was KX; spice mode was once per fragment; include unmodified fragments in output; enzyme was trypsin K/R, don't cleave at P; allowed up to three protein miscleavages; allowed up to one miscleavage in the “spice sequence”; output formatting was FASTA: single protein sequence; marked all cleaved site (J); retain comments in FASTA format; without line breaks in FASTA output. For SUMOylated site identification with MASCOT or Sequest, all MS/MS spectra were searched against a new FASTA file which was

created by ChopNSpice with the following parameters: mass tolerance of 10 ppm in MS mode and 0.8 Da in MS/MS mode; allow zero missed cleavage; variable modifications considered were methionine oxidation and cysteine carboxyamidomethylation; enzyme was cleavage J at N and C terminus for MASCOT or no enzyme must be used for Sequest. If the search was performed with the in-house MASCOT server, the file “quant_subs.pl” must be changed from 'J' => 0 to 'J' => 0.05 in line 3653). All MS/MS spectra were confirmed manually to identify the SUMO acceptor site. The symbol of the amino acid which was before and after the identified SUMO-conjugated peptide must be J. All high-abundance peaks had to be assigned to y- or b-ion series).

5.4 Result and Discussion

5.4.1 ChopNspice

A typical strategy in MS-based proteomics comprises enzymatic digestion of proteins with endoproteinases, separation of the resulting peptides by liquid chromatography (LC), and ionization and subsequent fragmentation of the peptides. Afterward the automated search of the fragment spectra against a database allows identification of the corresponding protein [35]. In addition, the identification of post-translational modifications by MS requires a highly accurate mass determination of the precursor and a sequence information of the peptide containing the modification. Search engines such as MASCOT and/or SEQUEST are commonly used among the MS-based proteomic researcher, and the output format of these search engines are widely accepted in the community. We developed a software tool ”ChopNSpice” that makes use of these search engines, implements new modified protein sequences to the standard databases and allows to identity modified sites.

Accordingly, our approach to identify SUMO acceptor sites is based on the fragmentation pattern of the SUMO-conjugated peptides, which are digested with trypsin. Such digestion results in the peptide in which a missed (i.e., non-cleaved, owing to SUMO modification) lysine residue is branched with a SUMO tryptic peptide shown in Figure 5.3A. In practice, we and others observed that the spectrum of MS/MS fragmentation pattern of such a branched peptide is similar to a tryptic linear peptide that has a miscleaved lysine residue and the SUMO peptide at its N-terminus shown in Figure 5.3 [180, 188]. Identification of SUMO acceptor lysines using such MS/MS spectra in a database search is only possible when the peptide sequences within the database are also modified by SUMO sequence. However, available protein database search engines for experimental fragment spectra do not include SUMO sequence as a putative modification at lysine residues. Simple addition of the molecular weight of the tryptic SUMO peptide to that of a lysine residue within the targeted protein, without obtaining sequence information, would create a large number of false positive hits in database searches. In addition, since protein SUMOylation can theoretically occur at every lysine

residue within a substrate protein, manual construction of such artificial peptides, calculation of molecular weight and matching MS/MS fragment pattern are time-consuming processes.

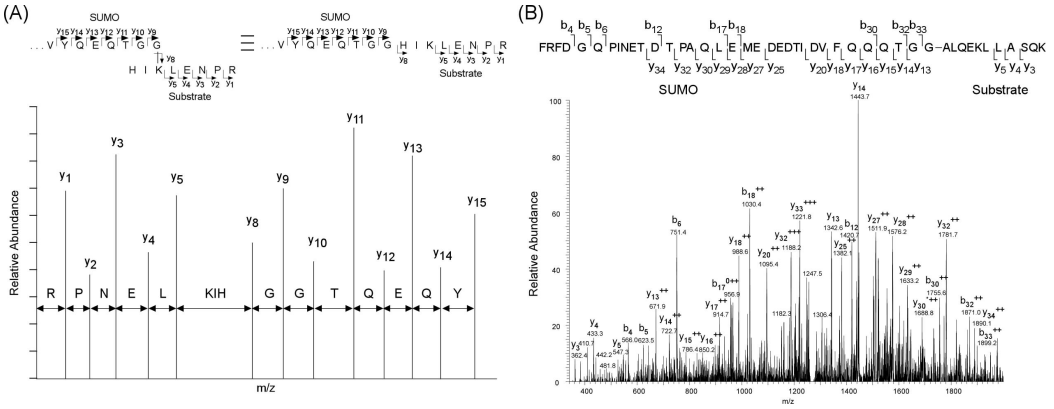


Figure 5.3. The fragment pattern of SUMO-conjugated peptide similar to linear peptide. (A) Branched tryptic peptide conjugated with tryptic SUMO fragment at its lysine acceptor site reveal a similar MS/MS fragmentation pattern as a linear peptide. The y-type ions in the artificial spectrum and in the peptide sequence are indicated. (B) The CID MS/MS spectrum of a tryptic SUMO-conjugated peptide recorded on an Orbitrap mass spectrometer. The Figure depicts the tryptic fragment of USP25 (encompassing positions 711-721) conjugated with SUMO2. The y-type ions in the MS/MS spectrum and in the peptide sequence are indicated.

Accordingly, we developed an algorithm to automatically generate such SUMO-conjugated FASTA sequences of proteins in silico shown in Figure 5.4A. Subsequently, the novel FASTA sequences are implemented in commonly used database search engines to identify lysine acceptor sites for SUMO conjugation shown in Figure 5.4B. More specifically, the FASTA sequence of a putatively SUMOylated protein is “chopped” into tryptic fragments (allowing 0, 1, 2 or n missed cleavages). The tryptic “spice” sequence (e.g. tryptic peptides from SUMO1 or any other ubiquitin-like protein) is attached to the N -terminus of each tryptic peptide that contains a lysine (K) as a missed cleavage site. It is of note that also the ubiquitin-like proteins are allowed to contain 0, 1, 2 or n miscleavage(s). To prevent the appearance of non-natural peptides, a virtual amino acid “J” is attached to the C -terminus of each tryptic fragment before ligation of the generated tryptic fragments into one large FASTA sequence. This large artificial protein sequence is submitted into the database search engine in which the virtual cleavage site “J” is recognized by an artificial endoproteinase that directly cleaves N- and C-terminally to “J” to generate the tryptic fragments for the selected missed cleavages. Subsequently the SUMO acceptor site can be identified by using the applied search engine (e.g. MASCOT, X!Tandem or Sequest). A workflow to set up a modified FASTA sequence in which certain

proteins (or entire databases) can be generated by a user-defined modifier is implemented in the program ChopNSpice (http://chopnspice.gwdg.de/).

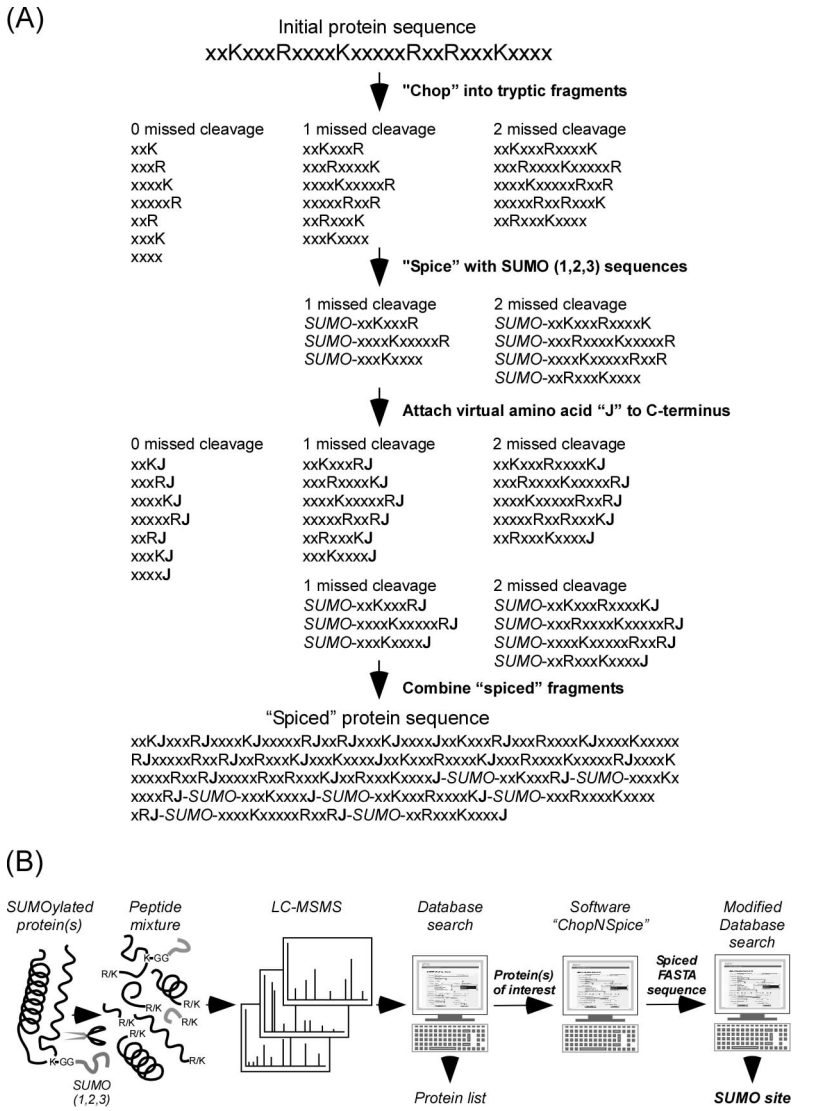


Figure 5.4. The concept of the "ChopNSpice" software. (A) The basic workflow of "ChopNSpice" to generate a "spiced" FASTA sequence from an initial protein sequence in which all lysine residues are putatively modified by SUMO1 or SUMO2/3. The "spiced" FASTA sequence is subsequently used in database searches. (B) The general workflow for identification of SUMO acceptor sites. Sumoylated proteins are digested with endoproteinases and analyzed by LC-MS/MS. The corresponding proteins are identified by a database search engines (MASCOT and/or Sequest). Putatively sumoylated protein sequences are "chopped and spiced" (see A) and the "spiced" FASTA sequences are added to the database. The protein database search is repeated to identify the SUMO-conjugated peptide with its corresponding lysine acceptor site (see text for details).

In practical terms, after enrichment of endogenous SUMO-conjugated proteins or proteins SUMOylated in vitro, putative SUMO substrates are identified by a standard MS-based protein identification, i.e. samples are digested with trypsin and the resulting tryptic peptides are separated by nano-LC, detected and sequenced in the MS. Corresponding proteins in the sample are identified by highly accurate mass of the peptide and matching the fragment spectra against a protein database using e.g. MASCOT, X!Tandem or Sequest as search engines. A second MS and MS/MS analysis under “high mass” conditions is performed where only those precursors exceed a certain mass, i.e. ≥ 2154 Da for SUMO-1 and ≥ 3568 Da for SUMO-2/3, are selected for sequencing. Once one or several putatively SUMOylated proteins have been identified in both the analyses after merging the data/results, MS and MS/MS data are resubmitted to search against the new database containing the virtual SUMOylated protein sequence generated by ChopNSpice (Figure 5.4B). In a subsequent experiment, the same sample can be reinvestigated by extended/modified LC-MS/MS analysis to identify the SUMO acceptor site(s).

Note that both of the search engines used in this study (MASCOT and Sequest) have some shortcomings. For instance, MASCOT does not efficiently search MS/MS fragment spectra that contain fragment ions with a charge state higher than 2; as a consequence, larger SUMOylated peptides with quadruply or higher charge states show a very low score in MASCOT result or are not identified at all. This problem can be circumvented by using either Sequest or other search engines (e.g. X!Tandem) or, alternatively, by using the software tool Raw2msn to deconvolute the higher charge stages of the fragment ions in the raw data to singly-charged fragment ions for MASCOT search [189]. However, a prerequisite for deconvolution process is that MS/MS spectra (generated either by CID or by HCD) are recorded in the Orbitrap analyzer with sufficient mass resolution for the charge state recognition; however, this in turn decreases sensitivity [190]. A comparison between the different systems for processing raw data and the different detection modes of the Orbitrap MS are shown in supplementary Figure 5.5. “Sequest” on the other hand does not allow for cleavage with endoproteinase “J” both N- and C-terminally to J, but rather either N-terminally or C-terminally. Therefore, cleavage of the FASTA sequence is performed unspecifically, i.e., no enzyme is used in silico and matched spectra are validated manually. Confidence in the results from the search engine is achieved by the high mass accuracy of the Orbitrap instrument (< 10 ppm), and by the fact that the validated sequence must be preceded or followed by the virtual amino acid J. Furthermore, all the abundant fragment ions must be assigned to y- and/or b-ion series. However, as a very simple alternative, the single concatenated peptide sequences can be submitted to the database without merging them into a single new FASTA sequence.

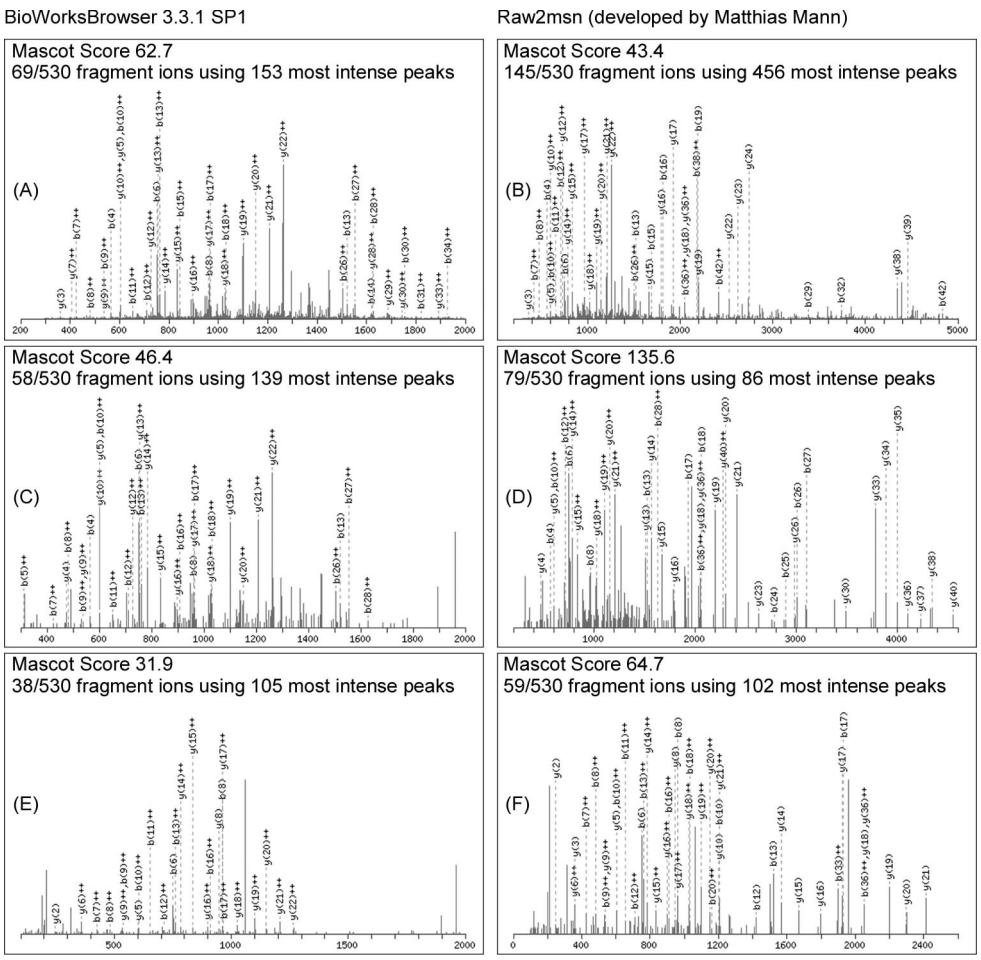


Figure 5.5. MS/MS spectra of USP25 peptide conjugated to SUMO2 at Lysine-754 (M.W.= 5304.5409, 5+, FRFDDQPINETDTPAQLEMEDEDTIDVFQQQTGG-HL**K**EETIQIITK). Left panels (A, C, E): Raw data processed with BioWorksBrowser (ThermoFisherScientific, mgf.file). Right panels (B, D, F): Raw data processed with Raw2msm (msm file - deisotope and deconvolute multiply charge to singly charged state). The software tool is available at: http://www.biochem.mpg.de/mann/publications/2006/0510_01/0510_01.html). **(**A) and (B) are collision-induced dissociation (CID) MS/MS scanning in the linear ion trap (LTQ) that show high sensitivity but a low resolution and additionally a low mass cut off. (C) and (D) are CID MS/MS scanning in the Orbitrap analyzer that shows a medium sensitivity but a high resolution but still a low mass cut off. (E) and (F) are high collision dissociation (HCD) scanning in the Orbitrap analyzer that have low sensitivity but high resolution and do not show any low mass cut off. The comparison of (D) and (C) and (F) and (E) reveals that deconvoluted spectra match to more y- and b-type ions in the MASCOT database search, obtaining high score. Since MASCOT does not support efficient match of MS/MS fragment spectra derived from quadruply and higher charge states, the use of Raw2msm is a pre-requisite for the comprehensive identification of SUMO-conjugated peptides recorded by Orbitrap analyzer.

5.4.2 Identification of SUMO Conjugation Sites in vitro

To validate our approach we applied Ran GTPase-activating protein 1 (RanGAP1), Nuclear autoantigen Sp-100 (Sp100), Cellular tumor antigen p53 (p53) and SUMO-activating enzyme subunit 2 (uba2) to an in vitro SUMOylation reaction with SUMO1 and SUMO2. Proteins migrating on SDS-PAGE with a higher apparent molecular weight than the original proteins were considered to be SUMOylated (Figure 5.6) and were then processed by nanoLC-MS/MS analysis as described above.

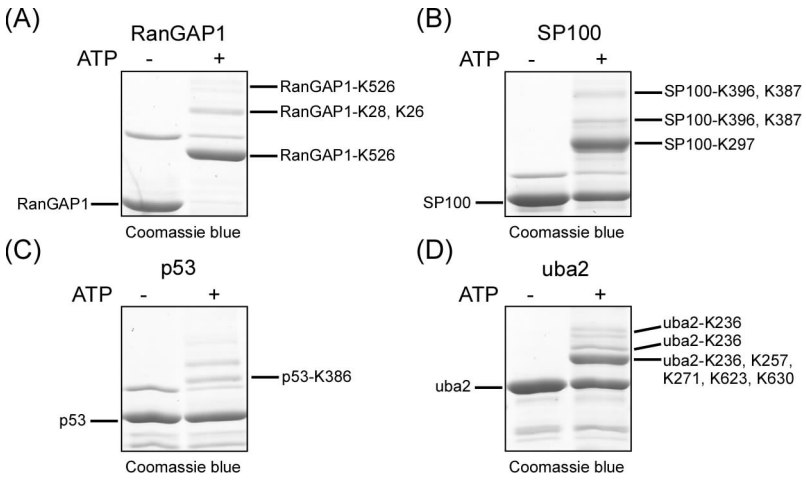


Figure 5.6. Detection of SUMO1-acceptor sites from in vitro sumoylated proteins. (A) In vitro sumoylation of 1 µg RanGAP1 with 100 ng Aos1/Uba2, 200 ng Ubc9, 2.5 µg SUMO1 for 1 hour at 30 ℃. Proteins were visualized by Coomassie blue staining. (B) In vitro sumoylation of 1 µg Sp100 as in (A). (C) In vitro sumoylation of 1 µg GST-p53 as in (A). (D) In vitro sumoylation of 1 µg Aos1-Uba2 as in (A). The acceptor sites identified are indicated at the protein band from which they were discovered.

For identification of SUMO-conjugated peptides, we first tested SUMO as a variable modification on lysines (2154 Da for SUMO 1; 3568 Da for SUMO2) using two commonly used protein identification tools MASCOT and Sequest. Like other groups [182], we were unable to identify any SUMOylated peptides by the standard nanoLC-MS proteomic procedure and subsequent database search. Whereas these search engines rely on the database that contain putative protein sequences and the extra mass added to a particular amino acid for PTMs identification. However, it is of the utmost importance to obtain sequence information not only from the substrate peptide but also from the modifier. Although manual identification of SUMO-conjugated sites was possible, it required laborious searching in the MS spectra for modified peptides [176]. In contrast, by using the ChopNSpice software on the identified protein sequences and subsequent database search with MASCOT and Sequest, we readily identified SUMO modification of RanGAP1 on lysine 526, of p53 on lysine 386 and of Sp100

on lysine 297. In addition, we observed several minor acceptor sites, also observed by others [191-193]. Furthermore, we discovered that numerous, so far unidentified, lysine residues within the SUMO E1 activating enzyme Uba2 are conjugated with SUMO1 and SUMO2 shown in Table 5.1. However, consistent with the identification of multiple acceptor sites, mutations of single lysine residues within Uba2 did not significantly impair its SUMOylation.

Table 5.1 The in vitro sumoylation assays of RanGap1, SP100, p53 and uba2. The sequence of the SUMOylated peptide and the positions of SUMO-acceptor sites as determined by MS and MS/MS using "ChopNSpice" in combination with MASCOT and SEQUEST as search engines are listed.

Modification	Protein Name	MASCOT Score	Sequest XCorr	Sequence	Site
SUMO1	RanGap1	110.29	6.45	ELGM$_{ox}$EEEDVIEVYQEQTGG-LLIHM$_{ox}$GLLKSEDK	K526
		61.9	4.81	ELGM$_{ox}$EEEDVIEVYQEQTGG-GKGLK	K28
		72.02	5.50	ELGM$_{ox}$EEEDVIEVYQEQTGG-TQVAGGQLSFKGK	K26
	GST-SP100	57.22	5.08	ELGM$_{ox}$EEEDVIEVYQEQTGG-LVDIKK	K297
		50.53	5.40	ELGMEEEDVIEVYQEQTGG-ESFKK	K396
		70.11	4.90	ELGMEEEDVIEVYQEQTGG-KLSTFR	K387
	p53	87.38	6.22	ELGMEEEDVIEVYQEQTGG-HM$_{ox}$FKTEGPDSD	K386
	uba2	99.43	5.97	ELGMEEEDVIEVYQEQTGG-ENLSAKR	K623
		46.63	4.25	ELGM$_{ox}$EEEDVIEVYQEQTGG-LFTKLFK	K257
		160.71	7.12	ELGMEEEDVIEVYQEQTGG-ASNEDGDIKR	K236
		124.39	7.65	ELGMEEEDVIEVYQEQTGG-YLLTMDKLWR	K271
		41.8	3.80	ELGM$_{ox}$EEEDVIEVYQEQTGG-SRIEQKEELDDVIALD	K630
SUMO2	RanGap1	118.17	8.27	FDGQPINETDTPAQLEMEDEDTIDVFQQQTGG-LLIHMGLLKSEDK	K526
	YFP-SP100	121.98	7.94	FDGQPINETDTPAQLEM$_{ox}$EDEDTIDVFQQQTGG-LVDIKK	K297
		80.63	6.97	FDGQPINETDTPAQLEM$_{ox}$EDEDTIDVFQQQTGG-ESFKK	K396
		136.33	7.16	FDGQPINETDTPAQLEM$_{ox}$EDEDTIDVFQQQTGG-KLSTFR	K387
		77.83	6.56	FRFDGQPINETDTPAQLEM$_{ox}$EDEDTIDVFQQQTGG-SKHGEK	K426
	GST-SP100	85.66	6.56	FDGQPINETDTPAQLEMEDEDTIDVFQQQTGG-ESFKK	K396
		132.34	7.25	FDGQPINETDTPAQLEMEDEDTIDVFQQQTGG-KLSTFR	K387
	p53	110.79	6.81	FDGQPINETDTPAQLEMEDEDTIDVFQQQTGG-HMFKTEGPDSD	K386
	uba2	152.29	9.75	FRFDGQPINETDTPAQLEMEDEDTIDVFQQQTGG-ASNEDGDIKR	K236
		45.79	5.77	FRFDGQPINETDTPAQLEM$_{ox}$EDEDTIDVFQQQTGG-SKAQVAK	K72
	SUMO2	73.74	6.74	FDGQPINETDTPAQLEM$_{ox}$EDEDTIDVFQQQTGG-GSHM$_{ox}$ADEKPK	K5
		85.81	7.98	FDGQPINETDTPAQLEM$_{ox}$EDEDTIDVFQQQTGG-EGVKTENNDHINLK	K11

5.4.3 Increasing Sensitivity by using “High Mass” Acquisition

In earlier cooperation with Prof. Dr. Frauke Melchior’s laboratory, we mapped two SUMO-conjugated sites within ubiquitin carboxyl-terminal hydrolase 25 (USP25) by which we identified lysine 141 using a mutagenesis approach while the other lysine 99 was identified using MS approach. It is of note that in our previous study we used a small fragment of USP25 that was conjugated with SUMO2 in bacteria followed by purification by gel-filtration and anion-exchange chromatography [176]. However, manual examination of full-length USP25 SUMOylated in vitro did not reveal any SUMO acceptor site. To test whether our ChopNSpice method has an increased sensitivity to identify the acceptor sites of this more complex sample, we conjugated full-length USP25 with SUMO2 in vitro, using the E3 ligase PiasXα, as described previously [176]. Next, the mixture was digested with trypsin in solution. Subsequently, to increase sensitivity for the identification of SUMO acceptor sites, we also used “high mass” MS/MS acquisition conditions (Figure 5.7A, compare the standard - upper panel - with the high mass - lower panel). Under these conditions, only peptides with a mass exceeding 2154 Da (for SUMO 1) or 3568 Da (for SUMO 2/3) are selected for MS/MS acquisition.

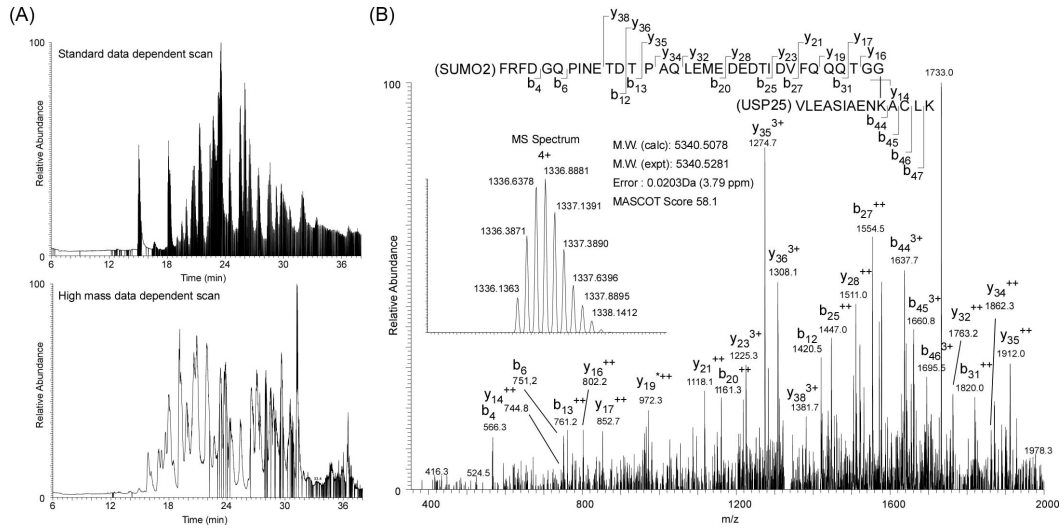


Figure 5.7. Increased sensitivity to discover SUMO acceptor sites. (A) Comparison of the Total Ion Count (TIC) under standard conditions (upper panel) with the TIC under “high mass” conditions (lower panel). The black lines indicate the MS/MS experiment performed. (B) MS/MS CID spectrum of a tryptic peptide (m/z = 1336.1363) derived from USP25 encompassing positions 132-145 with fragment ions recorded in the LTQ-Orbitrap MS. MS/MS in combination with database search of the modified USP25 sequence (using ChopNSpice) identified Lys-141 as actual SUMO site. Y- and b-type ions are shown in the spectrum and at their respective positions in the conjugated peptide. It is of note that Lys 141 was identified under the high mass conditions only.

This approach is highly suitable for the accurate detection and sequencing of larger peptides and additionally facilitates detection of lower-abundance SUMO conjugations. A database search against modified sequences (achieved by the program ChopNSpice in combination with MASCOT) demonstrated that SUMOylated peptides were enriched by high-mass MS/MS acquisition shown in Table 5.2. Using this strategy, we went on to identify several additional SUMO acceptor sites within full-length USP25, including lysine 141, which had previously been identified only by a mutational approach. In addition, we observed lysine 5 in SUMO2 as an acceptor site for chain formation, consistent with a previous report [180].

Table 5.2. The in vitro sumoylation assays of USP25 by using either standard data dependent acquisition (DDA) or high mass data dependent acquisition in the LTQ-Orbitrap MS. The sequence of the SUMOylated peptide and the positions of SUMO-acceptor sites as determined by MS and MS/MS using "ChopNSpice" in combination with MASCOT and SEQUEST as search engines are listed.

Protein Name	Mascot Score	Sequest XCorr	Sequence	Site	Stadnard DDA	High M.W. DDA
USP25	181.95	8.62	ACLKR	K145		X
	87.73	7.89	NQKEHR	K214		X
	187.12	8.89	LLASQKLR	K721	X	X
	128.21	9.48	ALQEKLLASQK	K715		X
	75.48	8.18	SVIHKPFTQSR	K520		X
	135.63	7.99	HLKEETIQIITK	K754		X
	142.99	9.70	VLEASIAENKACLK	K141		X
	17.93	5.68	SFHEPPKLPSYSTHELCER	K1059		X
SUMO2	93.67	8.49	GSHMADEKPK	K5		X

5.4.4 Identification of SUMO-Conjugated Sites in vivo

Although the identification of SUMO conjugation sites in endogenous proteins from yeast has been performed before [177], unbiased identification of SUMO acceptor sites in higher eukaryotes has remained a technical challenge. This can partly be accounted for by the high mass of SUMO after hydrolysis with trypsin in higher eukaryotes combined with the low abundance of post-translational modifications per se as compared with the amount of non-modified protein. Additionally, chemical enrichment for modifications with SUMO prior to MS has not been described, as is the case for instance with phosphorylation [75, 103, 194]. To examine the power of our strategy for identification of SUMO conjugation sites we purified endogenous SUMO1 conjugates from HeLa cells shown in Figure 5.8. Although the overall protein composition in the immunoprecipitation of SUMO1 conjugates seem indistinguishable from the control IP in Coomassie blue staining, the western blot clearly demonstrates enrichment of SUMO1 conjugates in the IP (Figure 5.8A). The gel was cut into slices and the proteins specifically present in the SUMO1 immunoprecipitation were identified by LC-MS/MS. One of the most prominent SUMO1 conjugates was found at 90kDa and represents RanGAP1

conjugated with SUMO1 [187]. By applying our ChopNSpice approach, we were able to identify lysine 524 in endogenous RanGAP1 with endogenous SUMO1 (Figure 5.8B).

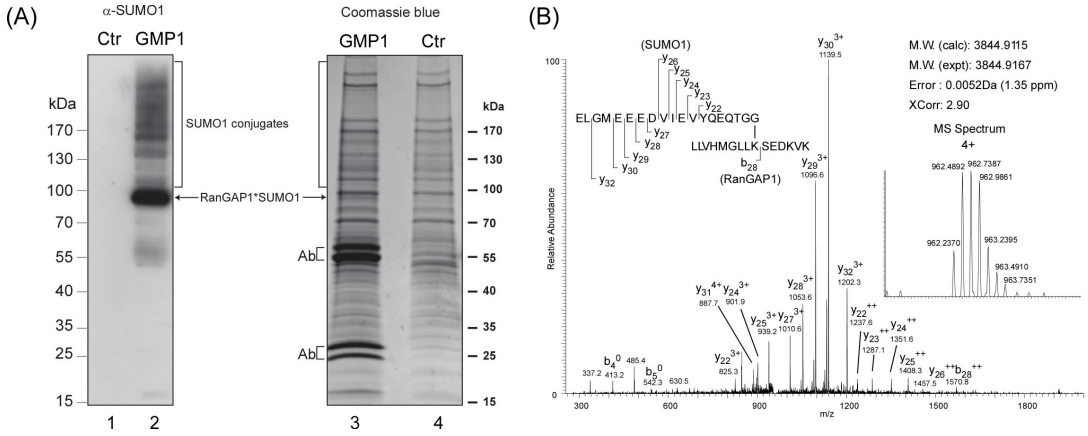


Figure 5.8. Identification of SUMO acceptor sites in endogenous proteins. (A) SUMO1 conjugated proteins were isolated from HeLa cells using SUMO1 antibodies coupled to protein G agarose or control protein G agarose. Immunoprecipitates were extensively washed and eluted with Sample buffer. Five percent of the sample was loaded to detect SUMO1 conjugated species by Western blot (left panel), the rest of the sample was used for to identify SUMO acceptor sites by MS (right panel). RanGAP1 conjugated with SUMO1 is indicated by the arrows. (B) MS/MS CID spectrum of a tryptic peptide (m/z = 962.2370) derived from RanGAP1 encompassing positions 516-530 with fragment ions recorded in the LTQ-Orbitrap MS. MS/MS in combination with database searches of the modified RanGAP1 sequence (by ChopNSpice) confirmed the known Lys-524 as actual SUMO site. Y- and b-type ions are shown in the spectrum and at their respective positions in the conjugated peptide. XCorr is the score in the database search using Sequest.

Importantly, in a subsequent experiment, we could additionally identify SUMO acceptor lysine residues in SUMO 1, SUMO 2/3, Ubc9, RanBP2, and others. Although several of these proteins were known as SUMO targets, the SUMO acceptor sites within RanBP2 have not been described before. Interestingly, in the SUMO1 immuno-precipitate we also observe SUMO2 conjugated to SUMO2 on Lysine 11. In total, 10 SUMOylated proteins corresponding to 17 distinct sites were identified. 15 SUMOylated sites were identified in this study appeared to be novel, which may provide a valuable resource to the biological research community. shown in Table 5.3. Thus, our MS approach proved to be highly reliable, and it easily and specifically identified SUMO acceptor sites both in vitro and in vivo. Thereby, our method increases the sensitivity of the identification of SUMO conjugation sites in mammalian cells.

Table 5.3. In vivo SUMOylated proteins derived after immunoprecipitation from HeLa cells using anti-SUMO 1 antibody. The sequence of the SUMOylated peptide and the positions of SUMO-acceptor sites as determined by MS and MS/MS using "ChopNSpice" in combination with MASCOT and Sequest as search engines are listed.

Modification	Conjugated Protein Name	Sequence	Site	Mascot Score	Sequest XC
SUMO1	Ran GTPase-activating protein 1	^{516}LLVHMGLLKSEDKVK^{530}	K524	70,29	6,10
	Small ubiquitin-related modifier 1	^{17}KEGEYIK^{23}	K17	92,87	5,60
	Small ubiquitin-related modifier 2	^{8}EGVKTENNDHINLK^{21}	K11	79,50	5,57
	SUMO-conjugating enzyme UBC9	^{142}VEYEKR^{147}	K146	59,84	5,18
	E3 SUMO-protein ligase RanBP2	^{100}IAELLCKNDVTDGRAKYWLER^{120}	-	14,31	4,34
		^{1408}FALVTPKK^{1415}	K1414	36,53	4,22
		^{1715}SGFEGMFTKK^{1724}	K1723	106,23	5,46
		^{2255}KNLFR^{2259}	K2255	35,55	2,86
		^{2424}FKLQDVADSFKK^{2434}	K2433	94,41	5,12
		^{2507}AVVSPPKFVFGSESVK^{2522}	K2513	29,65	2,59
		^{2581}NSDIEQSSDSKVK^{2594}	K2592	97,71	6,57
		^{2617}AKEK^{2620}	K2618	44,60	6,22
	RANBP2-like and GRIP domain-containing protein 4	^{691}KAEDIANDALSPEEQEECK^{709}	K691	7,29	-
	Chromodomain-helicase-DNA-binding protein 8	^{581}YTEDLDIKITDDEEEEEVDVTGPIK^{609}	K592	-	4,02
	Cytoplasmic dynein 1 heavy chain 1	^{3207}KIKETVDQVEELR^{3219}	K3207	20,56	4,21
	Very long-chain specific acyl-CoA dehydrogenase	^{633}NFKSISK^{639}	K635	46,13	-
	Bifunctional aminoacyl-tRNA synthetase	^{314}NPIEKNLQMWEEMK^{327}	K318	51,70	-
SUMO2	Small ubiquitin-related modifier 2	^{8}EGVKTENNDHINLK^{21}	K11	37,99	7,31

5.5 Conclusion

In this study, we present a freely available computational approach to identify SUMO acceptor sites in endogenous proteins by mass spectrometry that cannot easily be explored by using common search engines such as SEQUEST and/or MASCOT in a rapid and sensitive manner. The software "ChopNspice" for the identification of SUMO acceptor sites is unique in its ability to allow the user (I) to combine two protein sequences in a linear manner; (II) to generate any modified linear protein sequence that contains any modifications at the N-terminus of the novel fused sequence; (III) to introduce defined extra masses in either of the two protein sequences so that also peptide-peptide crosslink (using a crosslinking reagent) after enzymatic digestion of crosslinked proteins can be explored and identified, and (IV) to generate a m/z list of all linearly fused peptides. The latter is particularly useful when users do not have access to e.g. an Orbitrap MS or FT-IRC MS but instead would like to use a simple peptide mass fingerprint analysis by MALDI-MS of putatively SUMOylated proteins. In addition, the list serves as inclusion list in LC-MS/MS analysis such that predicted modified (e.g. SUMOylated) peptides

are chosen for fragmentation within the mass spectrometer. By using ChopNspice, a multiple SUMOylated FASTA sequence of a proteins is generated in silico. The new FASTA sequence is implemented in the MS-based database search engines to determine the SUMOylated sites. The high-mass data dependent acquisition is highly suitable for the accurate detection and sequencing of larger peptides and additionally facilitates detection of lower abundant SUMO conjugations. In conclusion, we demonstrate our approach is of value in MS-based analysis and subsequent database search for the identification of SUMO-conjugated sites in a more efficient time scale. We believe that this approach has the potential to be widely used in the foreseeable future.

References

[1] Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W.*, et al.*, The sequence of the human genome. *Science* 2001, *291*, 1304-1351.

[2] Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C.*, et al.*, Initial sequencing and analysis of the human genome. *Nature* 2001, *409*, 860-921.

[3] Wasinger, V. C., Cordwell, S. J., Cerpa-Poljak, A., Yan, J. X.*, et al.*, Progress with gene-product mapping of the Mollicutes: Mycoplasma genitalium. *Electrophoresis* 1995, *16*, 1090-1094.

[4] Wilkins, M. R., Pasquali, C., Appel, R. D., Ou, K.*, et al.*, From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis. *Nature Biotechnology* 1996, *14*, 61-65.

[5] James, P., Protein identification in the post-genome era: the rapid rise of proteomics. *Q Rev Biophys* 1997, *30*, 279-331.

[6] van Weeren, P. C., de Bruyn, K. M., de Vries-Smits, A. M., van Lint, J., Burgering, B. M., Essential role for protein kinase B (PKB) in insulin-induced glycogen synthase kinase 3 inactivation. Characterization of dominant-negative mutant of PKB. *J Biol Chem* 1998, *273*, 13150-13156.

[7] Thrower, J. S., Hoffman, L., Rechsteiner, M., Pickart, C. M., Recognition of the polyubiquitin proteolytic signal. *EMBO J* 2000, *19*, 94-102.

[8] Hay, R. T., SUMO: a history of modification. *Mol Cell* 2005, *18*, 1-12.

[9] Shapiro, A. L., Vinuela, E., Maizel, J. V., Jr., Molecular weight estimation of polypeptide chains by electrophoresis in SDS-polyacrylamide gels. *Biochem Biophys Res Commun* 1967, *28*, 815-820.

[10] Weber, K., Osborn, M., The reliability of molecular weight determinations by dodecyl sulfate-polyacrylamide gel electrophoresis. *J Biol Chem* 1969, *244*, 4406-4412.

[11] O'Farrell, P. H., High resolution two-dimensional electrophoresis of proteins. *J Biol Chem* 1975, *250*, 4007-4021.

[12] Gorg, A., Weiss, W., Dunn, M. J., Current two-dimensional electrophoresis technology for proteomics. *Proteomics* 2004, *4*, 3665-3685.

[13] Rabilloud, T., Two-dimensional gel electrophoresis in proteomics: old, old fashioned, but it still climbs up the mountains. *Proteomics* 2002, *2*, 3-10.

[14] Pappin, D. J., Hojrup, P., Bleasby, A. J., Rapid identification of proteins by peptide-mass fingerprinting. *Curr Biol* 1993, *3*, 327-332.

[15] Henzel, W. J., Billeci, T. M., Stults, J. T., Wong, S. C.*, et al.*, Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases. *Proc Natl Acad Sci U S A* 1993, *90*, 5011-5015.

[16] Mann, M., Hojrup, P., Roepstorff, P., Use of Mass-Spectrometric Molecular-Weight Information to Identify Proteins in Sequence Databases. *Biol Mass Spectrom* 1993, *22*,

338-345.

[17] James, P., Quadroni, M., Carafoli, E., Gonnet, G., Protein identification by mass profile fingerprinting. *Biochem Biophys Res Commun* 1993, *195*, 58-64.

[18] Yates, J. R., 3rd, Speicher, S., Griffin, P. R., Hunkapiller, T., Peptide mass maps: a highly informative approach to protein identification. *Anal Biochem* 1993, *214*, 397-408.

[19] Link, A. J., Eng, J., Schieltz, D. M., Carmack, E.*, et al.*, Direct analysis of protein complexes using mass spectrometry. *Nat Biotechnol* 1999, *17*, 676-682.

[20] Washburn, M. P., Wolters, D., Yates, J. R., 3rd, Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol* 2001, *19*, 242-247.

[21] Kim, S. C., Sprung, R., Chen, Y., Xu, Y.*, et al.*, Substrate and functional diversity of lysine acetylation revealed by a proteomics survey. *Mol Cell* 2006, *23*, 607-618.

[22] Ong, S. E., Mittler, G., Mann, M., Identifying and quantifying in vivo methylation sites by heavy methyl SILAC. *Nat Methods* 2004, *1*, 119-126.

[23] Ross, A. H., Baltimore, D., Eisen, H. N., Phosphotyrosine-containing proteins isolated by affinity chromatography with antibodies to a synthetic hapten. *Nature* 1981, *294*, 654-656.

[24] Cummings, R. D., Kornfeld, S., Fractionation of asparagine-linked oligosaccharides by serial lectin-Agarose affinity chromatography. A rapid, sensitive, and specific technique. *J Biol Chem* 1982, *257*, 11235-11240.

[25] Machida, K., Mayer, B. J., Nollau, P., Profiling the global tyrosine phosphorylation state. *Mol Cell Proteomics* 2003, *2*, 215-233.

[26] Oda, Y., Nagasu, T., Chait, B. T., Enrichment analysis of phosphorylated proteins as a tool for probing the phosphoproteome. *Nat Biotechnol* 2001, *19*, 379-382.

[27] Wells, L., Vosseller, K., Cole, R. N., Cronshaw, J. M.*, et al.*, Mapping sites of O-GlcNAc modification using affinity tags for serine and threonine post-translational modifications. *Mol Cell Proteomics* 2002, *1*, 791-804.

[28] Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F., Whitehouse, C. M., Electrospray ionization for mass spectrometry of large biomolecules. *Science* 1989, *246*, 64-71.

[29] Tanaka, K., The origin of macromolecule ionization by laser irradiation (Nobel lecture). *Angew Chem Int Ed Engl* 2003, *42*, 3860-3870.

[30] Karas, M., Bachmann, D., Hillenkamp, F., Influence of the Wavelength in High-Irradiance Ultraviolet-Laser Desorption Mass-Spectrometry of Organic-Molecules. *Analytical Chemistry* 1985, *57*, 2935-2939.

[31] Karas, M., Hillenkamp, F., Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal Chem* 1988, *60*, 2299-2301.

[32] Taylor, G., Disintegration of Water Drops in Electric Field. *Proc R Soc Lon Ser-A* 1964, *280*, 383-+.

[33] Wilm, M. S., Mann, M., Electrospray and Taylor-Cone Theory, Doles Beam of

Macromolecules at Last. *Int J Mass Spectrom* 1994, *136*, 167-180.

[34] Hillenkamp, F., Karas, M., Beavis, R. C., Chait, B. T., Matrix-assisted laser desorption/ionization mass spectrometry of biopolymers. *Anal Chem* 1991, *63*, 1193A-1203A.

[35] Aebersold, R., Mann, M., Mass spectrometry-based proteomics. *Nature* 2003, *422*, 198-207.

[36] Leary, J. J., Schmidt, R. L., Quadrupole mass spectrometers: An intuitive leak at the math. *J Chem Educ* 1996, *73*, 1142-1145.

[37] Steel, C., Henchman, M., Understanding the quadrupole mass filter through computer simulation. *J Chem Educ* 1998, *75*, 1049-1054.

[38] Domon, B., Aebersold, R., Mass spectrometry and protein analysis. *Science* 2006, *312*, 212-217.

[39] Vanberkel, G. J., Glish, G. L., Mcluckey, S. A., Electrospray Ionization Combined with Ion Trap Mass-Spectrometry. *Analytical Chemistry* 1990, *62*, 1284-1295.

[40] Jonscher, K. R., Yates, J. R., 3rd, The quadrupole ion trap mass spectrometer--a small solution to a big challenge. *Anal Biochem* 1997, *244*, 1-15.

[41] Weickhardt, C., Moritz, F., Grotemeyer, J., Time-of-flight mass spectrometry: State-of-the-art in chemical analysis and molecular science. *Mass Spectrometry Reviews* 1996, *15*, 139-162.

[42] Marshall, A. G., Hendrickson, C. L., Jackson, G. S., Fourier transform ion cyclotron resonance mass spectrometry: a primer. *Mass Spectrom Rev* 1998, *17*, 1-35.

[43] Makarov, A., Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis. *Anal Chem* 2000, *72*, 1156-1162.

[44] Hu, Q., Noll, R. J., Li, H., Makarov, A.*, et al.*, The Orbitrap: a new mass spectrometer. *J Mass Spectrom* 2005, *40*, 430-443.

[45] Chernushevich, I. V., Loboda, A. V., Thomson, B. A., An introduction to quadrupole-time-of-flight mass spectrometry. *J Mass Spectrom* 2001, *36*, 849-865.

[46] Medzihradszky, K. F., Campbell, J. M., Baldwin, M. A., Falick, A. M.*, et al.*, The characteristics of peptide collision-induced dissociation using a high-performance MALDI-TOF/TOF tandem mass spectrometer. *Analytical Chemistry* 2000, *72*, 552-558.

[47] Roepstorff, P., Fohlman, J., Proposal for a Common Nomenclature for Sequence Ions in Mass-Spectra of Peptides. *Biomed Mass Spectrom* 1984, *11*, 601-601.

[48] Johnson, R. S., Martin, S. A., Biemann, K., Stults, J. T., Watson, J. T., Novel Fragmentation Process of Peptides by Collision-Induced Decomposition in a Tandem Mass-Spectrometer - Differentiation of Leucine and Isoleucine. *Analytical Chemistry* 1987, *59*, 2621-2625.

[49] Eng, J. K., Mccormack, A. L., Yates, J. R., An Approach to Correlate Tandem Mass-Spectral Data of Peptides with Amino-Acid-Sequences in a Protein Database. *J Am Soc Mass Spectr* 1994, *5*, 976-989.

[50] Perkins, D. N., Pappin, D. J. C., Creasy, D. M., Cottrell, J. S., Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 1999, *20*, 3551-3567.

[51] Will, C. L., Luhrmann, R., In: Gesteland, R. F., Cech, J. F., and Atkins, J. F., eds. The RNA World, 3rd Ed., *Spliceosome Structure and Function*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. 2006, 369-400.

[52] Mermoud, J. E., Cohen, P. T., Lamond, A. I., Regulation of mammalian spliceosome assembly by a protein phosphorylation mechanism. *EMBO J* 1994, *13*, 5679-5688.

[53] Misteli, T., RNA splicing: What has phosphorylation got to do with it? *Curr Biol* 1999, *9*, R198-+.

[54] Bollen, M., Beullens, M., Signaling by protein phosphatases in the nucleus. *Trends Cell Biol* 2002, *12*, 138-145.

[55] Soret, J., Tazi, J., Phosphorylation-dependent control of the pre-mRNA splicing machinery. *Prog Mol Subcell Biol* 2003, *31*, 89-126.

[56] Hunter, T., Signaling--2000 and beyond. *Cell* 2000, *100*, 113-127.

[57] Mann, M., Ong, S. E., Gronborg, M., Steen, H.*, et al.*, Analysis of protein phosphorylation using mass spectrometry: deciphering the phosphoproteome. *Trends Biotechnol* 2002, *20*, 261-268.

[58] Mann, M., Jensen, O. N., Proteomic analysis of post-translational modifications. *Nat Biotechnol* 2003, *21*, 255-261.

[59] Kalume, D. E., Molina, H., Pandey, A., Tackling the phosphoproteome: tools and strategies. *Curr Opin Chem Biol* 2003, *7*, 64-69.

[60] Peters, E. C., Brock, A., Ficarro, S. B., Exploring the phosphoproteome with mass spectrometry. *Mini Rev Med Chem* 2004, *4*, 313-324.

[61] Conrads, T. P., Issaq, H. J., Veenstra, T. D., New tools for quantitative phosphoproteome analysis. *Biochem Biophys Res Commun* 2002, *290*, 885-890.

[62] Motoyama, A., Xu, T., Ruse, C. I., Wohlschlegel, J. A., Yates, J. R., 3rd, Anion and cation mixed-bed ion exchange for enhanced multidimensional separations of peptides and phosphopeptides. *Anal Chem* 2007, *79*, 3623-3634.

[63] Lim, K. B., Kassel, D. B., Phosphopeptides enrichment using on-line two-dimensional strong cation exchange followed by reversed-phase liquid chromatography/mass spectrometry. *Analytical Biochemistry* 2006, *354*, 213-219.

[64] Dai, J., Jin, W. H., Sheng, Q. H., Shieh, C. H.*, et al.*, Protein phosphorylation and expression profiling by Yin-yang multidimensional liquid chromatography (Yin-yang MDLC) mass spectrometry. *J Proteome Res* 2007, *6*, 250-262.

[65] Beausoleil, S. A., Jedrychowski, M., Schwartz, D., Elias, J. E.*, et al.*, Large-scale characterization of HeLa cell nuclear phosphoproteins. *Proc Natl Acad Sci U S A* 2004, *101*, 12130-12135.

[66] Ballif, B. A., Villen, J., Beausoleil, S. A., Schwartz, D., Gygi, S. P., Phosphoproteomic analysis of the developing mouse brain. *Mol Cell Proteomics* 2004, *3*, 1093-1101.

[67] Machida, M., Kosako, H., Shirakabe, K., Kobayashi, M.*, et al.*, Purification of phosphoproteins by immobilized metal affinity chromatography and its application to phosphoproteome analysis. *FEBS J* 2007, *274*, 1576-1587.

[68] Dubrovska, A., Souchelnytskyi, S., Efficient enrichment of intact phosphorylated proteins by modified immobilized metal-affinity chromatography. *Proteomics* 2005, *5*, 4678-4683.

[69] Zhang, X., Ye, J., Jensen, O. N., Roepstorff, P., Highly Efficient Phosphopeptide Enrichment by Calcium Phosphate Precipitation Combined with Subsequent IMAC Enrichment. *Mol Cell Proteomics* 2007, *6*, 2032-2042.

[70] Andersson, L., Porath, J., Isolation of phosphoproteins by immobilized metal (Fe3+) affinity chromatography. *Anal Biochem* 1986, *154*, 250-254.

[71] Muszynska, G., Dobrowolska, G., Medin, A., Ekman, P., Porath, J. O., Model studies on iron(III) ion affinity chromatography. II. Interaction of immobilized iron(III) ions with phosphorylated amino acids, peptides and proteins. *J Chromatogr* 1992, *604*, 19-28.

[72] Gruhler, A., Olsen, J. V., Mohammed, S., Mortensen, P.*, et al.*, Quantitative phosphoproteomics applied to the yeast pheromone signaling pathway. *Mol Cell Proteomics* 2005, *4*, 310-327.

[73] Pandey, A., Andersen, J. S., Mann, M., Use of mass spectrometry to study signaling pathways. *Sci STKE* 2000, *2000*, pl1.

[74] Corthals, G. L., Aebersold, R., Goodlett, D. R., Identification of phosphorylation sites using microimmobilized metal affinity chromatography. *Methods Enzymol* 2005, *405*, 66-81.

[75] Larsen, M. R., Thingholm, T. E., Jensen, O. N., Roepstorff, P., Jorgensen, T. J., Highly selective enrichment of phosphorylated peptides from peptide mixtures using titanium dioxide microcolumns. *Mol Cell Proteomics* 2005, *4*, 873-886.

[76] Sano, A., Nakamura, H., Titania as a chemo-affinity support for the column-switching HPLC analysis of phosphopeptides: Application to the characterization of phosphorylation sites in proteins by combination with protease digestion and electrospray ionization mass Spectrometry. *Anal Sci* 2004, *20*, 861-864.

[77] Sano, A., Nakamura, H., Chemo-affinity of titania for the column-switching HPLC analysis of phosphopeptides. *Anal Sci* 2004, *20*, 565-566.

[78] Kuroda, I., Shintani, Y., Motokawa, M., Abe, S., Furuno, M., Phosphopeptide-selective column-switching RP-HPLC with a titania precolumn. *Anal Sci* 2004, *20*, 1313-1319.

[79] Pinkse, M. W., Uitto, P. M., Hilhorst, M. J., Ooms, B., Heck, A. J., Selective isolation at the femtomole level of phosphopeptides from proteolytic digests using 2D-NanoLC-ESI-MS/MS and titanium oxide precolumns. *Anal Chem* 2004, *76*, 3935-3943.

[80] Li, Y., Liu, Y., Tang, J., Lin, H.*, et al.*, Fe3O4@Al2O3 magnetic core-shell microspheres for rapid and highly specific capture of phosphopeptides with mass spectrometry analysis. *J*

Chromatogr A 2007, *1172*, 57-71.

[81] Li, Y., Lin, H. Q., Deng, C. H., Yang, P. Y., Zhang, X. M., Highly selective and rapid enrichment of phosphorylated peptides using gallium oxide-coated magnetic microspheres for MALDI-TOF-MS and nano-LC-ESI-MS/MS/MS analysis. *Proteomics* 2008, *8*, 238-249.

[82] Zhou, H., Tian, R., Ye, M., Xu, S.*, et al.*, Highly specific enrichment of phosphopeptides by zirconium dioxide nanoparticles for phosphoproteome analysis. *Electrophoresis* 2007, *28*, 2201-2215.

[83] Cuccurullo, M., Schlosser, G., Cacace, G., Malorni, L., Pocsfalvi, G., Identification of phosphoproteins and determination of phosphorylation sites by zirconium dioxide enrichment and SELDI-MS/MS. *J Mass Spectrom* 2007, *42*, 1069-1078.

[84] Hsiao, H. H., Hsieh, H. Y., Chou, C. C., Lin, S. Y.*, et al.*, Concerted experimental approach for sequential mapping of peptides and phosphopeptides using C18-functionalized magnetic nanoparticles. *J Proteome Res* 2007, *6*, 1313-1324.

[85] Li, Y., Leng, T. H., Lin, H. Q., Deng, C. H.*, et al.*, Preparation of Fe3O4@ZrO2 core-shell microspheres as affinity probes for selective enrichment and direct determination of phosphopeptides using matrix-assisted laser desorption ionization mass spectrometry. *Journal of Proteome Research* 2007, *6*, 4498-4510.

[86] Chen, C. T., Chen, W. Y., Tsai, P. J., Chien, K. Y.*, et al.*, Rapid enrichment of phosphopeptides and phosphoproteins from complex samples using magnetic particles coated with alumina as the concentrating probes for MALDI MS analysis. *Journal of Proteome Research* 2007, *6*, 316-325.

[87] Kweon, H. K., Hakansson, K., Selective zirconium dioxide-based enrichment of phosphorylated peptides for mass spectrometric analysis. *Anal Chem* 2006, *78*, 1743-1749.

[88] Bodenmiller, B., Mueller, L. N., Mueller, M., Domon, B., Aebersold, R., Reproducible isolation of distinct, overlapping segments of the phosphoproteome. *Nat Methods* 2007, *4*, 231-237.

[89] Zheng, H., Hu, P., Quinn, D. F., Wang, Y. K., Phosphotyrosine proteomic study of interferon alpha signaling pathway using a combination of immunoprecipitation and immobilized metal affinity chromatography. *Mol Cell Proteomics* 2005, *4*, 721-730.

[90] Stupak, J., Liu, H., Wang, Z., Brix, B. J.*, et al.*, Nanoliter sample handling combined with microspot MALDI-MS for detection of gel-separated phosphoproteins. *J Proteome Res* 2005, *4*, 515-522.

[91] Kim, J. E., Tannenbaum, S. R., White, F. M., Global phosphoproteome of HT-29 human colon adenocarcinoma cells. *J Proteome Res* 2005, *4*, 1339-1346.

[92] Ficarro, S. B., McCleland, M. L., Stukenberg, P. T., Burke, D. J.*, et al.*, Phosphoproteome analysis by mass spectrometry and its application to Saccharomyces cerevisiae. *Nat Biotechnol* 2002, *20*, 301-305.

[93] Luhrmann, R., Appel, B., Bringmann, P., Rinke, J.*, et al.*, Isolation and characterization of

rabbit anti-m3 2,2,7G antibodies. *Nucleic Acids Res* 1982, *10*, 7103-7113.

[94] Bringmann, P., Luhrmann, R., Purification of the individual snRNPs U1, U2, U5 and U4/U6 from HeLa cells and characterization of their protein constituents. *EMBO J* 1986, *5*, 3509-3516.

[95] Behrens, S. E., Luhrmann, R., Immunoaffinity purification of a [U4/U6.U5] tri-snRNP from human cells. *Genes Dev* 1991, *5*, 1439-1452.

[96] Bessonov, S., Anokhina, M., Will, C. L., Urlaub, H., Luhrmann, R., Isolation of an active step I spliceosome and composition of its RNP core. *Nature* 2008, *452*, 846-850.

[97] Deckert, J., Hartmuth, K., Boehringer, D., Behzadnia, N.*, et al.*, Protein composition and electron microscopy structure of affinity-purified human spliceosomal B complexes isolated under physiological conditions. *Mol Cell Biol* 2006, *26*, 5528-5543.

[98] Zahler, A. M., Purification of SR protein splicing factors. *Methods Mol Biol* 1999, *118*, 419-432.

[99] Lee, C. L., Hsiao, H. H., Lin, C. W., Wu, S. P.*, et al.*, Strategic shotgun proteomics approach for efficient construction of an expression map of targeted protein families in hepatoma cell lines. *Proteomics* 2003, *3*, 2472-2486.

[100] Schwartz, D., Gygi, S. P., An iterative statistical approach to the identification of protein phosphorylation motifs from large-scale data sets. *Nat Biotechnol* 2005, *23*, 1391-1398.

[101] Schneider, T. D., Stephens, R. M., Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res* 1990, *18*, 6097-6100.

[102] Mazanek, M., Mituloviae, G., Herzog, F., Stingl, C.*, et al.*, Titanium dioxide as a chemo-affinity solid phase in offline phosphopeptide chromatography prior to HPLC-MS/MS analysis. *Nat Protoc* 2007, *2*, 1059-1069.

[103] Thingholm, T. E., Jorgensen, T. J., Jensen, O. N., Larsen, M. R., Highly selective enrichment of phosphorylated peptides using titanium dioxide. *Nat Protoc* 2006, *1*, 1929-1935.

[104] Stark, G. R., Stein, W. H., Moore, S., Reactions of Cyanate Present in Aqueous Urea with Amino Acids and Proteins. *J Biol Chem* 1960, *235*, 3177-3181.

[105] Stark, G. R., Reactions of Cyanate with Functional Groups of Proteins .3. Reactions with Amino and Carboxyl Groups. *Biochemistry* 1965, *4*, 1030-&.

[106] Stark, G. R., Modification of proteins with cyanate. *Methods Enzymol* 1967, *11*, 590-594.

[107] Yu, Y. Q., Gilar, M., Lee, P. J., Bouvier, E. S., Gebler, J. C., Enzyme-friendly, mass spectrometry-compatible surfactant for in-solution enzymatic digestion of proteins. *Anal Chem* 2003, *75*, 6023-6028.

[108] Ishihama, Y., Rappsilber, J., Mann, M., Modular stop and go extraction tips with stacked disks for parallel and multidimensional Peptide fractionation in proteomics. *J Proteome Res* 2006, *5*, 988-994.

[109] Rappsilber, J., Ishihama, Y., Mann, M., Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics.

Anal Chem 2003, *75*, 663-670.

[110] Yu, S. Y., Wu, S. W., Hsiao, H. H., Khoo, K. H., Enabling techniques and strategic workflow for sulfoglycomics based on mass spectrometry mapping and sequencing of permethylated sulfated glycans. *Glycobiology* 2009, *19*, 1136-1149.

[111] Jensen, S. S., Larsen, M. R., Evaluation of the impact of some experimental procedures on different phosphopeptide enrichment techniques. *Rapid Commun Mass Sp* 2007, *21*, 3635-3645.

[112] Stensballe, A., Jensen, O. N., Phosphoric acid enhances the performance of Fe(III) affinity chromatography and matrix-assisted laser desorption/ionization tandem mass spectrometry for recovery, detection and sequencing of phosphopeptides. *Rapid Commun Mass Spectrom* 2004, *18*, 1721-1730.

[113] Chen, H. S., Rejtar, T., Andreev, V., Moskovets, E., Karger, B. L., Enhanced characterization of complex proteomic samples using LC-MALDI MS/MS: exclusion of redundant peptides from MS/MS analysis in replicate runs. *Anal Chem* 2005, *77*, 7816-7825.

[114] Gnad, F., Ren, S., Cox, J., Olsen, J. V*., et al.*, PHOSIDA (phosphorylation site database): management, structural and evolutionary investigation, and prediction of phosphosites. *Genome Biol* 2007, *8*, R250.

[115] Manning, B. D., Cantley, L. C., Hitting the target: emerging technologies in the search for kinase substrates. *Sci STKE* 2002, *2002*, pe49.

[116] Manley, J. L., Tacke, R., SR proteins and splicing control. *Genes Dev* 1996, *10*, 1569-1579.

[117] Sanford, J. R., Longman, D., Caceres, J. F., Multiple roles of the SR protein family in splicing regulation. *Prog Mol Subcell Biol* 2003, *31*, 33-58.

[118] Xiao, S. H., Manley, J. L., Phosphorylation of the ASF/SF2 RS domain affects both protein-protein and protein-RNA interactions and is necessary for splicing. *Genes Dev* 1997, *11*, 334-344.

[119] Pearson, R. B., Kemp, B. E., Protein kinase phosphorylation site sequences and consensus specificity motifs: tabulations. *Methods Enzymol* 1991, *200*, 62-81.

[120] Zhou, S. Y., Blechner, S., Hoagland, N., Hoekstra, M. F*., et al.*, Use of an Oriented Peptide Library to Determine the Optimal Substrates of Protein-Kinases. *Curr Biol* 1994, *4*, 973-982.

[121] Schneider, M., Hsiao, H. H., Will, C. L., Giet, R*., et al.*, Human PRP4 kinase is required for stable tri-snRNP association during spliceosomal B complex formation. *Nature Structural & Molecular Biology* 2010, *17*, 216-U212.

[122] Oellerich, T., Gronborg, M., Neumann, K., Hsiao, H. H*., et al.*, SLP-65 phosphorylation dynamics reveals a functional basis for signal integration by receptor-proximal adaptor proteins. *Mol Cell Proteomics* 2009, *8*, 1738-1750.

[123] Luo, X., Hsiao, H. H., Bubunenko, M., Weber, G*., et al.*, Structural and functional analysis

of the E. coli NusB-S10 transcription antitermination complex. *Mol Cell* 2008, *32*, 791-802.

[124] Kuhn-Holsken, E., Lenz, C., Sander, B., Luhrmann, R., Urlaub, H., Complete MALDI-ToF MS analysis of cross-linked peptide-RNA oligonucleotides derived from nonlabeled UV-irradiated ribonucleoprotein particles. *RNA* 2005, *11*, 1915-1930.

[125] Manning, G., Whyte, D. B., Martinez, R., Hunter, T., Sudarsanam, S., The protein kinase complement of the human genome. *Science* 2002, *298*, 1912-1934.

[126] Gronborg, M., Kristiansen, T. Z., Stensballe, A., Andersen, J. S*., et al.*, A mass spectrometry-based proteomic approach for identification of serine/threonine-phosphorylated proteins by enrichment with phospho-specific antibodies: identification of a novel protein, Frigg, as a protein kinase A substrate. *Mol Cell Proteomics* 2002, *1*, 517-527.

[127] Kruger, M., Kratchmarova, I., Blagoev, B., Tseng, Y. H*., et al.*, Dissection of the insulin signaling pathway via quantitative phosphoproteomics. *Proc Natl Acad Sci U S A* 2008, *105*, 2451-2456.

[128] Wang, Y., Li, R., Du, D., Zhang, C*., et al.*, Proteomic analysis reveals novel molecules involved in insulin signaling pathway. *J Proteome Res* 2006, *5*, 846-855.

[129] Zhang, G., Neubert, T. A., Use of detergents to increase selectivity of immunoprecipitation of tyrosine phosphorylated peptides prior to identification by MALDI quadrupole-TOF MS. *Proteomics* 2006, *6*, 571-578.

[130] Tong, J., Taylor, P., Jovceva, E., St-Germain, J. R*., et al.*, Tandem immunoprecipitation of phosphotyrosine-mass spectrometry (TIPY-MS) indicates C19ORF19 becomes tyrosine-phosphorylated and associated with activated epidermal growth factor receptor. *J Proteome Res* 2008, *7*, 1067-1077.

[131] Xia, Q., Cheng, D., Duong, D. M., Gearing, M*., et al.*, Phosphoproteomic analysis of human brain by calcium phosphate precipitation and mass spectrometry. *J Proteome Res* 2008, *7*, 2845-2851.

[132] Schroeder, M. J., Shabanowitz, J., Schwartz, J. C., Hunt, D. F., Coon, J. J., A neutral loss activation method for improved phosphopeptide sequence analysis by quadrupole ion trap mass spectrometry. *Anal Chem* 2004, *76*, 3590-3598.

[133] Bateman, R. H., Carruthers, R., Hoyes, J. B., Jones, C*., et al.*, A novel precursor ion discovery method on a hybrid quadrupole orthogonal acceleration time-of-flight (Q-TOF) mass spectrometer for studying protein phosphorylation. *J Am Soc Mass Spectrom* 2002, *13*, 792-803.

[134] Leimgruber, R. M., Malone, J. P., Radabaugh, M. R., LaPorte, M. L*., et al.*, Development of improved cell lysis, solubilization and imaging approaches for proteomic analyses. *Proteomics* 2002, *2*, 135-144.

[135] Nielsen, P. A., Olsen, J. V., Podtelejnikov, A. V., Andersen, J. R*., et al.*, Proteomic mapping of brain plasma membrane proteins. *Mol Cell Proteomics* 2005, *4*, 402-408.

[136] Chen, E. I., Cociorva, D., Norris, J. L., Yates, J. R., 3rd, Optimization of mass

spectrometry-compatible surfactants for shotgun proteomics. *J Proteome Res* 2007, *6*, 2529-2538.

[137] Tazi, J., Kornstadt, U., Rossi, F., Jeanteur, P.*, et al.*, Thiophosphorylation of U1-70K protein inhibits pre-mRNA splicing. *Nature* 1993, *363*, 283-286.

[138] Woppmann, A., Patschinsky, T., Bringmann, P., Godt, F., Luhrmann, R., Characterisation of human and murine snRNP proteins by two-dimensional gel electrophoresis and phosphopeptide analysis of U1-specific 70K protein variants. *Nucleic Acids Res* 1990, *18*, 4427-4438.

[139] Kohtz, J. D., Jamison, S. F., Will, C. L., Zuo, P.*, et al.*, Protein-protein interactions and 5'-splice-site recognition in mammalian mRNA precursors. *Nature* 1994, *368*, 119-124.

[140] Raker, V. A., Plessel, G., Luhrmann, R., The snRNP core assembly pathway: identification of stable core protein heteromeric complexes and an snRNP subcore particle in vitro. *EMBO J* 1996, *15*, 2256-2269.

[141] Varki, A., Biological roles of oligosaccharides: all of the theories are correct. *Glycobiology* 1993, *3*, 97-130.

[142] Van den Steen, P., Rudd, P. M., Dwek, R. A., Opdenakker, G., Concepts and principles of O-linked glycosylation. *Crit Rev Biochem Mol Biol* 1998, *33*, 151-208.

[143] Calvete, J. J., Sanz, L., Analysis of O-glycosylation. *Methods Mol Biol* 2008, *446*, 281-292.

[144] Marshall, R. D., The nature and metabolism of the carbohydrate-peptide linkages of glycoproteins. *Biochem Soc Symp* 1974, 17-26.

[145] Gavel, Y., von Heijne, G., Sequence differences between glycosylated and non-glycosylated Asn-X-Thr/Ser acceptor sites: implications for protein engineering. *Protein Eng* 1990, *3*, 433-442.

[146] Hubbard, S. C., Ivatt, R. J., Synthesis and processing of asparagine-linked oligosaccharides. *Annu Rev Biochem* 1981, *50*, 555-583.

[147] Kornfeld, R., Kornfeld, S., Assembly of asparagine-linked oligosaccharides. *Annu Rev Biochem* 1985, *54*, 631-664.

[148] Steen, H., Kuster, B., Fernandez, M., Pandey, A., Mann, M., Detection of tyrosine phosphorylated peptides by precursor ion scanning quadrupole TOF mass spectrometry in positive ion mode. *Anal Chem* 2001, *73*, 1440-1448.

[149] 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.

[150] Schlosser, A., Pipkorn, R., Bossemeyer, D., Lehmann, W. D., Analysis of protein phosphorylation by a combination of elastase digestion and neutral loss tandem mass spectrometry. *Anal Chem* 2001, *73*, 170-176.

[151] Covey, T., Shushan, B., Bonner, R., Schröder, W., Hucho, F., In: H. Jornvall, J.-O. Hoog

and A.-M. Gustavsson, Editors, *Methods in protein sequence analysis*, Birkhäuser Verlag, Basel ; Boston 1991, 249-256.

[152] Huddleston, M. J., Bean, M. F., Carr, S. A., Collisional fragmentation of glycopeptides by electrospray ionization LC/MS and LC/MS/MS: methods for selective detection of glycopeptides in protein digests. *Anal Chem* 1993, *65*, 877-884.

[153] Palumbo, A. M., Reid, G. E., Evaluation of gas-phase rearrangement and competing fragmentation reactions on protein phosphorylation site assignment using collision induced dissociation-MS/MS and MS3. *Anal Chem* 2008, *80*, 9735-9747.

[154] Palumbo, A. M., Tepe, J. J., Reid, G. E., Mechanistic insights into the multistage gas-phase fragmentation behavior of phosphoserine- and phosphothreonine-containing peptides. *J Proteome Res* 2008, *7*, 771-779.

[155] Sullivan, B., Addona, T. A., Carr, S. A., Selective detection of glycopeptides on ion trap mass spectrometers. *Anal Chem* 2004, *76*, 3112-3118.

[156] Rosner, M. R., Robbins, P. W., Separation of glycopeptides by high performance liquid chromatography. *J Cell Biochem* 1982, *18*, 37-47.

[157] Geiss-Friedlander, R., Melchior, F., Concepts in sumoylation: a decade on. *Nat Rev Mol Cell Biol* 2007, *8*, 947-956.

[158] Kerscher, O., Felberbaum, R., Hochstrasser, M., Modification of proteins by ubiquitin and ubiquitin-like proteins. *Annu Rev Cell Dev Biol* 2006, *22*, 159-180.

[159] Meulmeester, E., Melchior, F., Cell biology: SUMO. *Nature* 2008, *452*, 709-711.

[160] Hershko, A., Ciechanover, A., The ubiquitin system. *Annu Rev Biochem* 1998, *67*, 425-479.

[161] Johnson, E. S., Protein modification by SUMO. *Annu Rev Biochem* 2004, *73*, 355-382.

[162] Verger, A., Perdomo, J., Crossley, M., Modification with SUMO. A role in transcriptional regulation. *EMBO Rep* 2003, *4*, 137-142.

[163] Bernier-Villamor, V., Sampson, D. A., Matunis, M. J., Lima, C. D., Structural basis for E2-mediated SUMO conjugation revealed by a complex between ubiquitin-conjugating enzyme Ubc9 and RanGAP1. *Cell* 2002, *108*, 345-356.

[164] Melchior, F., SUMO--nonclassical ubiquitin. *Annu Rev Cell Dev Biol* 2000, *16*, 591-626.

[165] Sampson, D. A., Wang, M., Matunis, M. J., The small ubiquitin-like modifier-1 (SUMO-1) consensus sequence mediates Ubc9 binding and is essential for SUMO-1 modification. *J Biol Chem* 2001, *276*, 21664-21669.

[166] Lin, D., Tatham, M. H., Yu, B., Kim, S.*, et al.*, Identification of a substrate recognition site on Ubc9. *J Biol Chem* 2002, *277*, 21740-21748.

[167] Melchior, F., Schergaut, M., Pichler, A., SUMO: ligases, isopeptidases and nuclear pores. *Trends Biochem Sci* 2003, *28*, 612-618.

[168] Hay, R. T., SUMO-specific proteases: a twist in the tail. *Trends Cell Biol* 2007, *17*, 370-376.

[169] Mukhopadhyay, D., Dasso, M., Modification in reverse: the SUMO proteases. *Trends Biochem Sci* 2007, *32*, 286-295.

[170] Mossessova, E., Lima, C. D., Ulp1-SUMO crystal structure and genetic analysis reveal conserved interactions and a regulatory element essential for cell growth in yeast. *Mol Cell* 2000, *5*, 865-876.

[171] Steinacher, R., Schar, P., Functionality of human thymine DNA glycosylase requires SUMO-regulated changes in protein conformation. *Curr Biol* 2005, *15*, 616-623.

[172] Baba, D., Maita, N., Jee, J. G., Uchimura, Y*., et al.*, Crystal structure of thymine DNA glycosylase conjugated to SUMO-1. *Nature* 2005, *435*, 979-982.

[173] 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.

[174] Pichler, A., Knipscheer, P., Oberhofer, E., van Dijk, W. J*., et al.*, SUMO modification of the ubiquitin-conjugating enzyme E2-25K. *Nat Struct Mol Biol* 2005, *12*, 264-269.

[175] Lin, D. Y., Huang, Y. S., Jeng, J. C., Kuo, H. Y*., et al.*, Role of SUMO-interacting motif in Daxx SUMO modification, subnuclear localization, and repression of sumoylated transcription factors. *Mol Cell* 2006, *24*, 341-354.

[176] Meulmeester, E., Kunze, M., Hsiao, H. H., Urlaub, H., Melchior, F., Mechanism and consequences for paralog-specific sumoylation of ubiquitin-specific protease 25. *Mol Cell* 2008, *30*, 610-619.

[177] Denison, C., Rudner, A. D., Gerber, S. A., Bakalarski, C. E*., et al.*, A proteomic strategy for gaining insights into protein sumoylation in yeast. *Mol Cell Proteomics* 2005, *4*, 246-254.

[178] Vertegaal, A. C., Andersen, J. S., Ogg, S. C., Hay, R. T*., et al.*, Distinct and overlapping sets of SUMO-1 and SUMO-2 target proteins revealed by quantitative proteomics. *Mol Cell Proteomics* 2006, *5*, 2298-2310.

[179] Hannich, J. T., Lewis, A., Kroetz, M. B., Li, S. J*., et al.*, Defining the SUMO-modified proteome by multiple approaches in Saccharomyces cerevisiae. *J Biol Chem* 2005, *280*, 4102-4110.

[180] Matic, I., van Hagen, M., Schimmel, J., Macek, B*., et al.*, In vivo identification of human small ubiquitin-like modifier polymerization sites by high accuracy mass spectrometry and an in vitro to in vivo strategy. *Mol Cell Proteomics* 2008, *7*, 132-144.

[181] Knuesel, M., Cheung, H. T., Hamady, M., Barthel, K. K., Liu, X., A method of mapping protein sumoylation sites by mass spectrometry using a modified small ubiquitin-like modifier 1 (SUMO-1) and a computational program. *Mol Cell Proteomics* 2005, *4*, 1626-1636.

[182] Pedrioli, P. G., Raught, B., Zhang, X. D., Rogers, R*., et al.*, Automated identification of SUMOylation sites using mass spectrometry and SUMmOn pattern recognition software. *Nat Methods* 2006, *3*, 533-539.

[183] Matunis, M. J., Coutavas, E., Blobel, G., A novel ubiquitin-like modification modulates the

partitioning of the Ran-GTPase-activating protein RanGAP1 between the cytosol and the nuclear pore complex. *J Cell Biol* 1996, *135*, 1457-1470.

[184] Bossis, G., Melchior, F., Regulation of SUMOylation by reversible oxidation of SUMO conjugating enzymes. *Mol Cell* 2006, *21*, 349-357.

[185] Pichler, A., Gast, A., Seeler, J. S., Dejean, A., Melchior, F., The nucleoporin RanBP2 has SUMO1 E3 ligase activity. *Cell* 2002, *108*, 109-120.

[186] Werner, A., Moutty, M. C., Moller, U., Melchior, F., Performing in vitro sumoylation reactions using recombinant enzymes. *Methods Mol Biol* 2009, *497*, 187-199.

[187] Mahajan, R., Delphin, C., Guan, T., Gerace, L., Melchior, F., A small ubiquitin-related polypeptide involved in targeting RanGAP1 to nuclear pore complex protein RanBP2. *Cell* 1997, *88*, 97-107.

[188] Maiolica, A., Cittaro, D., Borsotti, D., Sennels, L.*, et al.*, Structural analysis of multiprotein complexes by cross-linking, mass spectrometry, and database searching. *Mol Cell Proteomics* 2007, *6*, 2200-2211.

[189] Olsen, J. V., de Godoy, L. M., Li, G., Macek, B.*, et al.*, Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. *Mol Cell Proteomics* 2005, *4*, 2010-2021.

[190] Olsen, J. V., Macek, B., Lange, O., Makarov, A.*, et al.*, Higher-energy C-trap dissociation for peptide modification analysis. *Nat Methods* 2007, *4*, 709-712.

[191] Mahajan, R., Gerace, L., Melchior, F., Molecular characterization of the SUMO-1 modification of RanGAP1 and its role in nuclear envelope association. *J Cell Biol* 1998, *140*, 259-270.

[192] Rodriguez, M. S., Desterro, J. M., Lain, S., Midgley, C. A.*, et al.*, SUMO-1 modification activates the transcriptional response of p53. *EMBO J* 1999, *18*, 6455-6461.

[193] Sternsdorf, T., Jensen, K., Reich, B., Will, H., The nuclear dot protein sp100, characterization of domains necessary for dimerization, subcellular localization, and modification by small ubiquitin-like modifiers. *J Biol Chem* 1999, *274*, 12555-12566.

[194] Villen, J., Gygi, S. P., The SCX/IMAC enrichment approach for global phosphorylation analysis by mass spectrometry. *Nat Protoc* 2008, *3*, 1630-1638.

Appendixes

Appendix 1. Phosphorylation sites identified from spliceosomal proteins by using in-house TiO_2 microspin column enrichment. The underline indicates that the phosphorylation site was not identified previously.

Protein	Accession Number	Phosphorylation Sites	Unassigned phosphorylation sites
Sm proteins			
D2	gi|4759158	S9, T12, T40	
G	gi|4507133	T50	
D3	gi|4759160	S44	
E	gi|119589572	S89	
D1	gi|5902102	T57	
U1 snRNP			
U1 70K	gi|36100	S294, S403, S443, S445, S458, S470, S472, S497, S562, S587, Y303	
U1-C	gi|4507127	S17	
U1 snRNP associated			
RBM25	gi|887360	S107, S133, T154, T155	
17S U2 snRNP			
U2A'	gi|50593002	T180, S197	^{178}SKTFNPGAGLPTDKK^{192}+P, ^{236}SGPTDDGEEEMEEDTVTNGS^{255}+P
U2A'	gi|89063185		^{141}SGPTGDGEEEMEEDTVTNGS^{160}+P
SF3b145	gi|2498883	S266, S280, S284, S286, S339, S408, S412, S413, T288, T757	^{336}GSDSPAADVEIEYVTEEPEIYEPNFIFFKR^{365}+P
SF3a120	gi|5032087	S329, S359	
SF3b155	gi|15214275	S73, S129, S190, S194, S229, S322, S488, T142, T207, T211, T223, T227, T235, T244, T248, T257, T261, T267, T273, T279, T296, T303, T313, T326, T328, T426, T434, T436, T442	^{239}AKGSETPGATPGSK^{252}+P, ^{214}KLSSWDQAETPGHTPSLR^{231}+3P, ^{431}LTATPTPLGGMTGFHMQTEDR^{451}+3P,
SF3a60	gi|551450	S295	
SF3b14a/p14	gi|7706326	T121	
17S U2 related			
U2AF35	gi|4827046	S349, S384	

SF3b125	gi|45446743	S96, S185	
SPF45	gi|14249678	S155, S222	
hPRP5	gi|5410326	S803	
U2AF65	gi|6005926	S79	
SR140	gi|2224605	S484	
CHERP	gi|2058691	S781, S783, S785, T787	^{783}SRSPTPPSSAGLGSNSAPPIPDSR^{806}+3P
U5 snRNP			
200K	gi|40217847	S8, S17, S26, S42, S207, S225, S756, S1315, S2002, S2133, S2135, T393, T1572, T2131	^{187}EIQNMDDNIDETYGVNVQFESDEEEGDEDVYGEVR^{211}+2P
52K	gi|5174409	S49, S118, S195, T341	^{137}QASDSEEEDSLGQTSMSAQALLEGLLELLLPR^{168}+2P, ^{231}GLGCQTLGPHNPTPPPSLDMFAEELAEEELETPTPTQR^{268}+P,
116K	gi|41152056	S19, T86, S466, T478, S944	^{1}MDTDLYDEFGNYIGPELDSDEDDDELGR^{28}+2P, ^{1}MDTDLYDEFGNYIGPELDSDEDDDELGRETK^{31}+2P
100K	gi|67460585	S14, S16, S23, S39, S63, S65, S67, S107, S109, T25	^{106}SSLSPGRGK^{114}+2P
102K	gi|40807485	S143, T232, T235, S261, S263, T266	
102K	gi|4103604		^{204}QTQFGGLNTPYPGGLNTPYRGGMTPGLMTPGTGELDMR^{241}+P
220K	gi|73967172	S1358, S1411, S2079, T2042, T2044	^{2171}EMEPLGWIHTQPNESPQLSPQDVTTHAK^{2193}+P
U4/U6 snRNP			
90K	gi|2853287	S618	
61K	gi|4914604	S439, T441, T455	^{489}VKGEKSGLMST^{499}+P, ^{445}SSGTASSVAFTPLQGLEIVNPQAAEK^{470}+3P
60K	gi|2653736	S482	
U4/U6.U5 Tri-snRNP			
110K	gi|10863889	S84, S448, S463, S474, S486, S591, S596, S598, S623, T430, T764	^{587}DEERSANGGSESDGEENIGWSTVNLDEEK^{615}+2P, ^{591}SANGGSESDGEENIGWSTVNLDEEKQQQDFSASSTTILDEEPIVNR^{636}+P, ^{445}EPVPQPLPSDDTRVENMDISDEEEGGAPPPGSPQVLEEDEAELELQK^{501}+2P
TFIP11	gi|8393259	S59, S98, S210, S392	
65K	gi|13926071	S46, S58, S82, S97	
hPRP38	gi|14042816	S117, S118, S133, S150, S220	
27K	gi|24307919	S43, S45, S61, S63	59STSPSPSR66+3P, ^{1}MGRSRSRSPR^{11}+P

U11/U12			
65K	gi|14017895	S45, S132, S405	^{126}EQDRVHSPCPTSGSEKK^{142}+P
31K	gi|51243065	S155, S210, S216	
48K	gi|48427636	S207, S286	
SR proteins			
SC35	gi|337926	S26, S147, S149, S187, S189, S191, S206, S208, S212, S220, S221	^{25}TSPDTLRR^{32}+P, ^{185}SRSRSRSPPPVSK^{197}+3P, ^{142}SRYSRSKSR^{150}+3P,
ZNF265	gi|24980810	S120, S153, S188, S305, S307, T303	^{303}TRSRSPESQVIGENTK^{318}+2P
KIAA0853	gi|30908950	S77, S110, S198, S207, S209, S242, S265, S356, S387, S875, S877, S886, S993, S1208, S1210, S1366, S1423, S1438, T263, T354, T1033	^{1227}VLHSGSR^{1233}+P, ^{619}DSSFERR^{625}+P, ^{1277}RSSPESDR^{1284}+P, ^{638}DQRPSSPIR^{646}+P, ^{104}NTEESSSPVR^{113}+P, ^{370}SASPYPSHSLSSPQR^{384}+3P
SFRS1	gi|119614893	S181, S183, S187, S216, S220, S224, Y184	213SRGSPRYSPR222+2P
SFRS16	gi|3941326	S86, S270, S279, S320, S439, S441, S484, S486, S488	
RSRC1	gi|14714462	S6, S179	
SFRS18	gi|14042500	S211	^{286}SKFDSDEEEEDTENVEAASSGK^{307}+P
SRp55	gi|1049088	S45, S118, S119, S265, S272, S295, S297, S299, S301, S303, S314, S316	^{314}SVSPPPK^{320}+P, ^{297}SQSRSNSPLPVPPSK^{311}+P, ^{297}SQSRSNSPLPVPPSK^{311}+3P, ^{295}SRSQSRSNSPLPVPPSK^{311}+4P
LUC7L	gi|41393554	S363	
RSRC2	gi|38146107	S17, S27, S30, S32, S216, S218, S222, T220	^{216}SLSRTPSPPPFR^{227}+2P, ^{2}AASDTERDGLAPEKTSPDRDK^{22}+2P
SRp38	gi|16905517	S131, S133	^{140}RSYSPR^{145}+2P, ^{251}SRSWTSPK^{258}+2P,
SRp30c	gi|4506903	S189, S193, S204, S211, S216	
SFRS11	gi|34364678	S211, S322, S337, T213	
TRA-2 alpha	gi|9558733	S16, S18, S20, S46, S84, S86, S260, S262, T88	
hTra-2 beta/SFRS10	gi|4759098	S20, S22, S26, S29, S37, S39, S95, S97, S99, T33	
SFRS8	gi|3929376	S283	
9G8	gi|72534660	S163, S165, S167, S171, S173, S175, S181, S183, S192, S194, S196, S200, S202, S204, S208,	

		S209, S215, S231, S233	
SRP20	gi|4506901	S108, S138, S140, S158, S160	
SRP75	gi|21361282	S78, S269, S288, S290, S316, S431, S456, S458, S460, S482, S488, T468	^{482}SPSRSRSR^{489}+3P
CRKRS	gi|7706549	S249, S1083, T692, T893	
SRP40	gi|55640963	S212, S214, S228, S230, S232, S243, S244, S245, S247, S249, S252	^{241}RGSSSRSKSPASVDR^{255}+4P
SRP46	gi|118137604		^{25}TSPDSLRR^{32}+P
SRp25	gi|7619898	S129	
SR protein	gi|3253165	S27, S155, T161, S251	
rsr-1	gi|88952952	S95, S96	
SFRS12	gi|21040255		^{403}SSRSPSPRSR^{412}+2P
LUC7L2	gi|7022826	S281, S283, S336, S383, S384	
SR related proteins			
SRm300	gi|118572613	S295, S297, S323, S351, S353, S377, S387, S398, S404, S408, S440, S456, S484, S508, S510, S534, S536, S575, S637, S639, S702, S704, S706, S774, S778, S780, S790, S808, S820, S857, S871, S875, S876, S895, S908, S937, S952, S954, S974, S994, S1014, S1028, S1054, S1072, S1073, S1083, S1101, S1102, S1103, S1112, S1124, S1132, S1181, S1227, S1233, S1318, S1320, S1326, S1329, S1336, S1368, S1382, S1401, S1404, S1441, S1444, S1463, S1497, S1499, S1502, S1537, S1539, S1541, S1542, S1552, S1592, S1600, S1601, S1616, S1618, S1657, S1691, S1693, S1694, S1729, S1731, S1842, S1854, S1857, S1864, S1866, S1869, S1876, S1878, S1890, S1893,	^{1852}TSPAPWK^{1862}+P, ^{1693}SSPELTR^{1699}+P, ^{2044}SRSPLAIR^{2051}+P, ^{1854}SRTSPAPWK^{1862}+P, ^{1457}HSLSGSSPGMK^{1467}+P, ^{778}SLSGSSPCPK^{787}+2P, ^{1690}SSRSSPELTR^{1699}+2P, ^{2088}SRSATPPATR^{2097}+2P, ^{2407}SRTPPSAPSQSR^{2418}+P, ^{1497}SRSPSSPELNNK^{1508}+3P, ^{743}SNSSPEMKKSRISSR^{757}+2P, ^{1001}SSSPVTELASRSPIR^{1115}+P, ^{357}SSTGPEPPAPTPLLAER^{373}+2P, ^{1245}SEEPAGQILSHLSSELK^{1262}+P, ^{2098}NHSGSRTPPVALNSSR^{2113}+P, ^{433}HASSSPESPKPAPAPGSHR^{451}+3P, ^{1539}SGSSQELDVKPSASPQER^{1556}+3P, ^{1449}DGSGTPSRHSLSGSSPGMK^{1467}+3P, ^{979}VKPETPPRQSHSGSISPYPK^{998}+2P, ^{852}QGSITSPQANEQSVTPQRR^{870}+2P, ^{1138}FQSDSSSYPTVDSNSLLGQSR^{1158}+P, ^{304}GEGDAPFSEPGTTSTQRPSSPETATK^{329}+2P, ^{346}SATRPSPSPERSSTGPEPPAPTPLLAER^{373}+2P, ^{374}HGGSPQPLATTPLSQEPVNPPSEASPTR^{401}+4P,

		S1913, S1916, S1923, S1925, S1972, S1975, S1984, S1987, S2020, S2032, S2044, S2046, S2067, S2071, S2090, S2100, S2102, S2132, S2272, S2310, S2365, S2368, S2382, S2398, S2407, S2449, S2581, S2675, S2677, S2684, S2688, S2690, S2692, S2694, S2702, S2706, T326, T364, T384, T400, T577, T792, T810, T866, T903, T983, T1003, T1043, T1208, T1413, T1434, T1472, T1492, T1511, T1531, T1569, T1844, T1856, T1880, T1892, T1927, T1986, T2022, T2034, T2069, T2092, T2104, T2289, T2302, T2316, T2329, T2381, T2397, T2409, T2583	
SRm160	gi|23274133	S260, S306, S308, S310, S389, S391, S393, S402, S414, S429, S431, S450, S452, S463, S465, S478, S549, S551, S560, S562, S583, S590, S592, S597, S605, S607, S616, S626, S628, S636, S638, S652, S654, S683, S685, S695, S696, S713, S715, S725, S738, S740, S756, S769, S775, S797, S874, T220, T327, T406, T416, T572, T574, T581, T614, T727, T846, T872	^{763}KPPAPPSPVQSQSPSTNWSPAVPVK^{787}+2P
SRm160	gi|119615546	S538	
SFRS2IP	gi|74717376	S293, S481, T287	
hnRNP			
hnRNP H1	gi|5031753	S104, S310	
hnRNP A1	gi|36102	S6	
hnRNP U	gi|32358	S252, S668	
hnRNP F	gi|4826760	S310	

hnRNP G	gi|3256007	S208, S352	^{299}GPPPSYGGSSR^{309}+P
Lsm proteins			
LSM11	gi|27735089	S15, S21, S28	
Lsm3	gi|7657315		^{2}ADDVDQQQTTNTVEEPLDLIR^{22}+P
hPrp19/CDC5L complex			
CDC5L	gi|11067747	S64, S303, S317, S393, S437, S463, T61, T355, T373, T377, T385, T396, T404, T424, T430, T438, T442, T510	^{427}SGTTPKPVINSTPGR^{441}+2P, ^{400}QVVQTPNTVLSTPFR^{414}+2P, ^{427}SGTTPKPVINSTPGRTPLR^{445}+2P, ^{362}ILQEAQNLMALTNVDTPLK^{380}+P, ^{427}SGTTPKPVINSTPGRTPLR^{445}+3P, ^{427}SGTTPKPVINSTPGRTPLR^{445}+4P, ^{427}SGTTPKPVINSTPGRTPLR^{445}+5P
CCAP2	gi|6841518	T110, S121, T173	
PRCC	gi|14714625	S157, S159, S267	^{207}KPSDGSPDTKPSR^{219}+P
Npw38BP	gi|7706501	S353, S361, S364	
PRL1	gi|4505895		^{117}MPSESAAQSLAVALPLQTK^{135}+P
hPrp19/CDC5L related proteins			
SKIP	gi|6912676	S33, S182, S224, S232, S389, S415	^{401}TSNEVQYDQR^{410}+P
hSYF1	gi|10566459	S851	
CRNKL1	gi|11055967	S834	
Cyp-E	gi|73917051	T126	
hIsy1	gi|6330157	S267	
hRES complex proteins			
SNIP1	gi|10434110	S35, S52, S54, S74, S76, S153, S202, S394, S396	^{220}EKPSFELSGALLEDTNTFR^{240}+P
MGC13125	gi|14249338	S18, S57, S139, S151, S163, S175, S197, S201, S214, S222, S226, S235, S240, S248, S258, S271, S281, S325, S354, S357, S358, S364, S370, S372, S375, S385, S386, S388, S391, S401, S402, S404, S407, T135, T147, T159, T174	^{170}HDSDTSPPR^{178}+P, ^{121}HDTPDSSPR^{129}+P, ^{351}ATDSDLSSPR^{360}+2P,
CGI-79	gi|118600973	S188	^{178}EVQAEQPSSSSPR^{190}+P, ^{229}TAYSGGAEDLER^{240}+P, ^{272}SSDAHSSWYNGR^{283}+P

Proteins recruited to A-complex			
tat SF1	gi|21361437	S387, S579, S616, S624, S642, S676, S702	
RBM10	gi|1469167	S170, S803, S813, S816, S818, S877, S925, Y812	
Splicing factor HCC1	gi|4757926	S97, S106, S127, S129, S136, S337	
SF1	gi|785996	S80, S82	
RBM5/LUCA15	gi|1244404	S59, S621, S624, Y620	
Proteins recruited to B-complex			
PRPF4B	gi|23831382	S23, S32, S87, S93, S142, S144, S232, S239, S241, S257, S277, S301, S303, S328, S341, S354, S356, S366, S368, S387, S394, S396, S410, S411, S427, S431, S437, S443, S445, S458, S460, S518, S519, S520, S569, S578, S580, T576, Y849	^{408}RLSSPR^{413}+P, ^{332}DASSGKENRSPSR^{344}+P, ^{276}ARSPTDDKVKIEDK^{289}+P, ^{514}VEQESSSDDNLEDFDVEEEDEEALIEQR^{541}+2P
GCFC	gi|22035565	S16, S62, S155, S262, S295, S557, S558	
THRAP3	gi|114555524	S22, S28, S34, S36, S53, S55, S119, S139, S141, S143, S237, S240, S243, S248, S253, S315, S320, S379, S534, S535, S672, S682, S928, S939, T874	^{922}WAHDKFSGEEGEIEDDESGTENR^{944}+P, ^{803}EESTTGFDK^{811}+P, ^{203}EQTFSGGTSQDTK^{215}+P
MFAP1	gi|1709012	S52, S53, S94, S116, S118, S132, S133, T267	
HSPB1	gi|4504517	S82	
MATR3	gi|9956070	S158	
PRPF4B	gi|405749	S7, S9, Y19	
CCDC16	gi|14290546	S217	
KIAA1604	gi|10047283	S56, S57, S68, S90, S93, S120, S122, S815, S821, S858, S860	^{127}NPETSVTQSSSAQDEPATK^{145}+P, ^{55}NSSPEDRYEEQERSPR^{70}+3P, ^{433}EILDEGDTDSNTDQDAGSSEEDEEEEEEEGEEDEEGQK^{470}+4P
DHX16	gi|4503293	S103, S106, S107, S160	

FBP21	gi|6005948	S220, S262, S277	
RNF113A	gi|5902158	S84, S85, S253	
PPIL4	gi|20911035	S178	
RED	gi|5901878	S409	
CCDC12	gi|21389497	S149, S152	
PABPN1	gi|4758876		[79]APPGAPGPGPGSGAPGSQEEEEEPGLVEGDPGDGAIED PELEAIK[123]+P
Proteins recruited to C-complex			
PPWD1	gi|559713	S38	
Abstrakt	gi|21071032	S21, S23, S66, S405, S618, S619	
c19orf29	gi|122937392	S139, S476, T162	
C1orf55	gi|21751360	S278	[176]VVNTDHGSPEQLQIPVTDSGR[196]+P
A/B proteins			
DDX9	gi|1082769	S87	
NFAR	gi|9663121	S29	
NFAR	gi|1770458		[475]DSSKGEDSAEETEAKPAVVAPAPVVEAVSTPSAAFPSDAT AEQGPILTK[523]+P
YB-1	gi|27807361	S176, S314	[2]SSEAETQQPPAAPPAAPALSAADTKPGTTGSGAGSGGPG GLTSAAPAGGDK[52]+P
HNRPCL1	gi|14249959	S220, S240, S247, S286	
HNRPCL1	gi|13937888	S253, S260	
Potential C-complex specific proteins			
CDC2L1	gi|507160	S47, S232, S262, S268, S571, T577, S734	
TOE1 (FLJ13949)	gi|10436256	S5, S428	
PPIG	gi|62988846	S71, S73, S74, S76, S107, S173, S214, S504, S513, S533, S534, T175	[531]IRSSVEK[537]+P, [71]SASSESEAENLEAQPQSTVRPEEIPPIPENR[103]+3P
C9orf78	gi|7706557	S15, S17, S213	
NKAP	gi|47938198	S9, S149	[2]APVSGSRSPDREASGSGGR[20]+P
ZCCHC8	gi|7018505	S189, S411, T410	[353]SEAGHASSPDSEVTSLCQK[371]+P
DGCR14	gi|12804313	S291, S390, S394, T332, T338, T385, T449, T452	[429]TPASGLQTPTSTPAPGSATRTPLTQDPASITDNLLQLPAR[468]+4P
CCDC130	gi|13540614	S286, S306, S362	
ZCCHC10	gi|8923106	S68	

Step 2 factors			
hPRP17	gi|4102713	S40, S42	^{40}SPSSKPSLAVAVDSAPEVAVKEDLETGVHLDPAVK^{74}+P
hPRP22	gi|127797813	S129, S374, S377, S460	
hSLU7	gi|119581960	S225, S245, S476	
hPRP16	gi|3123906	S1198	
EJC/mRNP			
Acinus	gi|7513059	S155, S182, S267, S304, S323, S325, S327, S349, S429, S544, S594, S596, S649, S668, S754, S837, S943	^{762}KISVVSATK^{770}+P
Pinin	gi|3021392	S94, S341, S437	
RNPS1	gi|88973230	S251	
pre-mRNA/mRNA binding proteins			
LOC124245	gi|31377595	S13, S32, S34, S46, S53, S67, S74, S78, S83, S95, S118, S487, S532, S534, S536, S836, S842, S852, S857, S868, S893, T162, T677, T851	^{576}SSSYSSYSSR^{585}+2P, ^{109}TSDLRDEASSVTR^{121}+2P,
BCLAF1	gi|7661958	S25, S27, S29, S153, S177, S222, S268, S285, S290, S385, S397, S496, S512, S531, S648, S658, T402, T840, Y150	^{141}SSSSRSSSPYSK^{152}+P, ^{751}SSSSSASPSSPSSR^{764}+P, ^{205}SSATSGDIWPGLSAYDNSPR^{224}+2P,
ARS2	gi|46812675	S3, S10, T480	^{496}NITDYLIEEVSAEEEELLGSSGGAPPEEPPK^{526}+P
ARS2B	gi|13383501	S67	
DBPA	gi|16198465	S34	
DDX3	gi|2580550	S2	
RBM7	gi|7023641	S137	
CBP80	gi|4505343		^{21}TSDANETEDHLESLICK^{37}+P
ELG	gi|8923771	S135	
TREX			
THOC2	gi|52486999	S1278, S1302	
THOC1	gi|4826882	S560	
Nucleus (Location)			
COIL	gi|189514	S100, S109, T119, S316, S318, S395	^{391}LIIESPSNTSSTEPA^{405}+2P
NCL	gi|189306	S67, S145, S153, S184, S206,	^{177}AAAAAPASEDEDDEDDEDDEDDDDDEEDDSEEEAMET

		S563	TPAK[217]+2P
ZBTB7A	gi|7705375	S526, S549	[275]GGEEEAASLSEAAPEPGDSPGFLSGAAEGEDGDGPDVD GLAASTLLQQMMSSVGR[329]+P
SMN1	gi|736411	S27, S30	
EAPP	gi|73919272	S17, S24, S25, S26, S109, S111, S175	
DDX21	gi|2135315	S107, S139	
DKC1	gi|2737894	S494, S513	
DDX20	gi|5359631	S187, S532, S677, S678, S703, S714	
BCDIN3	gi|47271406	S60, S69, T75, S254, S330, S530	[519]KRSCFPASLTASR[531]+2P
CCNL1	gi|9945320	S65, S335, S342, S352, S504	[64]LSPTPSMQDGLDLPSETDLR[83]+P
NPM1	gi|825671	S56, S111	
SRPK1	gi|630737	S51, S311	[346]DTEGGAAEINCNGVIEVINYTQNSNNETLR[375]+P
SAFB	gi|1213639	T137, S293, S719	[122]ELPEQLQEHAIEDKETINNLDTSSSDFTILQEIEEPSLEPEN EK[165]+P
NOL5A	gi|119630989	S580, S629, S630	
NOL1	gi|12653741	S732, S786	
G3BP1	gi|5031703	S149, S232	
LARP7	gi|109809739	S261, S300, S337, T338	
NIPBL	gi|47458031	S915, S2658	
SPEN	gi|14790190	S309, S725, S727, S1425, S1918, S2120, S2412, T2421	
DDX51	gi|21756727	S83	
RBM15	gi|14041646	S51, S97, S128, S292, S294, S622, S656, S659, S670, S674, S708, S741, S765	[257]SRSPLDKDTYPPSASVVGASVGGHR[281]+2P
ARGLU1	gi|84000355	S56, S58, S60, S77	
SAP30BP	gi|9994179	S22	[6]NVLSSLAVYAEDSEPESDGEAGIEAVGSAAEEK[38]+2P
TAF15	gi|4507353		[219]TDADSESDNSDNNTIFVQGLGEGVSTDQVGEFFK[252]+P
UBR5	gi|15147337	S1549	
KPNA3	gi|1934907	S60	
NOLC1	gi|4758860	S698	
SSRP1	gi|4507241	S444	
HD	gi|4586876	S1261	
RBM14	gi|5454064	S147, T148, T206	[263]TQPMTAQAASYR[374]+P
SP1	gi|339518	S523	

ZC3H14	gi|28207879	S430	
TOP2A	gi|105857	S1376, S1392	
KIAA1429	gi|7243239	S171, S220, S1577	
MPHOSPH10	gi|2230873	S233	
CHD8	gi|34328020	T1189	
NEK2	gi|62898267	S354, Y355	
GEMIN8	gi|8923481		^{112}EDQALSKEEEMETESDAEVECDLSNMEITEELR^{144}+2P
MLF2	gi|4885487	S238	
NOC2L	gi|7512721	S54	
TSPYL2	gi|11545835	S204	
ZNF788	gi|34535580	T254	
PRKRIP1	gi|13375901		^{159}EQGSSSSAEASGTEEEEEVPSFTMGR^{184}+P
NPM1	gi|63101692	S42, S61	
SSB	gi|337457	S313	
ZBTB11	gi|7657703	Y319	
ZNF167	gi|10434411	T157, T161	
GEMIN5	gi|22001417	S778, Y416	
CROP	gi|7023491	S333, S335, S367, S425, S431	
YTHDC1	gi|16551831	S95, S257, S446, T97	^{66}LSSSASREPYK^{76}+P
CLK3	gi|4502885	S135	^{78}SPSFGEDYYGPSR^{88}+P
CLK1	gi|1705918	S140	^{323}VVDFGSATYDDEHHSTLVSTR^{343}+2P
JARID1D	gi|1871160	S114, S116	
RNF8	gi|34304336	S431	
NOL5	gi|6841462	S438	
ANP32E	gi|119573964	T243	
MGEA5	gi|10645186	S485	
BLM	gi|4557365	T321, S328	
OTUD7B	gi|9367763		^{30}STGAEPGLAR^{39}+P
SLC4A1AP	gi|8922557	S466	
GTF2B	gi|254934		^{49}TFSNDKATKDPSR^{61}+P
RBM15B	gi|54607124	S109, S265, S267, S552, S562, S609	
ZC3H7B	gi|119580830		^{64}QSPACPTLPCSSR^{76}+P
HOXB3	gi|123268	S181	
AHR	gi|6330736		^{476}NSPISHPPSPSPSAYSSR^{493}+3P
ZNF264	gi|4585643	T163	
RFC1	gi|296908	S69, S71	
NUP98	gi|4545099	S1312	

FIP1L1	gi|15079403	S226	
DEK	gi|4503249	S32, S227, S230, S231, S232, S243, S244, S251, S301, S303, S306, S307	^{299}KESESEDSSDDEPLIK^{314}+3P
PRPF40A	gi|5360087	S347, S397	^{385}DSGNWDTSGSELSEGELEK^{403}+2P
NUMA1	gi|35121	S1743	
WTAP	gi|10334526	S303	
PRPF38B	gi|8922358	S527, S529	
SMC4	gi|6807671	S41, S359	
MED24	gi|8699628	S862	
TRIM28	gi|5032179		^{33}STAPSAAASASASAAASSPAGGGAEALELLEHCGVCR^{69}+P
YY1	gi|186768	S118	
SUB1	gi|62088150	S16, S17, S18, S19, S20, S24, S125	^{13}ELVSSSSSGSDSDSEVDKK^{31}+6P
TCEA1	gi|313223	S100	
PDS5B	gi|7657269	S1358	
INTS1	gi|7243278		^{513}SSPEQPIGQGR^{523}+P
DMAP1	gi|7243231	T473	
TCOF1	gi|1587477	S1284	
ZNF91	gi|549839	T1038, T1039	
POLR1A	gi|49256401	S931	
DBP5 (DDX19)	gi|17046381	S154, S283, S1002, S1556, S1697, S1948, S1950, S1952, S2009, S2011, S2013	
BAT2	gi|18375626	S1089, S1219	^{597}EGPEPPEEVPPPTTPPVPK^{615}+P
ZC3HAV1	gi|119604309	S19	
AKAP8	gi|5031579	S323, S328, S339	
EXOSC9	gi|179287	S222	
IGF2BP1	gi|56237027	S181	
SAF-A2	gi|52545896	S40, S107, T44	
SAFB	gi|62244004	S271, S535, S572, S730	
CARF	gi|8923040	S10, Y8	^{378}SSSQTSGSLVSK^{386}+P, ^{161}TSAKTER^{167}+P
CLK4	gi|10190706	S138, S140	
RBMY1F	gi|113430736	S164	
ZNF638	gi|1374698	S490, S1401, T1487	
SAFB2	gi|7661936	S287	^{189}NTLETSSLNFK^{199}+P
PPHLN1	gi|7020106	S21	
ESR2	gi|3091286		^{133}CASPVTGPGSK^{143}+P, ^{199}KSCQACR^{205}+P

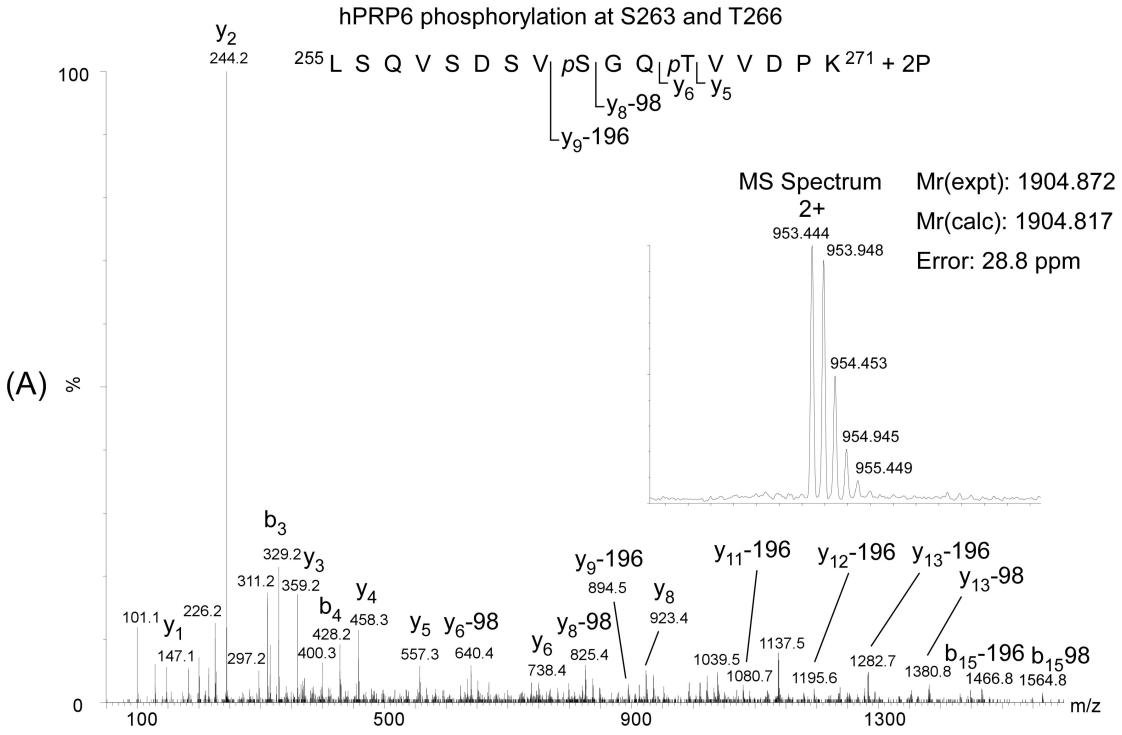
UTY	gi|2580576		^{380}RCSNTSTLAAR^{390}+P
DDX50	gi|16551055	S48	
HIST1H2AD	gi|89037003		^{120}KTESHLK^{126}+P
FRA10AC1	gi|14189976	S151, S156, S161	^{155}NSDEEESASESELWK^{169}+3P
WIZ	gi|4056411		^{124}TPLALAGSPTPK^{135}+P
ZNF136	gi|4507987	S292, T294	
TGS1	gi|14278850	S85, S89	
MYCBP2	gi|3319326	S2797	^{2794}SSPSGASSPR^{2803}+2P, ^{2766}MPSSSRAESPGPGSR^{2780}+P
TTN	gi|1212992	S16047, T18743, T18744	
CCNL2	gi|33150646	S15, S46, S102, S104	
TSEN54	gi|31565520		^{203}RSSSSPR^{209}+2P
BIVM	gi|7020076	S254, S255	
CASZ1	gi|78191040		^{1550}QFSSSADCAVPDCK^{1563}+P
HN1L	gi|21700763	S188	
SNRPN	gi|36495	T8	
RANBP2	gi|857368	S778, T779	
BAT2D1	gi|5541863	S1249, S2105	
Membrane (Location)			
NKTR	gi|6631100	S463, S613, S1077, S1146	
PARD3	gi|13491612	S201	^{466}KGTEGLGFSITSR^{478}+P
SEPT9	gi|5106557	S12	
MFI2	gi|136204	S51	
CLEC4E	gi|7657333		^{6}SSETQCTER^{14}+P
SDK1	gi|32351274	S522	
EXT1	gi|1168162		^{127}IAESYQNILAAIEGSR^{142}+2P
ABCB10	gi|9961244	S34	
ABCA7	gi|119589965	S1255	
KCNS2	gi|6329973	S475	
ARTS1	gi|6381989		^{714}QTWTDEGSVSERMLR^{728}+P
SIGLEC15	gi|119621867		^{468}RCNVVCATSPER^{479}+P
BCL2L7P1	gi|2493275	S3	
ENPP1	gi|119568426	T707, T713	
NDUFB3	gi|4505361	T27	
TAPT1	gi|22759976	Y99	
KIR3DL1	gi|995757	S243, S245, S246	
KIAA0716	gi|21757978	S121	
GPRIN1	gi|21739648	S660, S663	

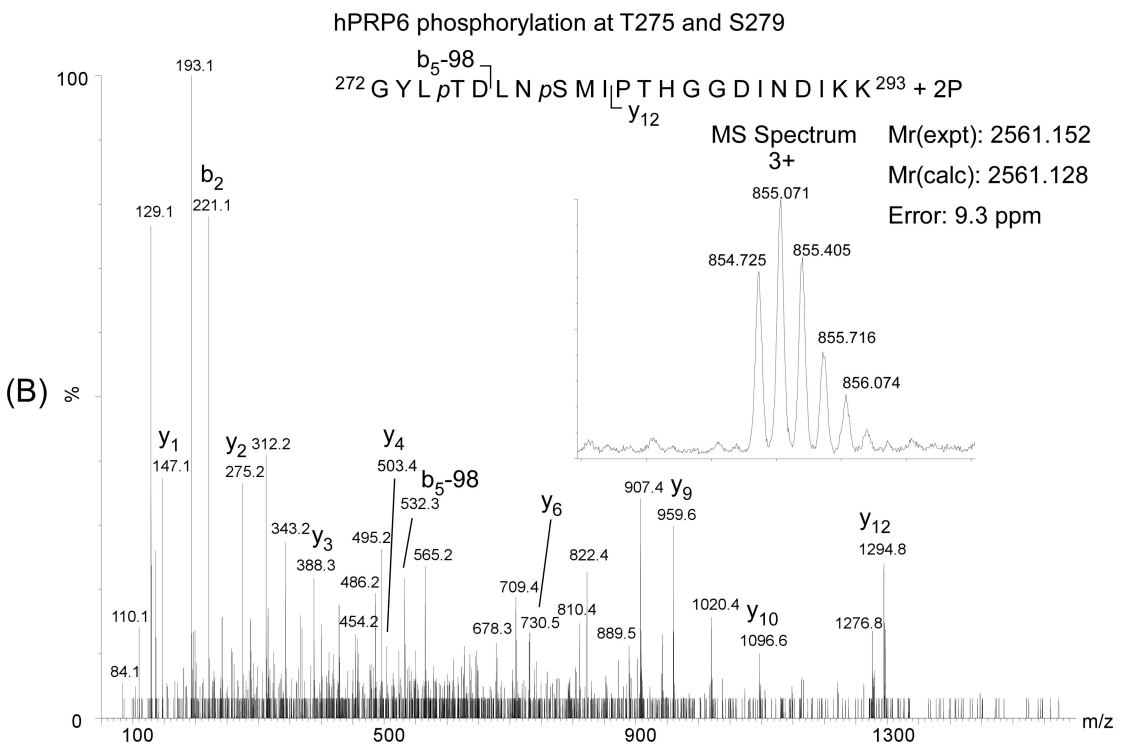
TMEM61	gi|32698902	S92	
RAB43	gi|50234889	S166, S167	
ALG11	gi|34535907	S169, S172	
Cytoplasm			
(Location)			
BSN	gi|2662149	S634	
ABLIM1	gi|57162149	S96	
AMBRA1	gi|10435015		^{533}SSERPGTSR^{541}+P
ARHGDIA	gi|36038		^{179}SRFTDDDK^{186}+P
HSP90AB1	gi|306891	S255	
CAB39	gi|42543739	T40, Y50	
SPTBN1	gi|119620543	S2016, S2255	
CTTN	gi|2498954	S418	^{399}TQTPPVSPAPQPTEERLPSSPVYEDAASFK^{428}+P
FXR1	gi|1730139	T5	
DBN1	gi|2498313	S142	
DSP	gi|1147813	S333	
SHB	gi|406738		^{448}QSSPSPSR^{455}+P
DIXDC1	gi|12698015	T436	
SPTBN5	gi|17369320	S766, S767, S2501, S3052	
Others			
WDR79	gi|8922396	S54, S85, S90, S491	
GPATCH8	gi|50962882	S653, S740, S758, S898, S981, S1009, S1014, S1033, S1035, S1107	
LMO7	gi|17225574	S533, S654, S657, S692, S1259	
PRSAP2	gi|89059414	S472, T626	^{619}KCRTGSMTHGLK^{630}+P, ^{1771}ASTXASTATASRTMRXK^{1787}+P
FNBP4	gi|6808095	S16, S112, S427, S460	
RPLP1	gi|4506669	S101, S104	
CCNK	gi|8980825	S324, S340	^{322}KPSPQPSSPR^{331}+2P
TNRC15	gi|7023190	S26	
SEPT2	gi|1040689	S218	
PLEC1	gi|41322923		^{19}TSSEDNLYLAVLR^{31}+P
BCR	gi|44238463	S60	
AFAP1L1	gi|10441465	S668	
LOC729639	gi|113425854	S348	
LOC731115	gi|113429485	T8	
GRHPR	gi|10439099	S192	

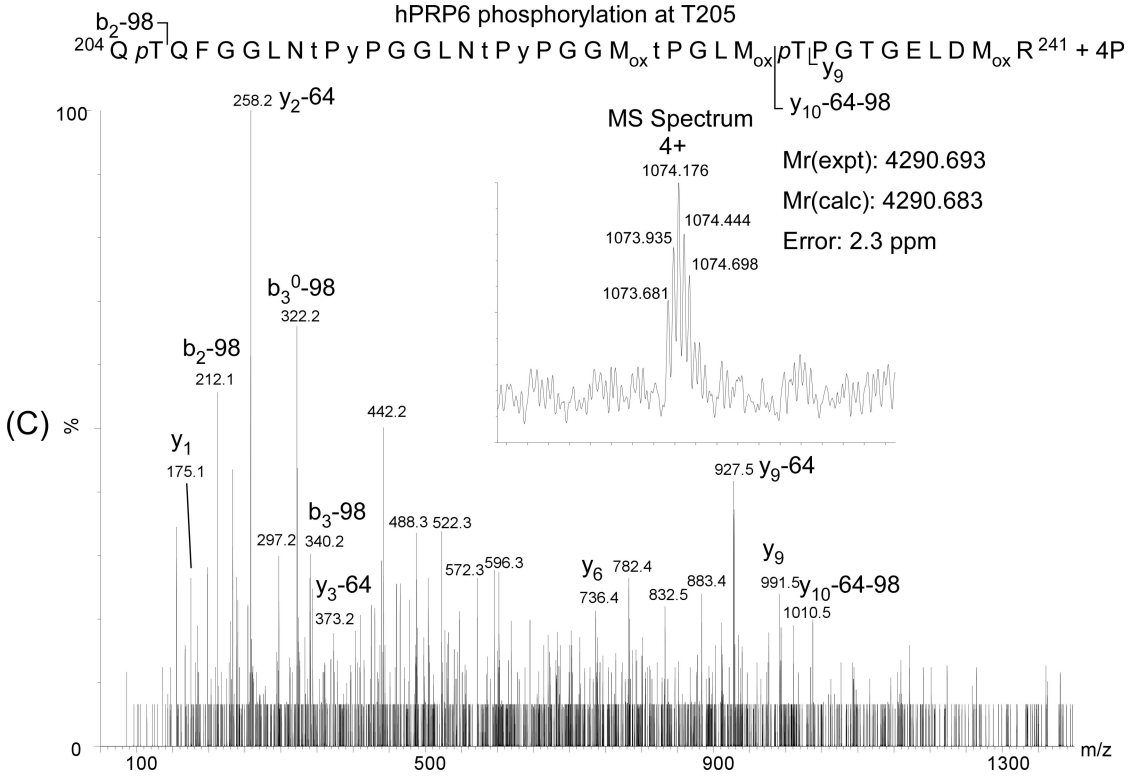
SPECC1L	gi|2280485		^{524}ETERSDMK^{531}+P
KIAA1783	gi|89043071	S279	
INPP5B	gi|1019103	S163, T165	
TRIML1	gi|31542779	T275	
hCG1820449	gi|28207851	S87, S88	
hCG1654449	gi|119602732	S10	
ANKRD26	gi|34532267	Y34, Y38	
SPARCL1	gi|21359871	T628	
hCG2036616	gi|88989580	T95	
LOC400870	gi|46409568	S92	
FAM112B	gi|21389379	T50	
LOC729518	gi|113426178	S88	
TTC28	gi|16741254	S1073	
GPATCH1	gi|21361684	S6, S8	
hCG2025834	gi|119603486	S148	
TSSC4	gi|4567068	S132	^{319}SSSPEDPGAEV^{329}+P
YME1L1	gi|7657689	S163	
EEF1D	gi|38522	T256	
ZCCHC3	gi|23396534	T3	
TTTY12	gi|27805759	S78	
TNS1	gi|111599618	T1506	
hCG22052	gi|119568106	S156, Y157, S159	
KRI1	gi|10434729	S59	
LIMCH1	gi|20521766	S736	
hCG1817914	gi|119571393		^{632}SGSTGSLKHLR^{642}+P
RPS3	gi|32532	T221	
NRD1	gi|2462485	S106, Y113	
LOC731726	gi|113422156	T132	
hCG1993567	gi|119570834		^{1}MVELIFIPTNSDSR^{14}+P
ZC3H4	gi|4884368	S283	
C9orf131	gi|6599155	T102	
MMTAG2	gi|13236559		^{210}RPAEATSSPTSPERPR^{225}+P
TMOD4	gi|21751358	T118	
myeloid/lymphoid	gi|13376343		^{33}SSPIPPFSSR^{42}+P
hCG2038818	gi|119622863	S135	
RPL14	gi|1710488	S139	
C12orf11	gi|6807949	S59	
C17orf47	gi|34194295	S483, S486	

TCTEX1D4	gi|61966721	S24	
EIF3C	gi|1931584	S39	
CCDC131	gi|29421174	S37, S958	
TUBA1C	gi|340021	S48, S439	
USP42	gi|51094456	S856, S936, S1263	^{747}DAEPQPGSPAAESLEEPDAAAGLSSTK^{773}+P
RBBP6	gi|15705403	S340, S554, S582, S584, S594, S596, S685, S697, S1003, S1045, S1101, S1152, S1449, T1292	
PCYT1B	gi|3153239	S271	
hCG1993516	gi|119593833	S42, S44	
BOK	gi|6456033	T69	
EIF3B	gi|1778051		^{142}ALENGDADEPSFSDPEDFVDDVSEEELLGDVLK^{174}+P
EIF5	gi|1229140	S389, S390	
CDC2L5	gi|14110387	S317, S325	
RBM12B	gi|118722349	S638	^{226}SSSPIPR^{232}+P
SFRP4	gi|4506895		^{226}SSSPIPR^{232}+P
C19orf34	gi|109452617		^{16}ATVITTR^{22}+P
TUBB5	gi|18088719	S172	
DHX34	gi|38158022	S749, S750	
HSPA2	gi|4204880	T164	
ZNHIT3	gi|703110	S77	
S22900	gi|107784	S85	
SGIP1	gi|51476757	T227	
BAB15370	gi|10438882	S4, S6	
hCG1981358	gi|119602855	S3, T4	
EXOC1	gi|7023220	Y400	
FGA	gi|223918	S3	
FBXO10	gi|6103649	S127, S464	^{202}KRSSSSPR^{209}+P
CXorf40A	gi|31377535	S32, S125, T122	
PIK3AP1	gi|20987486	S241	
SYBU	gi|50949414	S52, S55	
LCA10	gi|116517307	T97	
SNX32	gi|16550234	S250	
C1orf74	gi|22749019	S125, S128	
ZC3H6	gi|34528944	T313	
BAC85541	gi|34528618	S4, S7	
HARSL	gi|15029520	S161	

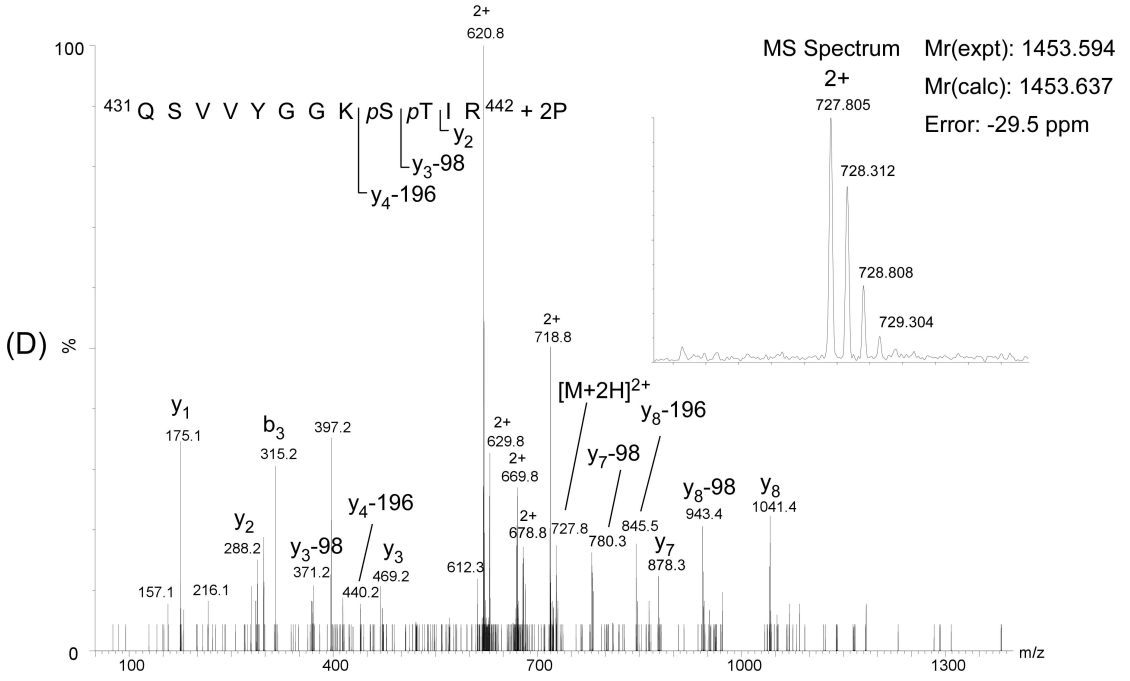
Appendix 2. MS and MS/MS spectra of phosphopeptides derived from human PRP6 (A-C) and PRP31 (D-G). The experimentally measured precursor masses and the charge states are listed in the MS spectra (small boxes). The sequences of the phosphopeptides are listed within the MS/MS spectra. Selected y- and b-type ions that led to the identification of the actual phosphorylation sites are indicated within the sequences and at the corresponding fragment peaks in the MS/MS spectra. Unambiguously identified phosphorylation sites are listed as pS, pT and pY, potentially phosphorylated amino acids are in small letters (s, t, y). Neutral loss of phosphoric acid (-98) of y- and/or b-type ions is indicated within the sequences and MS/MS spectra



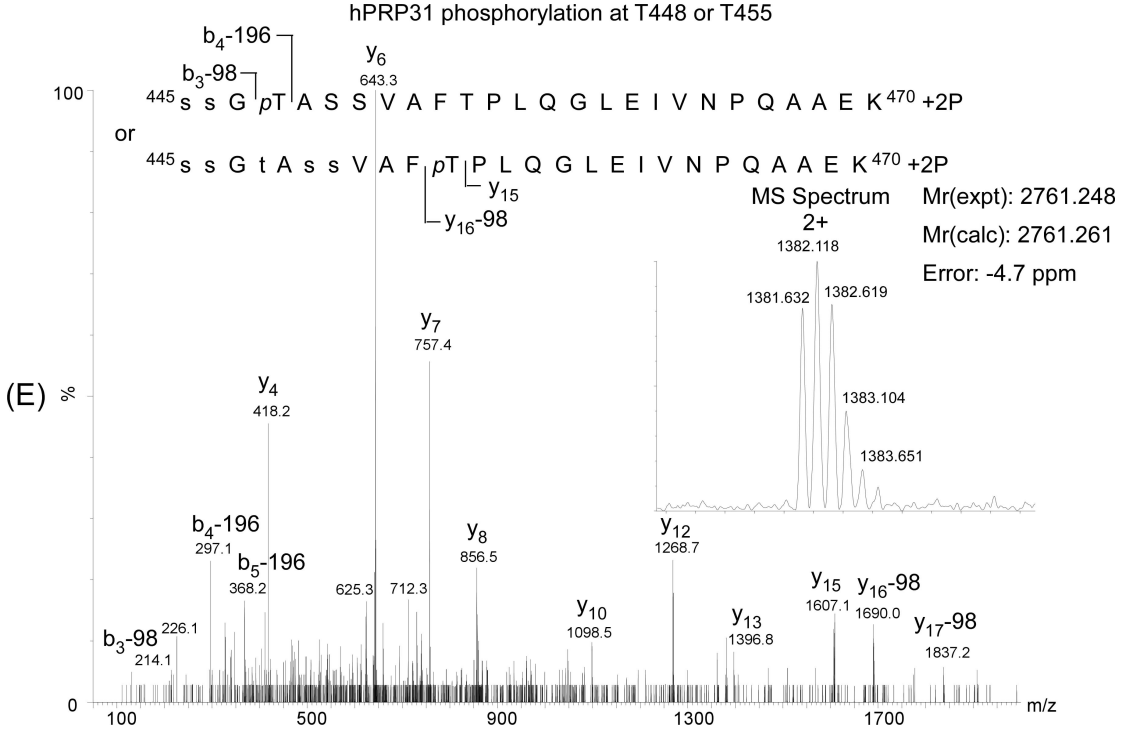

hPRP6 phosphorylation at T275 and S279
b5-98
272 G Y L pT D L N pS M I P T H G G D I N D I K K 293 + 2P
y12
MS Spectrum
3+
Mr(expt): 2561.152
Mr(calc): 2561.128
Error: 9.3 ppm
(B)
%
m/z
100
500
900
1300
193.1
b2
221.1
129.1
855.071
854.725
855.405
855.716
856.074
y1
147.1
y2 312.2
275.2
y4
503.4
b5-98
532.3
y6
907.4 y9
959.6
y12
1294.8
343.2 y3
388.3
495.2
565.2
822.4
486.2
709.4
810.4
1020.4 y10
1096.6
1276.8
110.1
454.2
678.3
730.5
889.5
84.1

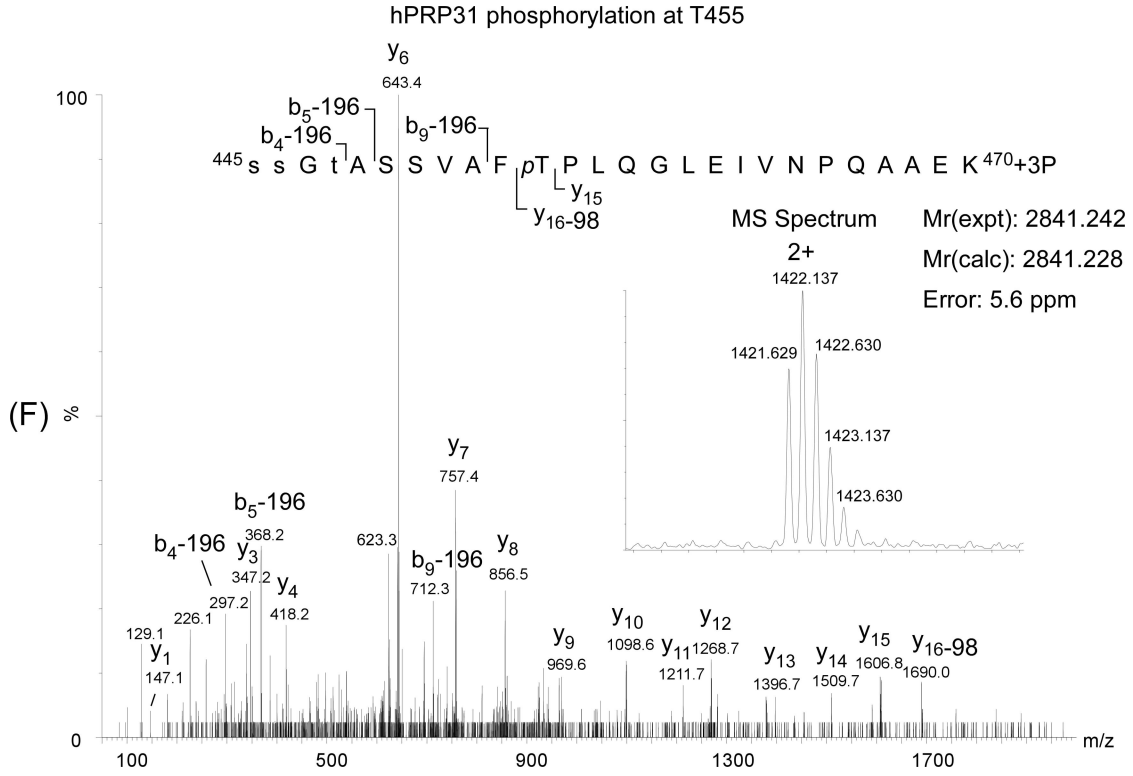

hPRP6 phosphorylation at T205
b2-98
204 Q pT Q F G G L N t P y P G G L N t P y P G G Mox t P G L Mox pT P G T G E L D Mox R 241 + 4P
y9
y10-64-98
MS Spectrum
4+
Mr(expt): 4290.693
Mr(calc): 4290.683
Error: 2.3 ppm
(C)
%
m/z
100
500
900
1300
258.2 y2-64
1074.176
1074.444
1073.935
1074.698
1073.681
b3 0-98
322.2
b2-98
212.1
442.2
y1
175.1
927.5 y9-64
b3-98
297.2
340.2
488.3
522.3
572.3
596.3
y6
782.4
736.4
832.5
883.4
y9
991.5
y3-64
373.2
y10-64-98
1010.5

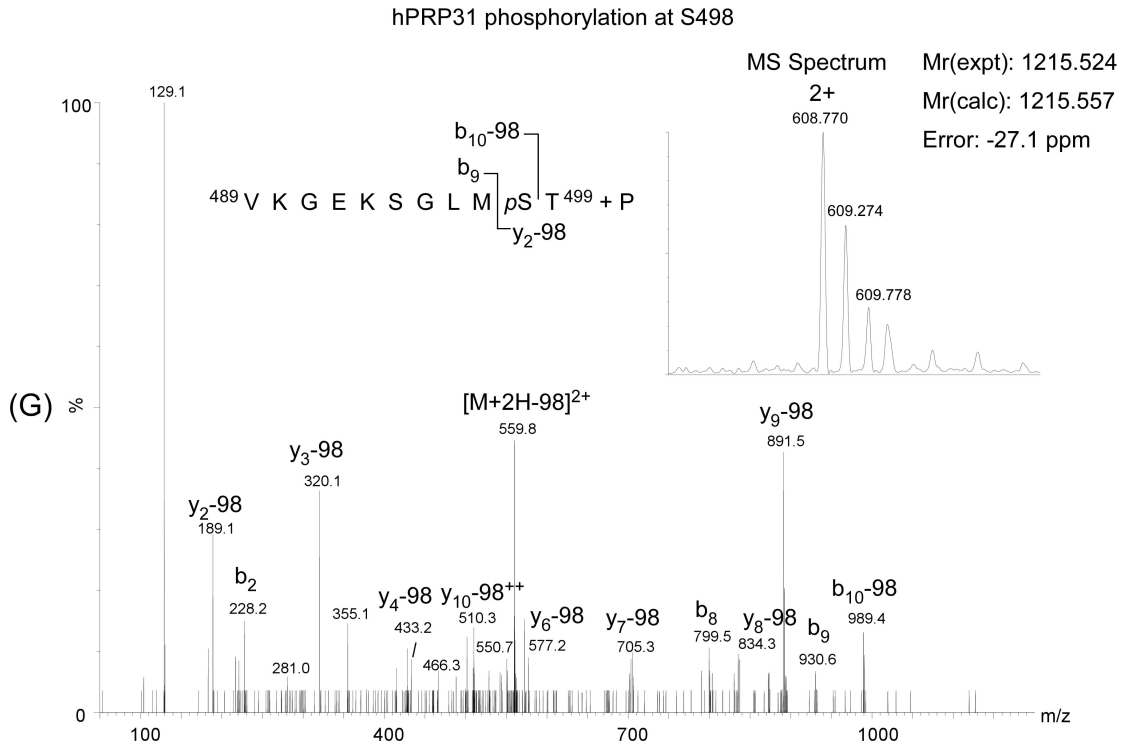
hPRP31 phosphorylation at S439 and T440



hPRP31 phosphorylation at T448 or T455




hPRP31 phosphorylation at T455
(F)
445 s s G t A S S V A F pT P L Q G L E I V N P Q A A E K 470 +3P
MS Spectrum
2+
Mr(expt): 2841.242
Mr(calc): 2841.228
Error: 5.6 ppm
m/z
%


hPRP31 phosphorylation at S498
(G)
489 V K G E K S G L M pS T 499 + P
MS Spectrum
2+
Mr(expt): 1215.524
Mr(calc): 1215.557
Error: -27.1 ppm
m/z
%

Appendix 3. Phosphorylation sites identified from U1 snRNPs by using CPP method in combination with in-house TiO_2 microspin column enrichment. The underline indicates that the phosphorylation site was not identified previously.

Protein	Accession Number	Phosphorylation sites	Potential phosphorylation sites
U1 snRNPs			
70K	IPI00290204	Y38; Y112; S117; Y126; Y206; S211; S226; S245; S257; S259; S266; S268; S281; S293; S295; S385; S410	
A	IPI00012382	S115	
C	IPI00013396	Y5; S17	
Sm proteins			
B/B'	IPI00027285	T30	
D2	IPI00017963	T12; T40	
D3	IPI00017964	S2; S44	
F	IPI00220528	S2	
SR proteins			
SFRS1	IPI00215884	S199	^{194}VDGPR*p*SP*p*[SY]GR^{204}
SC35	IPI00005978	S26; Y144; S147; S149; S187; S189; S191; S202; S204; S206; S208; S212	
SRP20	IPI00010204	S108; S148; S150; S152	
SRP40	IPI00012341	S229; S231; S233; S248; S253	
SRp30c	IPI00012340	S204; S211	
SFRS11	IPI00464952	S366; S368; S370; S434; S456	
SFRS16	IPI00432061	S101; S285; S335; Y453; S456	
LUC7L2	IPI00006932	S444; S450	
RSRC1	IPI00395014	S6; S61	
SR related proteins			
SRm160	IPI00328293	T220; S260; S308; Y309; S341; S343; S380; S389; S391; S393; S402; T406; S431; S463; S561; S563; T584; T586; S602; S604; S609; S617; S619; T626; S628; S638; S640; S648; S650; S665; S667; S708; S785; S886	^{714}RGA*p*[SSS]PQRR^{723}
SRm160	IPI00647720		^{426}SRVSV*p*SPGR*p*[TS]GK^{438}

SRm300	IPI00782992	S295; S297; S404; S408; S472; S474; T476; S575; T577; S596; T598; S890; S892; S894; S895; S908; T983; T1003; S1320; S1791; S1798; S2343; S2355; S2449; S2581; T2583; S2702	^{1041}*p*[SST]PPGESYFGVSSLQLK^{1058}; ^{2397}*p*[TS]PPLLDR^{2404}; ^{1204}D*p*[TLRT]PPRERSGAG*p*[SS]PETK^{1223}; ^{1984}SR*p*[TS]PVTR^{1991}; ^{979}VKPE*p*TPPRQ*p*[SHSGSIS]PYPK^{998}; ^{2578}RVP*p*[SPT]PAPK^{2587}; ^{307}DAPFSEPGTTSTQRP*p*[SS]PETATK^{329}; ^{889}HSCSG*p*[SS]PPR^{898}; ^{2687}DSRSLSYSPVERRRP*p*[SPQPS]PR^{2708}; ^{2407}*p*[SRT]PP*p*[SAPS]QSR^{2418}; ^{810}TPPRRSRSS*p*[SS]PPPK^{824}; ^{2144}*p*[TPMS]VLQQAGGSMMDGPGPR^{2163}; ^{374}HGGSPQPLATTPLSQEPVNPP*p*[SEAS]PTR^{401}; ^{883}*p*[SRTPS]RH*p*SC*p*SG*p*S*p*SPPRVKSS*p*[TPPRQ SPSRSSS]PQPKVK^{920}
U5 snRNPs			
52K	IPI00006103	S49; S139; S141; S195; T243	^{188}KGPGQP*p*[SS]PQR^{198}
100K	IPI00006725	S14; T25; S107; S109	^{104}KR*p*[SSLS]PGR^{112}
116K	IPI00003519	S19	^{1}MD*p*[TDLY]DEFGNYIGPELD*p*SDEDDDELGR^{28}
200K	IPI00420014	S207; S225; S1799; S1803	
U4/U6.U5 Tri-snRNPs			
65K	IPI00419844	S82	
110K	IPI00021417	S65; S448	^{64}G*p*SGRRGAEAEAR*p*[SST]HGR^{81}
TFIP11	IPI00015924	S210	
17S U2 snRNPs			
SF3a120	IPI00017451	S329; S359	
17S U2 related protein			
SPF45	IPI00176706	S155	^{222}*p*[SPT]GPSNSFLANMGGTVAHK^{241}
U11/U12 snRNPs			
65K	IPI00514393	S21; S108	
pre-mRNA/mRNA binding proteins			
BCLAF1	IPI00886854	S175; S383; S510; S646; S656	^{282}*p*[YS]PSQNSPIHHIPSR^{296}
DBPA	IPI00031801	S34	
Recruited to B Complex			
PRPF4B	IPI00013721	S227; S257; S277; S292; S294; S354; S356; Y849	^{845}DITP*p*[YLVS]RFYRAPEIIIGK^{864}; ^{324}KPIKp[SPS]K^{331}

THRAP3	IPI00104050	S320; S535; S682	
Potential C Complex			
TOE1	IPI00549516	S5	^{23}*p*[STTS]GEELVVQVPVVDVQSNNFK^{45}
EJC/mRNP			
Acinus	IPI00911038	S118; S286; S290; S380; S507; Y572; S802; S908	
Nucleus			
ARGLU1	IPI00844406	S77	
Coilin	IPI00006442		^{559}DPRLIIE*p*[SPS]NT*p*[SST]EPA^{576}
DKC1	IPI00221394	S21; S451; S453; S455; S494	^{401}HGKP*p*[TDST]PATWK^{413}
SRPK1	IPI00290439	S51; S309; S311	
HNRNPUL1	IPI00013070	S194	
Others			
SRRT	IPI00220038	S67; S74	
SNRNP40	IPI00006723	S292	
RRBP1	IPI00215743	S765	
NOP2	IPI00294891	S82; T88; S148; S150	
C19orf43	IPI00031526		^{41}DEEVSGAG*p*[SS]PVSGGVNLFAN^{61}

Appendix 4A. MASCOT searching result against swissprot bovine database by using regular data-dependent acquisition (Figure 4.7A).

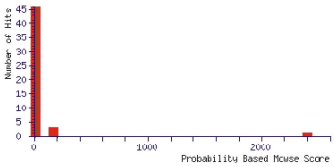
Mascot Search Results

User : He-Hsuan
Email : hhsiao@gwdg.de
Search title :
MS data file : D:\Data in MPIbpc\Orbi\Glyco&phospho\Hsiao_b_casein_10fmole_BSA_DDA_20100204.mgf
Database : SwissProt 20091103 (512205 sequences; 180277873 residues)
Taxonomy : Bos taurus (5730 sequences)
Timestamp : 3 Jun 2010 at 20:04:16 GMT
Protein hits : P02769 Serum albumin OS=Bos taurus GN=ALB PE=1 SV=4
Q2UVX4 Complement C3 OS=Bos taurus GN=C3 PE=1 SV=2
Q3SZR3 Alpha-1-acid glycoprotein OS=Bos taurus GN=ORM1 PE=2 SV=1
P02672 Fibrinogen alpha chain OS=Bos taurus GN=FGA PE=1 SV=5
P02662 Alpha-S1-casein OS=Bos taurus GN=CSN1S1 PE=1 SV=2
Q08DK9 Mutated melanoma-associated antigen 1 OS=Bos taurus GN=MUM1 PE=2 SV=1
P02666 Beta-casein OS=Bos taurus GN=CSN2 PE=1 SV=2

	SwissProt	Decoy	False discovery rate
Peptide matches above identity threshold	83	1	1.20 %
Peptide matches above homology or identity threshold	83	1	1.20 %

Probability Based Mowse Score

Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 33 indicate identity or extensive homology (p<0.001).
Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.



Peptide Summary Report

Format As | Peptide Summary | Help
Significance threshold p< 0.001 | Max. number of hits AUTO
Standard scoring | MudPIT scoring | Ions score or expect cut-off 10 | Show sub-sets 0
Show pop-ups | Suppress pop-ups | Sort unassigned Decreasing Score | Require bold red
Select All | Select None | Search Selected | Error tolerant | Archive Report

1. P02769 Mass: 69248 Score: 2400 Queries matched: 157 emPAI: 10.12
Serum albumin OS=Bos taurus GN=ALB PE=1 SV=4
Check to include this hit in error tolerant search or archive report

Query	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Rank	Peptide
2	351.2034	700.3922	700.3942	-2.74	0	(23)	0.0045	1	K.GACLLPK.I
13	356.6895	711.3644	711.3664	-2.83	0	27	0.0024	1	K.SEIAHR.F
48	376.6816	751.3487	751.3500	-1.83	0	17	0.029	1	K.NYQEAK.D
53	379.7141	757.4137	757.4156	-2.59	0	35	0.00029	1	K.GACLLPK.I + Carbamidomethyl (C)
54	379.7144	757.4142	757.4156	-1.95	0	(16)	0.023	1	K.GACLLPK.I + Carbamidomethyl (C)
55	379.7147	757.4148	757.4156	-1.06	0	(12)	0.12	1	K.GACLLPK.I + Carbamidomethyl (C)
56	379.7150	757.4154	757.4156	-0.25	0	(30)	0.0018	1	K.GACLLPK.I + Carbamidomethyl (C)
88	395.2375	788.4605	788.4644	-4.93	0	34	0.00085	1	K.LVTDLTK.V
89	395.2388	788.4630	788.4644	-1.76	0	(13)	0.12	1	K.LVTDLTK.V
90	395.2393	788.4641	788.4644	-0.36	0	(19)	0.016	1	K.LVTDLTK.V
115	409.7144	817.4142	817.4181	-4.85	0	(19)	0.027	1	K.ATEEQLK.T
116	409.7144	817.4142	817.4181	-4.77	0	(15)	0.076	1	K.ATEEQLK.T
117	409.7146	817.4147	817.4181	-4.18	0	28	0.0041	1	K.ATEEQLK.T
141	424.2538	846.4931	846.4963	-3.77	1	15	0.034	1	R.LSQKFPK.A
175	443.7086	885.4027	885.4080	-5.91	0	48	2.5e-005	1	K.DDSPDLPK.L
179	449.7428	897.4710	897.4742	-3.60	0	43	7.7e-005	1	R.LCVLHEK.T + Carbamidomethyl (C)
199	461.7465	921.4785	921.4807	-2.44	0	56	5.4e-006	1	K.AEFVEVTK.L
203	464.2484	926.4822	926.4861	-4.29	0	38	0.00022	1	K.YLYEIAR.R
237	487.7300	973.4454	973.4505	-5.23	0	(25)	0.0069	1	K.DLGEEHFK.G
238	487.7305	973.4464	973.4505	-4.23	0	30	0.0022	1	K.DLGEEHFK.G
268	501.7935	1001.5725	1001.5757	-3.17	0	52	1.4e-005	1	K.LVVSTQTALA.-
278	507.8115	1013.6084	1013.6121	-3.59	0	47	2.8e-005	1	K.QTALVELLK.H
304	517.7377	1033.4609	1033.4651	-4.05	0	28	0.004	1	R.NECFLSHK.D + Carbamidomethyl (C)
351	534.7220	1067.4294	1067.4342	-4.49	0	31	0.00095	1	K.QNCDQFEK.L + Carbamidomethyl (C)
358	536.7548	1071.4951	1071.5019	-6.33	0	54	7.2e-006	1	K.SHCIAEVEK.D + Carbamidomethyl (C)
359	358.1733	1071.4980	1071.5019	-3.61	0	(23)	0.011	1	K.SHCIAEVEK.D + Carbamidomethyl (C)
401	554.2579	1106.5013	1106.5066	-4.81	0	(28)	0.0029	1	K.EACFAVEGPK.L + Carbamidomethyl (C)
402	554.2587	1106.5028	1106.5066	-3.48	0	(36)	0.00038	1	K.EACFAVEGPK.L + Carbamidomethyl (C)

☑	403	554.2588	1106.5031	1106.5066	-3.16	0	44	6.6e-005	1	K.EACFAVEGPK.L + Carbamidomethyl (C)
☑	404	554.2593	1106.5040	1106.5066	-2.38	0	(26)	0.0036	1	K.EACFAVEGPK.L + Carbamidomethyl (C)
☑	405	554.2593	1106.5041	1106.5066	-2.27	0	(20)	0.016	1	K.EACFAVEGPK.L + Carbamidomethyl (C)
☑	406	554.2596	1106.5046	1106.5066	-1.83	0	(35)	0.00052	1	K.EACFAVEGPK.L + Carbamidomethyl (C)
☑	407	554.2598	1106.5051	1106.5066	-1.39	0	(15)	0.052	1	K.EACFAVEGPK.L + Carbamidomethyl (C)
☑	408	554.2601	1106.5057	1106.5066	-0.84	0	(19)	0.022	1	K.EACFAVEGPK.L + Carbamidomethyl (C)
☑	409	554.2609	1106.5072	1106.5066	0.48	0	(19)	0.027	1	K.EACFAVEGPK.L + Carbamidomethyl (C)
☑	451	569.7513	1137.4880	1137.4907	-2.34	0	35	0.0005	1	K.CCTESLVNR.R + 2 Carbamidomethyl (C)
☑	456	571.8580	1141.7015	1141.7070	-4.83	1	57	1.8e-006	1	K.KQTALVELLK.H
☑	457	381.5753	1141.7040	1141.7070	-2.62	1	(53)	5.6e-006	1	K.KQTALVELLK.H
☑	495	582.3162	1162.6179	1162.6234	-4.71	0	(37)	0.00045	1	K.LVNELTEFAK.T
☑	496	582.3166	1162.6187	1162.6234	-3.98	0	(22)	0.017	1	K.LVNELTEFAK.T
☑	497	582.3168	1162.6191	1162.6234	-3.66	0	(53)	1.3e-005	1	K.LVNELTEFAK.T
☑	498	582.3170	1162.6195	1162.6234	-3.34	0	61	1.5e-006	1	K.LVNELTEFAK.T
☑	499	582.3173	1162.6200	1162.6234	-2.93	0	(23)	0.0098	1	K.LVNELTEFAK.T
☑	500	582.3183	1162.6220	1162.6234	-1.14	0	(28)	0.0025	1	K.LVNELTEFAK.T
☑	501	582.3183	1162.6220	1162.6234	-1.14	0	(36)	0.00047	1	K.LVNELTEFAK.T
☑	502	582.3186	1162.6227	1162.6234	-0.61	0	(30)	0.0016	1	K.LVNELTEFAK.T
☑	503	582.3191	1162.6236	1162.6234	0.23	0	(34)	0.00066	1	K.LVNELTEFAK.T
☑	504	582.3192	1162.6238	1162.6234	0.33	0	(20)	0.017	1	K.LVNELTEFAK.T
☑	505	582.3194	1162.6242	1162.6234	0.75	0	(24)	0.0082	1	K.LVNELTEFAK.T
☑	618	417.2099	1248.6080	1248.6139	-4.73	1	55	9e-006	1	R.FKDLGEEHFK.G
☑	619	625.3123	1248.6100	1248.6139	-3.13	1	(51)	3.1e-005	1	R.FKDLGEEHFK.G
☑	620	625.3138	1248.6130	1248.6139	-0.69	1	(41)	0.00026	1	R.FKDLGEEHFK.G
☑	656	428.5737	1282.6992	1282.7034	-3.25	0	34	0.00065	1	R.HPEYAVSVLLR.L
☑	673	646.3022	1290.5898	1290.5948	-3.85	0	50	2.6e-005	1	K.ECCDKPLLEK.S + 2 Carbamidomethyl (C)
☑	692	653.3594	1304.7042	1304.7088	-3.55	0	(13)	0.12	1	K.HLVDEPQNLIK.Q
☑	693	653.3595	1304.7044	1304.7088	-3.36	0	(25)	0.007	1	K.HLVDEPQNLIK.Q
☑	694	653.3598	1304.7051	1304.7088	-2.90	0	(23)	0.012	1	K.HLVDEPQNLIK.Q
☑	695	653.3604	1304.7063	1304.7088	-1.96	0	41	0.00019	1	K.HLVDEPQNLIK.Q
☑	696	435.9095	1304.7067	1304.7088	-1.62	0	(22)	0.015	1	K.HLVDEPQNLIK.Q
☑	698	653.3621	1304.7097	1304.7088	0.66	0	(10)	0.2	1	K.HLVDEPQNLIK.Q
☑	699	653.3633	1304.7121	1304.7088	2.53	0	(17)	0.041	1	K.HLVDEPQNLIK.Q
☑	700	653.3635	1304.7124	1304.7088	2.71	0	(14)	0.087	1	K.HLVDEPQNLIK.Q
☑	792	700.3485	1398.6825	1398.6853	-2.06	0	(49)	3.3e-005	1	K.TVMENFVAFVDK.C
☑	816	708.3457	1414.6769	1414.6803	-2.41	0	61	2.7e-006	1	K.TVMENFVAFVDK.C + Oxidation (M)
☑	819	473.9007	1418.6802	1418.6864	-4.40	0	(29)	0.0041	1	K.SLHTLFGDELCK.V + Carbamidomethyl (C)
☑	820	473.9010	1418.6811	1418.6864	-3.75	0	(25)	0.0099	1	K.SLHTLFGDELCK.V + Carbamidomethyl (C)
☑	821	473.9010	1418.6812	1418.6864	-3.68	0	(41)	0.00024	1	K.SLHTLFGDELCK.V + Carbamidomethyl (C)
☑	822	473.9011	1418.6814	1418.6864	-3.56	0	(20)	0.029	1	K.SLHTLFGDELCK.V + Carbamidomethyl (C)
☑	823	473.9012	1418.6819	1418.6864	-3.17	0	(36)	0.00072	1	K.SLHTLFGDELCK.V + Carbamidomethyl (C)
☑	824	710.3484	1418.6822	1418.6864	-2.94	0	57	6.2e-006	1	K.SLHTLFGDELCK.V + Carbamidomethyl (C)
☑	826	473.9016	1418.6830	1418.6864	-2.39	0	(35)	0.00083	1	K.SLHTLFGDELCK.V + Carbamidomethyl (C)
☑	827	473.9016	1418.6830	1418.6864	-2.39	0	(20)	0.028	1	K.SLHTLFGDELCK.V + Carbamidomethyl (C)
☑	828	710.3489	1418.6832	1418.6864	-2.25	0	(50)	2.6e-005	1	K.SLHTLFGDELCK.V + Carbamidomethyl (C)
☑	829	473.9018	1418.6835	1418.6864	-2.07	0	(22)	0.018	1	K.SLHTLFGDELCK.V + Carbamidomethyl (C)
☑	830	473.9019	1418.6838	1418.6864	-1.82	0	(18)	0.048	1	K.SLHTLFGDELCK.V + Carbamidomethyl (C)
☑	870	480.6063	1438.7970	1438.8045	-5.15	1	70	1.3e-007	1	R.RHPEYAVSVLLR.L
☑	871	480.6065	1438.7977	1438.8045	-4.70	1	(30)	0.0012	1	R.RHPEYAVSVLLR.L
☑	872	480.6069	1438.7989	1438.8045	-3.87	1	(58)	2.5e-006	1	R.RHPEYAVSVLLR.L
☑	873	720.4069	1438.7992	1438.8045	-3.68	1	(24)	0.0055	1	R.RHPEYAVSVLLR.L
☑	874	480.6071	1438.7996	1438.8045	-3.37	1	(61)	9.6e-007	1	R.RHPEYAVSVLLR.L
☑	875	720.4071	1438.7997	1438.8045	-3.34	1	(28)	0.0018	1	R.RHPEYAVSVLLR.L
☑	876	480.6072	1438.7999	1438.8045	-3.18	1	(28)	0.002	1	R.RHPEYAVSVLLR.L
☑	877	480.6075	1438.8007	1438.8045	-2.60	1	(33)	0.00063	1	R.RHPEYAVSVLLR.L
☑	878	360.7075	1438.8010	1438.8045	-2.42	1	(23)	0.0062	1	R.RHPEYAVSVLLR.L
☑	879	480.6081	1438.8025	1438.8045	-1.33	1	(32)	0.00075	1	R.RHPEYAVSVLLR.L
☑	880	360.7080	1438.8028	1438.8045	-1.15	1	(24)	0.0044	1	R.RHPEYAVSVLLR.L
☑	891	722.3222	1442.6299	1442.6347	-3.39	0	53	1.4e-005	1	K.YICDNQDTISSK.L + Carbamidomethyl (C)
☑	892	722.3235	1442.6325	1442.6347	-1.53	0	(18)	0.046	1	K.YICDNQDTISSK.L + Carbamidomethyl (C)
☑	893	722.3251	1442.6357	1442.6347	0.67	0	(26)	0.0072	1	K.YICDNQDTISSK.L + Carbamidomethyl (C)
☑	894	722.3268	1442.6390	1442.6347	2.96	0	(11)	0.26	1	K.YICDNQDTISSK.L + Carbamidomethyl (C)
☑	925	488.5322	1462.5748	1462.5817	-4.74	0	(30)	0.002	1	K.TCVADESHAGCEK.S + 2 Carbamidomethyl (C)
☑	926	732.2965	1462.5785	1462.5817	-2.20	0	72	1.2e-007	1	K.TCVADESHAGCEK.S + 2 Carbamidomethyl (C)
☑	937	489.5728	1465.6966	1465.7017	-3.49	1	46	7.1e-005	1	K.VTKCCTESLVNR.R + 2 Carbamidomethyl (C)
☑	949	739.7635	1477.5125	1477.5160	-2.31	0	83	7.6e-009	1	R.ETYGDMADCCEK.Q + 2 Carbamidomethyl (C)
☑	954	493.9355	1478.7847	1478.7881	-2.31	0	57	6.7e-006	1	K.LGEYGFQNALIVR.Y
☑	955	740.3997	1478.7848	1478.7881	-2.29	0	(50)	3.1e-005	1	K.LGEYGFQNALIVR.Y
☑	973	747.7617	1493.5088	1493.5109	-1.42	0	(58)	1.7e-006	1	R.ETYGDMADCCEK.Q + 2 Carbamidomethyl (C); Oxidation (M)
☑	992	751.8061	1501.5976	1501.6065	-5.88	0	(51)	1.7e-005	1	K.EYEATLEECCAK.D + 2 Carbamidomethyl (C)
☑	993	501.5403	1501.5990	1501.6065	-4.95	0	(13)	0.12	1	K.EYEATLEECCAK.D + 2 Carbamidomethyl (C)
☑	995	751.8116	1501.6087	1501.6065	1.51	0	(13)	0.12	1	K.EYEATLEECCAK.D + 2 Carbamidomethyl (C)
☑	996	751.8119	1501.6092	1501.6065	1.84	0	52	1.6e-005	1	K.EYEATLEECCAK.D + 2 Carbamidomethyl (C)
☑	1007	504.6174	1510.8305	1510.8355	-3.35	0	(34)	0.0008	1	K.VPQVSTPTLVEVSR.S
☑	1008	756.4265	1510.8384	1510.8355	1.87	0	80	1.9e-008	1	K.VPQVSTPTLVEVSR.S
☑	1053	511.5959	1531.7658	1531.7738	-5.20	1	(34)	0.00097	1	K.LKECCDKPLLEK.S + 2 Carbamidomethyl (C)
☑	1054	766.8904	1531.7663	1531.7738	-4.88	1	72	1.5e-007	1	K.LKECCDKPLLEK.S + 2 Carbamidomethyl (C)
☑	1055	383.9491	1531.7673	1531.7738	-4.26	1	(21)	0.02	1	K.LKECCDKPLLEK.S + 2 Carbamidomethyl (C)
☑	1091	777.8277	1553.6408	1553.6457	-3.11	0	54	8.7e-006	1	K.DDPHACYSTVFDK.L + Carbamidomethyl (C)
☑	1092	518.8878	1553.6416	1553.6457	-2.61	0	(17)	0.043	1	K.DDPHACYSTVFDK.L + Carbamidomethyl (C)

☑	1114	784.3754	1566.7362	1566.7354	0.47	0	80	4.1e-008	1	K.DAFLGSFLYEYSR.R
☑	1127	526.2579	1575.7520	1575.7603	-5.27	0	(14)	0.19	1	K.LKPDPNTLCDEFK.A + Carbamidomethyl (C)
☑	1128	526.2584	1575.7533	1575.7603	-4.46	0	(25)	0.014	1	K.LKPDPNTLCDEFK.A + Carbamidomethyl (C)
☑	1130	526.2599	1575.7580	1575.7603	-1.44	0	(10)	0.42	1	K.LKPDPNTLCDEFK.A + Carbamidomethyl (C)
☑	1131	788.8867	1575.7589	1575.7603	-0.89	0	53	1.9e-005	1	K.LKPDPNTLCDEFK.A + Carbamidomethyl (C)
☑	1132	526.2604	1575.7595	1575.7603	-0.51	0	(26)	0.011	1	K.LKPDPNTLCDEFK.A + Carbamidomethyl (C)
☑	1133	526.2607	1575.7602	1575.7603	-0.04	0	(29)	0.0048	1	K.LKPDPNTLCDEFK.A + Carbamidomethyl (C)
☑	1189	547.3148	1638.9226	1638.9305	-4.79	1	(38)	0.00025	1	R.KVPQVSTPTLVEVSR.S
☑	1190	547.3152	1638.9239	1638.9305	-4.01	1	(48)	2.7e-005	1	R.KVPQVSTPTLVEVSR.S
☑	1191	547.3153	1638.9241	1638.9305	-3.90	1	(65)	5.8e-007	1	R.KVPQVSTPTLVEVSR.S
☑	1192	547.3154	1638.9245	1638.9305	-3.67	1	(37)	0.00033	1	R.KVPQVSTPTLVEVSR.S
☑	1193	547.3155	1638.9246	1638.9305	-3.56	1	(65)	5.1e-007	1	R.KVPQVSTPTLVEVSR.S
☑	1194	547.3156	1638.9248	1638.9305	-3.45	1	(32)	0.001	1	R.KVPQVSTPTLVEVSR.S
☑	1195	547.3160	1638.9261	1638.9305	-2.67	1	(30)	0.0016	1	R.KVPQVSTPTLVEVSR.S
☑	1196	547.3162	1638.9267	1638.9305	-2.33	1	(28)	0.0027	1	R.KVPQVSTPTLVEVSR.S
☑	1197	547.3165	1638.9278	1638.9305	-1.66	1	(24)	0.0062	1	R.KVPQVSTPTLVEVSR.S
☑	1198	547.3168	1638.9287	1638.9305	-1.10	1	(21)	0.012	1	R.KVPQVSTPTLVEVSR.S
☑	1199	547.3168	1638.9287	1638.9305	-1.10	1	(17)	0.03	1	R.KVPQVSTPTLVEVSR.S
☑	1200	820.4717	1638.9288	1638.9305	-1.02	1	71	1.1e-007	1	R.KVPQVSTPTLVEVSR.S
☑	1201	547.3173	1638.9300	1638.9305	-0.32	1	(35)	0.00047	1	R.KVPQVSTPTLVEVSR.S
☑	1270	575.6150	1723.8231	1723.8273	-2.41	0	(31)	0.0036	1	R.MPCTEDYLSLILNR.L + Carbamidomethyl (C)
☑	1271	862.9195	1723.8244	1723.8273	-1.66	0	(54)	1.6e-005	1	R.MPCTEDYLSLILNR.L + Carbamidomethyl (C)
☑	1290	870.9164	1739.8182	1739.8222	-2.30	0	56	1e-005	1	R.MPCTEDYLSLILNR.L + Carbamidomethyl (C); Oxidation (M)
☑	1291	580.9475	1739.8207	1739.8222	-0.87	0	(19)	0.045	1	R.MPCTEDYLSLILNR.L + Carbamidomethyl (C); Oxidation (M)
☑	1298	874.3553	1746.6961	1746.6978	-0.93	0	83	1.1e-008	1	K.YNGVFQECCQAEDK.G + 2 Carbamidomethyl (C)
☑	1301	583.8901	1748.6484	1748.6553	-3.93	0	(43)	8.6e-005	1	K.ECCHGDLLECADDR.A + 3 Carbamidomethyl (C)
☑	1303	583.8910	1748.6511	1748.6553	-2.35	0	(45)	5.5e-005	1	K.ECCHGDLLECADDR.A + 3 Carbamidomethyl (C)
☑	1304	875.3336	1748.6526	1748.6553	-1.54	0	60	1.5e-006	1	K.ECCHGDLLECADDR.A + 3 Carbamidomethyl (C)
☑	1448	627.6428	1879.9066	1879.9138	-3.83	0	(21)	0.036	1	R.RPCFSALTPDETYVPK.A + Carbamidomethyl (C)
☑	1449	627.6431	1879.9074	1879.9138	-3.44	0	(16)	0.11	1	R.RPCFSALTPDETYVPK.A + Carbamidomethyl (C)
☑	1450	627.6432	1879.9077	1879.9138	-3.24	0	37	0.00079	1	R.RPCFSALTPDETYVPK.A + Carbamidomethyl (C)
☑	1451	627.6432	1879.9077	1879.9138	-3.24	0	(17)	0.085	1	R.RPCFSALTPDETYVPK.A + Carbamidomethyl (C)
☑	1453	627.6447	1879.9123	1879.9138	-0.81	0	(14)	0.16	1	R.RPCFSALTPDETYVPK.A + Carbamidomethyl (C)
☑	1454	940.9636	1879.9126	1879.9138	-0.67	0	(33)	0.002	1	R.RPCFSALTPDETYVPK.A + Carbamidomethyl (C)
☑	1455	627.6448	1879.9127	1879.9138	-0.61	0	(25)	0.014	1	R.RPCFSALTPDETYVPK.A + Carbamidomethyl (C)
☑	1460	630.3117	1887.9133	1887.9195	-3.30	0	37	0.0011	1	R.HPYFYAPELLYYANK.Y
☑	1461	944.9653	1887.9160	1887.9195	-1.87	0	(36)	0.0015	1	R.HPYFYAPELLYYANK.Y
☑	1469	634.6274	1900.8605	1900.8625	-1.06	1	(30)	0.0039	1	R.NECFLSHKDDSPDLPK.L + Carbamidomethyl (C)
☑	1470	951.4387	1900.8629	1900.8625	0.19	1	65	1.3e-006	1	R.NECFLSHKDDSPDLPK.L + Carbamidomethyl (C)
☑	1475	636.6434	1906.9083	1906.9135	-2.74	0	21	0.041	1	K.LFTFHADICTLPDTEK.Q + Carbamidomethyl (C)
☑	1497	652.6561	1954.9466	1954.9524	-2.97	0	(37)	0.0011	1	K.DAIPENLPPLTADFAEDK.D
☑	1498	978.4829	1954.9513	1954.9524	-0.56	0	(51)	4.1e-005	1	K.DAIPENLPPLTADFAEDK.D
☑	1500	978.4841	1954.9536	1954.9524	0.63	0	61	3.6e-006	1	K.DAIPENLPPLTADFAEDK.D
☑	1501	978.4871	1954.9597	1954.9524	3.75	0	(17)	0.09	1	K.DAIPENLPPLTADFAEDK.D
☑	1543	512.0107	2044.0136	2044.0206	-3.43	1	(19)	0.056	1	R.RHPYFYAPELLYYANK.Y
☑	1544	682.3462	2044.0167	2044.0206	-1.90	1	63	2.1e-006	1	R.RHPYFYAPELLYYANK.Y
☑	1545	512.0118	2044.0183	2044.0206	-1.16	1	(12)	0.28	1	R.RHPYFYAPELLYYANK.Y
☑	1546	1023.0171	2044.0196	2044.0206	-0.49	1	(39)	0.00056	1	R.RHPYFYAPELLYYANK.Y

2. Q2UVX4 **Mass:** 187135 **Score:** 149 **Queries matched:** 3 **emPAI:** 0.04
Complement C3 OS=Bos taurus GN=C3 PE=1 SV=2
☐ Check to include this hit in error tolerant search or archive report

	Query	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Rank	Peptide
☑	9	354.1674	706.3202	706.3221	-2.63	0	22	0.006	1	K.FPCQR.R + Carbamidomethyl (C)
☑	258	498.7390	995.4634	995.4672	-3.87	0	66	6e-007	1	K.AGQYSSDLR.K
☑	1244	844.8911	1687.7676	1687.7698	-1.33	0	61	3.8e-006	1	K.AFLDCCEYITQLR.Q + 2 Carbamidomethyl (C)

3. Q3SZR3 **Mass:** 23168 **Score:** 147 **Queries matched:** 5 **emPAI:** 0.31
Alpha-1-acid glycoprotein OS=Bos taurus GN=ORM1 PE=2 SV=1
☐ Check to include this hit in error tolerant search or archive report

	Query	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Rank	Peptide
☑	157	432.2455	862.4765	862.4800	-4.03	0	14	0.062	1	K.EFLDVIK.C
☑	293	513.2410	1024.4674	1024.4713	-3.81	0	46	6.5e-005	1	R.EYQTIEDK.C
☑	463	573.7883	1145.5620	1145.5658	-3.34	0	31	0.0021	1	K.WFYIGSAFR.N
☑	570	400.8908	1199.6507	1199.6550	-3.61	0	20	0.025	1	R.EHFVDLLLSK.H
☑	1541	680.3389	2037.9948	2038.0007	-2.92	0	36	0.0011	1	K.NVGVSFYADKPEVTQEQK.K

4. P02672 **Mass:** 66971 **Score:** 98 **Queries matched:** 2 **emPAI:** 0.05
Fibrinogen alpha chain OS=Bos taurus GN=FGA PE=1 SV=5
☐ Check to include this hit in error tolerant search or archive report

	Query	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Rank	Peptide
☑	235	487.2624	972.5102	972.5127	-2.65	0	19	0.029	1	K.ELLIDNEK.V
☑	1388	898.4380	1794.8615	1794.8636	-1.12	0	80	4.8e-008	1	R.TGLAPEFAALGESGSSSSK.T

5. P02662 **Mass:** 24513 **Score:** 51 **Queries matched:** 2 **emPAI:** 0.14

Alpha-S1-casein OS=Bos taurus GN=CSN1S1 PE=1 SV=2
Check to include this hit in error tolerant search or archive report

Query	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Rank	Peptide
638	634.3542	1266.6939	1266.6972	-2.56	0	34	0.00068	1	R.YLGYLEQLLR.L
779	692.8685	1383.7225	1383.7227	-0.17	0	16	0.062	1	R.FFVAPFPEVFGK.E

6. Q08DK9 **Mass:** 71064 **Score:** 38 **Queries matched:** 1 **emPAI:** 0.05
Mutated melanoma-associated antigen 1 OS=Bos taurus GN=MUM1 PE=2 SV=1
Check to include this hit in error tolerant search or archive report

Query	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Rank	Peptide
52	379.2497	756.4848	756.4858	-1.23	0	38	0.00027	1	R.GISVLLR.R

7. P02666 **Mass:** 25091 **Score:** 37 **Queries matched:** 3
Beta-casein OS=Bos taurus GN=CSN2 PE=1 SV=2
Check to include this hit in error tolerant search or archive report

Query	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Rank	Peptide
42	371.7274	741.4403	741.4425	-2.93	0	11	0.074	1	R.GPFPIIV.-
47	374.6889	747.3633	747.3625	1.04	0	12	0.15	1	K.EMPFPK.Y
82	390.7502	779.4859	779.4905	-5.89	0	14	0.042	1	K.VLPVPQK.A

Search Parameters

```
Type of search         : MS/MS Ion Search
Enzyme                 : Trypsin
Variable modifications : Carbamidomethyl (C),Oxidation (M),Phosphc-STY (STY)
Mass values            : Monoisotopic
Protein Mass           : Unrestricted
Peptide Mass Tolerance : ± 7 ppm
Fragment Mass Tolerance: ± 0.6 Da
Max Missed Cleavages   : 3
Instrument type        : ESI-TRAP
Number of queries      : 1617
```

Mascot: http://www.matrixscience.com/

Appendix 4B. MASCOT searching result against swissprot bovine database by using pseudo-neutral-loss acquisition (Figure 4.7B).

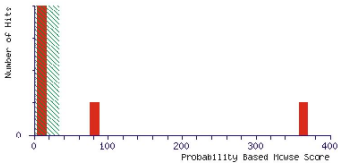
Mascot Search Results

```
User            : He-Hsuan
Email           : hhsiao@gwdg.de
Search title    :
MS data file    : D:\Data in MPIbpc\Orbi\Glyco&phospho\Hsiao_b_casein_10fmole_BSA_masstag_20100204.mgf
Database        : SwissProt 20091103 (512205 sequences; 180277873 residues)
Taxonomy        : Bos taurus (5730 sequences)
Timestamp       : 3 Jun 2010 at 20:05:15 GMT
Protein hits    : P02769 Serum albumin OS=Bos taurus GN=ALB PE=1 SV=4
                  P02666 Beta-casein OS=Bos taurus GN=CSN2 PE=1 SV=2
```

	SwissProt	Decoy	False discovery rate
Peptide matches above identity threshold	8	0	0.00 %
Peptide matches above homology or identity threshold	8	0	0.00 %

Probability Based Mowse Score

Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 33 indicate identity or extensive homology (p<0.001).
Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.



Peptide Summary Report

Format As: Peptide Summary — Help

Significance threshold p< 0.001 — Max. number of hits AUTO

Standard scoring ◉ MudPIT scoring ○ — Ions score or expect cut-off 10 — Show sub-sets 0

Show pop-ups ◉ Suppress pop-ups ○ — Sort unassigned Decreasing Score — Require bold red ☑

Select All | Select None | Search Selected | ☐ Error tolerant | Archive Report

1. P02769 Mass: 69248 Score: 364 Queries matched: 7
Serum albumin OS=Bos taurus GN=ALB PE=1 SV=4
☐ Check to include this hit in error tolerant search or archive report

Query	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Rank	Peptide
☑ 6	449.7429	897.4713	897.4742	-3.20	0	32	0.00098	1	R.LCVLHEK.T + Carbamidomethyl (C)
☑ 9	461.7490	921.4834	921.4807	2.92	0	52	1.3e-005	1	K.AEFVEVTK.L
☑ 29	625.3127	1248.6109	1248.6139	-2.36	1	42	0.00022	1	R.FKDLGEEHFK.G
☑ 41	739.7637	1477.5128	1477.5160	-2.14	0	75	4.4e-008	1	R.ETYGDMADCCEK.Q + 2 Carbamidomethyl (C)
☑ 51	788.8876	1575.7606	1575.7603	0.20	0	49	5.9e-005	1	K.LKPDPNTLCDEFK.A + Carbamidomethyl (C)
☑ 56	820.4708	1638.9271	1638.9305	-2.07	1	63	8e-007	1	R.KVPQVSTPTLVEVSR.S
☑ 63	652.6569	1954.9488	1954.9524	-1.85	0	52	3.2e-005	1	K.DAIPENLPPLTADFAEDK.D

2. P02666 Mass: 25091 Score: 85 Queries matched: 1
Beta-casein OS=Bos taurus GN=CSN2 PE=1 SV=2
☐ Check to include this hit in error tolerant search or archive report

Query	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Rank	Peptide
☑ 64	1031.4186	2060.8226	2060.8212	0.70	0	85	7.8e-009	1	K.FQSEEQQQTEDELQDK.I + Phospho-STY (STY)

Search Parameters

```
Type of search          : MS/MS Ion Search
Enzyme                  : Trypsin
Variable modifications  : Carbamidomethyl (C),Oxidation (M),Phosphc-STY (STY)
Mass values             : Monoisotopic
Protein Mass            : Unrestricted
Peptide Mass Tolerance  : ± 7 ppm
Fragment Mass Tolerance : ± 0.6 Da
Max Missed Cleavages    : 3
Instrument type         : ESI-TRAP
Number of queries       : 65
```

Mascot: http://www.matrixscience.com/

Curriculum Vitae

Name: He-Hsuan Hsiao (蕭鶴軒)

Address: Am Fassberg 11, 37075, Göttingen, Germany.

Tel: +49 551 201 1500 / 1060

Fax: +49 551 201 1197

E-mail: hhsiao@gwdg.de

I. PERSONAL INFORMATION

Date of Birth:	1977/08/22
Gender:	Male
Nationality:	Taiwan
Current Position:	PhD Student in Dr. Henning Urlaub's Lab Max Planck Institute for Biophysical Chemistry, Germany.

II. EDUCATIONAL BACKGROUND

02/07 - Now	PhD Student Faculty of Biology, Georg-August-Universität Göttingen, Germany
09/99 - 06/01	Master of Science Department of Chemistry, National ChungHsing University, Taiwan
Major:	Analytical Chemistry
Thesis:	Study of the Renal Function: Utilizing HPLC/MS to Analyze Creatine, Creatinine and Guanidineacetic acid in Biofluids. Develop a Sensitive Analytical Technique to Monitor Trace Ether in Water.
09/95 - 06/99	Bachelor of Science Department of Chemistry, National ChungHsing University, Taiwan

III. PROFESSIONAL EXPERIENCE

07/01 - 01/07	Research Assistant / Mass Spectrometry Laboratory Specialist Core Facilities for Proteomics Research, Institute of Biological Chemistry, Academia Sinica, Taiwan

IV. PUBLICATIONS

1. **Hsiao, H. H.**, Urlaub, H. "Efficient Enrichment of Intact Phosphoproteins prior to Mass Spectrometric Analysis." ***In Preparation***.
2. **Hsiao, H. H.**, Urlaub, H. **"**Pseudo-Neutral-Loss Scan for Selective Detection of Phosphopeptides and N-Glycopeptides using Liquid Chromatography Coupled with a Hybrid Linear Ion-Trap / Orbitrap Mass Spectrometer." ***PROTEOMICS,*** Technical Brief, ***Accepted***.
3. Wu, H. J., Seib, K. L., Srikhanta, Y. N., Edwards, J., Kidd, S. P., Maguire, T. L., Hamilton, A.,

Pan, K. T., **Hsiao, H. H.**, Yao, C. W., Grimmond, S. M., Apicella, M. A., McEwan, A. G., Wang, A. H., and Jennings, M. P. (2010) “Manganese regulation of virulence factors and oxidative stress resistance in Neisseria gonorrhoeae.” ***JOURNAL OF PROTEOMICS*** 73, 899-916.

4. Schneider, M., **Hsiao, H. H.**, Will, C. L., Giet, R., Urlaub, H., and Luhrmann, R. (2010) “Human PRP4 kinase is required for stable tri-snRNP association during spliceosomal B complex formation.” ***NATURE STRUCTURAL & MOLECULAR BIOLOGY*** 17, 216-221.
5. Kuhn-Holsken, E., Lenz, C., Dickmanns, A., **Hsiao, H. H.**, Richter, F. M., Kastner, B., Ficner, R., and Urlaub, H. (2010) “Mapping the binding site of snurportin 1 on native U1 snRNP by cross-linking and mass spectrometry.” ***NUCLEIC ACIDS RESEARCH***, In Press.
6. Yu, S. Y., Wu, S. W., **Hsiao, H. H.**, and Khoo, K. H. (2009) “Enabling techniques and strategic workflow for sulfoglycomics based on mass spectrometry mapping and sequencing of permethylated sulfated glycans.” ***GLYCOBIOLOGY*** 19, 1136-1149.
7. Oellerich, T., Gronborg, M., Neumann, K., **Hsiao, H. H.**, Urlaub, H., and Wienands, J. (2009) “SLP-65 phosphorylation dynamics reveals a functional basis for signal integration by receptor-proximal adaptor proteins.” ***MOLECULAR & CELLULAR PROTEOMICS*** 8, 1738-1750.
8. Martin Richter, F., **Hsiao, H. H.**, Plessmann, U., and Urlaub, H. (2009) “Enrichment of protein-RNA crosslinks from crude UV-irradiated mixtures for MS analysis by on-line chromatography using titanium dioxide columns.” ***BIOPOLYMERS*** 91, 297-309.
9. **Hsiao, H. H.**, Meulmeester, E., Frank, B. T., Melchior, F., and Urlaub, H. (2009) “"ChopNSpice," a mass spectrometric approach that allows identification of endogenous small ubiquitin-like modifier-conjugated peptides.” ***MOLECULAR & CELLULAR PROTEOMICS*** 8, 2664-2675.
10. Meulmeester, E., Kunze, M., **Hsiao, H. H.**, Urlaub, H., and Melchior, F. (2008) “Mechanism and consequences for paralog-specific SUMOylation of ubiquitin-specific protease 25.” ***MOLECULAR CELL*** 30, 610-619.
11. Luo, X., **Hsiao, H. H.**, Bubunenko, M., Weber, G., Court, D. L., Gottesman, M. E., Urlaub, H., and Wahl, M. C. (2008) “Structural and functional analysis of the E. coli NusB-S10 transcription antitermination complex.” ***MOLECULAR CELL*** 32, 791-802.
12. **Hsiao, H. H.**, Hsieh, H. Y., Chou, C. C., Lin, S. Y., Wang, A. H., and Khoo, K. H. (2007) “Concerted experimental approach for sequential mapping of peptides and phosphopeptides using C18-functionalized magnetic nanoparticles.” ***JOURNAL OF PROTEOME RESEARCH*** 6, 1313-1324.
13. Chen, Y. Y., Lin, S. Y., Yeh, Y. Y., **Hsiao, H. H.**, Wu, C. Y., Chen, S. T., and Wang, A. H. (2005) “A modified protein precipitation procedure for efficient removal of albumin from serum.” ***ELECTROPHORESIS*** 26, 2117-2127.

14. Tsai, J. M., Wang, H. C., Leu, J. H., **Hsiao, H. H.**, Wang, A. H., Kou, G. H., and Lo, C. F. (2004) "Genomic and proteomic analysis of thirty-nine structural proteins of shrimp white spot syndrome virus." ***JOURNAL OF VIROLOGY*** 78, 11360-11370.
15. Chung, T. L., **Hsiao, H. H.**, Yeh, Y. Y., Shia, H. L., Chen, Y. L., Liang, P. H., Wang, A. H., Khoo, K. H., and Shoei-Lung Li, S. (2004) "In vitro modification of human centromere protein CENP-C fragments by small ubiquitin-like modifier (SUMO) protein: definitive identification of the modification sites by tandem mass spectrometry analysis of the isopeptides." ***JOURNAL OF BIOLOGICAL CHEMISTRY*** 279, 39653-39662. (equal contribution as the first autho)
16. Lee, C. L., **Hsiao, H. H.**, Lin, C. W., Wu, S. P., Huang, S. Y., Wu, C. Y., Wang, A. H., and Khoo, K. H. (2003) "Strategic shotgun proteomics approach for efficient construction of an expression map of targeted protein families in hepatoma cell lines." ***PROTEOMICS*** 3, 2472-2486.
17. **Hsaio, H. H.**, Lee, M. R., (2000) "Purge and Trap Extraction Technique." **Instruments Today** 119, 22 (3), 83-91

V. POSTER PRESENTATIONS

1. **Hsiao, H. H.**, Urlaub, H. "Efficient enrichment of intact phosphoproteins prior to mass spectrometric analysis." 58th ASMS Conference, May 23- 27, 2010, Salt Lake City, Utah.
2. **Hsiao, H. H.**, Meulmeester, E., Frank, B. T. C., Melchior, F., Urlaub, H. "Mapping endogenous SUMO sites: A novel approach using ESI-MS and modified database search with common search engines." 57th ASMS Conference, May 31- June 4, 2009, Philadelphia, Pennsylvania.
3. Richter, F., **Hsiao, H. H.**, Paul, N., Luo, X., Agafonov, D., Lührmann, R., Wahl, M., Zavolan, M., Urlaub, H. "An Efficient Strategy for the Identificiation of UV-induced Protein-RNA Crosslinks by ESI Mass Spectrometry." 57th ASMS Conference, May 31- June 4, 2009, Philadelphia, Pennsylvania.
4. **Hsiao, H. H.**, Gronborg, M., Lührmann, R., Urlaub, H. "A high-throughput method for phosphopeptide enrichment of spliceosomal proteins." 56th ASMS Conference, June 1- 5, 2008, Denver, Colorado.
5. **Hsiao, H. H.**, Hsieh, H. Y., Lin, S. Y., Chou, C.C., Wang, A.H., Khoo, K. H. "Highly specific enrichment of phosphorylated peptides from peptides mixtures utilizing magnetic nanoparticles." 54th ASMS Conference, May 28- Jun 1, 2006, Seattle, Washington.
6. Lin S. Y., **Hsiao, H. H.**, Chen, Y. Y., Fan, Y. Y., Lin, C. W., Yeh, Y. Y., Hsieh, H. Y., Chen, S. T., Khoo, K. H., Wang, A. H. "Proteomic and Glycomic Analyses of Serum and Ascitic Fluids in Relation to Colonic Tumor Progression in Inbred Mouse Model." HUPO 4th Annual World Congress, Aug 29-Sep 1, 2005, Munich, Germany.
7. Wu, H. J., Pan, K. T., **Hsiao, H. H.**, Chen, Y. H., Jennings, M. P., McEwan, A. G., Yao, C. W., Wang, A. H. "A Proteomic Study of Manganese Regulation of Gene Expression in

Neisseria Gonorthoeae" HUPO 4th Annual World Congress, Aug 29-Sep 1, 2005, Munich, Germany.

8. **Hsiao, H. H.**, Lin, S. Y., Chen, Y. Y., Fan, Y. Y., Lin, C. W., Yeh, Y. Y., Hsieh, H. Y., Chen, S. T., Khoo, K. H., Wang, A. H. "Concerted Serum Proteomics and Glycomics Mapping of Colonic Tumor Progression in Inbred Mouse Mode." 53th ASMS Conference, Jun 5-9, 2005, San Antonio, Texas.
9. Lin, S. Y., Lee, C. L., **Hsiao, H. H.**, Shih, R. Y., Yeh, Y. Y., Wu, C. Y., Liao, C. L., Wang, A. H. "Proteomic Analysis of Human Plasma/Serum." HUPO 2nd Annual World Congress, Oct 8-11, 2003, Montreal, Canada.
10. Lee, C. L., Shyu, W. T., Lin, C. Y., **Hsiao, H. H.**, Chen, C. L., Wu, C. Y., Wu, Y. T., Khoo, K. H., Wang, A. H. "Strategic Experimental Approach Towards Establishing Model Hepatoma Cell Lines for the Global Human Liver Proteome Initiatives" HUPO 1st Annual World Congress, Nov 21-24, 2002, Verailles, France.
11. Kuo, C. W., **Hsiao, H. H.**, Shyu, W. T., Chen, C. M., Lee, Y. C., Chu, S. T., Khoo, K. H. "Proteomic and Glycomic Analyses of Mouse Uterine Luminal Fluid." HUPO 1st Annual World Congress, Nov 21-24, 2002, Verailles, France.
12. **Hsiao, H. H.**, Lin, H. J., Yeh, Y. Y., Wang, A. H., Khoo, K. H. "Mass Spectrometry Based Strategies for Analysis of Phosphoprotein and Identification of Phosphorylation Site." 3rd Autumn Symposium of Taiwan Society for Biochemistry and Molecular Biology, Tainan, Nov 14-16, 2003.
13. **Hsiao, H. H.**, Lee, M. R. "Analysis of Creatine and Creatinine in Serum and Urine by LC/MS." 2000 Annual Meeting of Chemical Society Located in Taipei.
14. **Hsiao, H. H.**, Lee, M. R. "Study of Quinolinic Acid by Mass Spectrometry." 8th Analytical Chemistry Techniques Conference Located in Hsinchu.
15. **Hsiao, H. H.**, Lee, M. R. "Solid Phase Micro-Extraction Coupled with Cryo-Trap GC/MS for the Analysis of Methyl-t-Butyl Ether (MTBE) in Water." 1999 Annual Meeting of Chemical Society Located in Taipei.