We are IntechOpen,
the world's leading publisher of Open Access books
Built by scientists, for scientists

4,800
Open access books available

122,000
International authors and editors

135M
Downloads

154
Countries delivered to

Our authors are among the
TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Proteomic Strategies for Comprehensive Identification of Post-Translational Modifications of Cellular Proteins Including Low Abundant and Novel Modifications

Jaeho Jeong, Jae-Jin Lee and Kong-Joo Lee
*The Center for Cell Signaling & Drug Discovery Research, College of Pharmacy,
Division of Life & Pharmaceutical Sciences, Department of Bioinspired Science, Ewha
Womans University,
Republic of Korea*

1. Introduction

Post-translational modifications (PTMs) enable proteins carryout multiple biological functions. PTMs facilitate and regulate various protein functions by changing the chemical and physical characteristics of a protein, which affect its stability, cellular localization, and its interaction with other proteins as well as non-protein molecules which endow the protein-protein interactions, cellular localization and with new or altered biological activities. Understanding how and to what extent cellular proteins are post-translationally modified, at which sites on the protein, how the modifications affect the stoichiometry of the protein sequence, and what the functional consequences of each of the modifications are, therefore becomes indispensible for understanding cellular function and regulation of a protein. Recent advances that facilitated combining affinity-based enrichment of low abundant and novel PTMs with peptide sequencing with tandem mass spectrometric analysis have increased its speed and sensitivity of analytical approaches. Although it has been demonstrated that mass spectrometry can be an ideal tool for both qualitative and quantitative analysis of protein modifications, comprehensive identification of PTMs using mass spectrometry (MS) in a high-throughput manner remains a highly challenging task because of the diversity, dynamic complexities, low abundance and heterogeneity of PTMs. In addition, difficulties in interpreting tandem MS spectra for peptide sequencing, poor peptide fragmentation, and surprising appearance of totally unexpected modifications, among others, have compounded this challenge and limited the application of mass spectrometry to the identification of a few types of PTMs.

We will describe in this chapter, a strategy we have been employing successfully for rapid, efficient and sensitive identification of diverse and novel PTMs occurring in mammalian biological systems *in vivo*. Our strategy includes: separation of unmodified and modified proteins on 2-dimensional gel electrophoresis (2D-PAGE); and detection of low abundant peptide modifications employing a selectively excluded mass screening analysis (SEMSA) of unmodified peptides during LC-ESI-q-TOF MS/MS through replicated runs (Seo *et al.*, 2008), in conjunction with a searching algorithm MOD[i] (Kim *et al.*, 2006; Na *et al.*, 2008) and

DBond (Choi *et al.*, 2010), which we developed. We will also provide examples of novel some low abundant cysteine modifications which we identified in cellular proteins employing our strategy (Lee *et al.*, 2009; Hwang *et al.*, 2009; Jeong *et al.*, 2011).

Since it is difficult in many cases, to predict the type and position of modifications from autotranslated DNA sequences, it is highly important in the postgenomic era, to experimentally obtain this information about PTMs. For, the type and position of PTMs differ, depending on the organism, cell type, signaling process and its kinetics. Proteomics allows deciphering the global picture of protein-related processes in the cell. But proteomics cannot be successfully employed without defining the dynamic PTM maps, which in turn requires the availability of a fast, reliable, and sensitive procedure for PTM characterization. We believe that the highly sensitive and efficient strategies for unrestrictive blind PTM identification that will be described in this chapter will help develop such PTM maps.

2. Diversity of post-translational modifications and their identification using mass spectrometry

Structural changes in proteins can occur from various events including genomic mutations, alternative splicing, proteolytic cleavage, chemical and enzymatic modifications in amino acid side chains. These changes occurring inside cells and tissues cannot be recognized in genomic level, other than by proteomic analysis which makes it possible to identify these alternations. Since these modifications are highly diverse, many proteomic approaches are required to identify them. This review will describe the diverse characteristics of PTMs in proteins and the analytical approaches including mass spectrometry employed to identify them.

2.1 Cellular protein modifications

Modifications in the side chain of amino acids of proteins can alter the protein's charge, polarity and spatial features, and induce conformational changes which in turn cause changes in a variety of protein characteristics. These modifications can occur reversibly or irreversibly, and be mediated by enzymes or non-enzymatic way. In addition, many changes at the genomic level such as point mutation, deletion and insertion, alternative splicing etc. induce alterations in amino acid sequences. Identification of these changes may require different analytical MS methodologies, depending on the type of the modification.

2.1.1 Amino acid substitution

Amino acid substitution, which can occur by point mutation in the genome, or chemical conversion of one amino acid to another, is involved in 346 of the listed 900 modifications in the database of protein modifications for mass spectrometry, UniMod (www.unimod.org). The Human Gene Mutation Database (www.hgmd.org), as of August 2011, lists more than 113,000 mutations of 4,122 human genes, which include missense/nonsense mutation, splicing, small and gross deletion/insertion, complex rearrangement and repeat variation. Although a single amino acid substitution usually does not seem affect a protein's structure in most cases, several proteins are known in which a change in single amino acid change caused significant changes in the protein structure and stability. An example is a single mutation in hemoglobin causes sickle cell anemia, one of several monogenic diseases.

Another way to occur amino acid substitution is by chemical conversion of one amino acid to another, for example, Asn and Gln can be converted to Asp and Glu by deamination. Cys residues are converted to Ser during oxidative stress (Jeong *et al*., 2011). Mass spectrometry has been extremely useful in identifying unexpected amino acid substitutions and deletion and insertion of amino acids (Fig. 1). However, finding unknown substitutions with routine peptide sequencing in conjunction with tandem MS is not to easy, because MS/MS spectra searching algorithm matches only known sequences. In order to find these unexpected amino acid changes, in MS/MS spectra, it is necessary to carry out *de novo* partial sequencing of a peptide from an MS/MS spectrum. To find unknown amino acid changes, we suggest using the searching algorithm MOD[i] combined with peptide sequencing using tandem MS.

2.1.2 Alternative splicing

Insertion and deletion of large amino acid sequences can markedly affect the protein structure, in contrast to changes in single amino acid residues which mostly do not. Alternative splicing can occur in a number of ways including exon skipping, exon insertion, alternative 5′ initiation, 3′ termination and even intron inclusions. About 74% of multi-exon genes produce alternatively spliced protein variants in humans. These PTMs result in proteins differing in binding affinity, enzymatic activity, localization within the cell, and stability. Alternative splicing therefore, is a common phenomenon in biology. Existence of the alternatively spliced variants can be detected only at the protein level. However, it is not easy to find spliced variants without separating each variant population. 2D-PAGE is an useful tool to separate the different variants because of their differences in their charges (isoelectric points) and molecular weights (Fig. 1). Separation of each population of protein variants is a prerequisite for employing mass spectrometry to identify the sequence changes. By comparing the chromatogram of each population, it is possible to find peptide products generated from alternative splicing, and then *de novo* sequencing the differential peptides.

2.1.3 Proteolytic cleavage

Proteolytic cleavage is obviously an irreversible protein modification. Proteins can be cleaved by various proteolytic enzymes, which are classified into serine proteases, cysteine proteases, aspartic acid proteases and metalloprotease, depending on their active sites of action. Employing genome sequence analysis, 553 proteases were identified in the human genome: 203 Ser proteases, 143 Cys proteases, 21 Asp proteases and 186 metalloproteases. Since the substrate specificity of each protease is not known in many cases, the proteolytic products can be further identified only at protein level. Unexpected cleavage products are not easy to detect without mass spectrometry (Fig. 1). Full peptide sequencing with tandem MS is the only way to characterize the cleaved products using the same strategy employed to detect the alternative splicing variants.

2.1.4 Enzymatic modifications

Cellular protein modifications are designed by nature to initiate and regulate essential cellular processes. The mechanisms for PTM regulation are not fully understood because of their complexity. More than 200 kinds of enzymes (>5% of total proteins) have been shown to be involved in catalyzing the various chemical modifications of protein side chains. In the human genome, it has been shown that more than 500 proteases, more than 500 protein

```
Wild type              ABCDE FGHIJ KLMNO PQRST UVWXY Z

AA substitution        ABCDE XGHIJ KLMNO PQRST UVWXY Z
AA deletion            ABCDE -GHIJ KLMNO PQRST UVWXY Z
AA insertion           ABCDE FXGHIJ KLMNO PQRST UVWXY Z

Alternative splicing         FGHIJ KLMNO PQRST UVWXY Z
                       ABCDE -------- KLMNO PQRST UVWXY Z

Proteolytic cleavage              HIJ KLMNO PQRST UVWXY Z
                                       ST UVWXY Z
```

Fig. 1. Protein amino acid sequence changes with various ways generate unexpected peptide fragments in MS analysis.

kinases, more than 150 protein phosphatases, 5 class methyltransferases, a series of acetyltransferase and deacetylase, oxidoreductases, E1, E2, E3 for ubiquitination, sumoylation and neddylation, operate among others. Since most enzyme induced protein modifications are reversible, they are readily removed during the biological processes after they function as signalling molecules.

Enzymatic modification of amino acid side chains occurs in various ways depending on the species of amino acids, as shown in Table 1. For example, phosphorylation at –OH in Ser, Thr and Tyr residues, can be promoted by various protein kinases and one third of total human proteome is estimated to comprise substrates of various proteins kinases, predicted to number more than 500. Phosphorylated proteins serve as substrates of kinases. They are readily dephosphorylated by phosphatases. Tyrosine kinase receptors (e.g. PDGF receptor, VEGF receptor) are auto-phosphorylated by ligand binding and transduce the signalling as turn 'on' switch, and later dephosphorylated by phosphatases, as turn 'off' switch. These reversible modifications act as on/off switches in various signalling pathways by controlling phosphorylation and dephosphorylation. Thus it becomes important to identify, which residues are phosphorylated, which kinase acts on them, how long signalling is turned on and which phosphatases are involved in the turn off.

In addition to kinases and phosphatases which signal phosphorylation/dephosphorylation, other enzyme pairs are involved in reversible modifications; e.g. acetyltransferase and deacetylase, acyltransferase and deacylase, methyltransferase and demethylase, ubiquitinating and deubiquitinating enzymes, chemical oxidation and peroxidase and reductase, glycosylase and deglycosylase, among others. Irreversible modifications are also possible, for example, crosslinkiing of the proteins by transglutaminase. Although many modifications have been identified thus far, an enormous number of unknown protein modifications are waiting to be discovered. Mass spectrometry offers an useful analytical approach to identifying and quantifying the unknown novel modifications, modification sites, and the accompanying structural changes. The relevant methodologies are detailed in section 2.2.

2.1.5 Non-enzymatic chemical modifications

Protein modifications can occur inside cells by chemical reactions rather than by enzymatic action. Oxidation of Cys, Met, Pro, Trp residues is promoted by reactive oxygen species (ROS) resulting in loss of protein activity or alteration of the protein's biological function by modifying its cellular localization and interactions with other proteins. Nitrosylation,

formation of metabolic protein adducts e.g. 4-hydoxynonenal (HNE) adducts involving His, Cys, Lys, chemical adducts with biotin, lipoic acid and phosphopantetheine are well defined modifications (Table 1). Many metabolite adducts of proteins will be further investigated using *de novo* sequencing using tandem MS and searching algorithm.

PTM type	Modified amino acid residue	Monoisotopic mass change	Remarks
Phosphorylation	S T Y H D	+79.966331	Appearance of ΔM = -97.976896 Da peak for dehydration in the phosphorylation sites of Ser and Thr. Use phosphatase inhibitor in sample
Cysteine oxidation	C		Careful for artifactual oxidation
Dehydroalanine		-33.987721	
Cys to Ser		-15.977156	
Disulfide		-2.0145	
Sulfenic acid		+15.994915	
Sulfinic acid		+31.98983	
Sulfonic acid		+47.984745	
Thiosulfonic acid		+63.961901	
Selenylation		+79.916520	Cys-S-SeH
C-nitrosylation		+28.990164	
Glutathionylation		+305.068156	
Acetylation	K N-term	+42.010565	Change (+) charge to neutral Use deacetylase inhibitors in sample
Acylation			Increase hydrophobicity of protein to be anchored into membrane
Octanoylation	S T	+126.104465	
Farnesylation	C	+204.187801	
Palmitoylation	C K N-term	+238.229666	
Geranyl-geranylation	C K N-term	+272.250401	
Retinal	K	+266.203451	
Diacylglycrol	C	+576.511761	
Ubiquitination	K	+114.042927	Ubiquitin C-term Gly-Gly adduct Use proteasome inhibitor in sample
Glycosylation			
Fucose	S T	+146.057909	
Hexoseamine	N T	+161.068808	
Hexose	N T	+162.052824	
O-GlcNAcylation	N S T	+203.0743	N-acetylhexoseamine
Methylation			Sterically bulky without charge change
Mono-methylation	K R	+14.01565	
Di-methylation	K R	+28.0313	
Tri-methylation	K R	+42.04695	
Sulfonation	S T Y C	+79.956815	
Deamidation	N Q	+0.984016	Amino acid substitution to D E
Chemical adduct			
Allysine	K	-1.031634	Lys oxid to aminoadipic semialdehyde
Oxidation	P D K N P Y R	+15.994915	Oxidation or hydroxylation
Hypusine	K	+87.068414	
4-hydroxynonenal	C H K	+156.115030	HNE
Lipoyl	K	+188.032956	
N-term			
Met-loss	M	-131.040485	N-term Met removal by aminopeptidase
Met-loss + Acetyl	M	-89.029920	Acetylation in new N-term after Met loss
Artifactual adduct			
Propionamide	C K N-term	+71.037114	Acrylamide adduct in PAGE
Carbamidomethyl	C K N-term H D	+57.021464	Iodoacetamide derivative

Table 1. The list of identified post-translational modifications.

2.2 Identification of post-translational modifications by mass spectrometry

Protein modifications play key role in protein structure, stability, its interactions and cellular localizations. MS is an ideal analytical tool for analyzing protein sequences, identifying many unknown post-translational modifications, sites modified, and amino acid sequences inserted, deleted or repalced. Studies employing MS along with other techniques, have recorded more than 900 PTMs in UniMod database (www.unimod.org). Number of PTM list in this database is rapidly increased. This informations on the nature of PTMs should facilitate a fuller understanding protein structure and function. A list of reported PTMs and the corresponding sites of amino acids, monoisotopic mass change by PTM, and sample preparation for identifying each PTM is shown in Table 1. These proteomic studies for PTM analysis can be divided into two groups. One group consists of large scale analysis of one kind PTM after enriching the modified proteins and identifying PTMs, necessitated by low abundance of modified proteins. The second group comprises comprehensive analyses of multiple modifications in one protein because the diversity of modifications in the chosen protein (Fig. 2).

2.2.1 Large scale analysis of same type of PTMs in many proteins

Large scale identification of PTMs is not easy problem to be solved because each modification has its distinct and unique chemical property including chemical affinity, solubility, charge and hydrophobicity. In order to identify one kind of PTM in many proteins with high throughput analysis, the enrichment of modified proteins or peptides was combined with peptide sequencing with tandem MS. Affinity-based enrichment of post-translationally modified proteins and peptides can provide to increase the relative abundance of a selected PTM in the sample.

Several studies have reported the systemic identification of same PTM to understand wide signaling cascade (Table 2). Phosphorylation, the most extensively studied PTM, has been reported in more than 10,000 proteins, variously concerned with specific signaling, in cell cycle, cancer, receptor function, and stress responses (Olsen *et al.*, 2010; Dephoure *et al.*, 2008; Rikova *et al.*, 2007; Kim *et al.*, 2002, Kim *et al.*, 2007a and 2007b). Intriguingly, although phosphorylation seems to play a pivotal role in various signal cascades, phosphorylated peptides constitute minor fraction of the total protein mileau. Due to the low abundance of phosphopeptides and low degrees of phosphorylation, enrichment phosphopeptides is essential prior to MS analysis (Fig. 2A). There are several strategies for enrichment of phosphoproteins, including immobilized metal ion affinity chromatography (IMAC), titanium dioxide (TiO_2) chromatography, and immunoaffinity chromatography with anti-pY antibodies (Thingholm *et al.*, 2009). These enrichment techniques themselves have advantages and disadvantages, related to different specificities of the proteins under study. Immunoaffinity chromatography is a traditional biochemical technique, based on the binding of proteins or peptides to phospho-specific antibodies. Highly selective phospho-specific antibodies are available that can be utilized for the enrichment of phosphorylated proteins prior to analysis by MS. IMAC takes advantage of the affinity of chelated (Fe^{3+}, Al^{3+}, Ga^{3+}, or Co^{2+}) ions towards the negative phosphate group of phosphopeptides. Titanium dioxide chromatography utilizes the affinity for phosphate ions in aqueous solutions. Titanium dioxide chromatography sometimes can be combined with IMAC in which mono- and multiply phosphorylated peptides are efficiently enriched from highly complex samples. When combined with other enrichment methods, the method offers an efficient enrichment strategy for phosphoproteomic studies. In another approach, cellular

Proteomic Strategies for Comprehensive Identification
of Post-Translational Modifications of Cellular Proteins Including Low...

93

proteins with or without treatment, were separated on 2D-PAGE and phosphorylated proteins on Tyr residues were detected by anti-pY antibody, and each spot was analyzed with tandem MS (Kim *et al.*, 2007a and 2007b). Separation of modified protein from abundant unmodified proteins on 2D-PAGE is an informative way to obtain only modified proteins.

However, attempts to enrich other PTMs have met with limited success. Acetylation of Lys residue is another abundant PTM with fundamentally important regulatory function. Acetylation occurs as a co-translational and post-translational modification of proteins, such as histones, p53 and tubulins. The acetyl group can become attached to either the α-amino group at the protein N-terminus or the ε-amino groups of Lys residues and eliminate the positive charge of the amino group to make it uncharged. Thus, Lys acetylation, a reversible PTM, regulates protein interaction with negatively charged DNA, thereby playing a key regulatory role in gene expression. For example, acetylation of histone or p53 inhibits the DNA binding and renders DNA more relaxed, and deacetylation reverses this process. However, enrichment methods, except for immunoaffinity purification, have not been developed thus far. There remains a need for robust methods that can address the complexity and dynamic range of the cellular proteome.

In contrast to phosphorylation and acetylation as small chemical modifications, ubiquitination, a PTM involving the covalent attachment of ubiquitin, a 76-residue polypeptide, to Lys residue, induce bulk change of protein. Depending on the nature and site of linkage, ubiquitination regulates protein degradation, signal transduction, intracellular localization and DNA repair. Ubiquitination is readily detected by peptide sequencing with tandem MS, because Gly-Gly adduct of ubiquitin C-terminal (+114 Da increase) can easily be detected at Lys ubiquitinated residues in tryptic peptides. Most common enrichment methodology for ubiquinated proteins is immunoaffinity purification employing exogenously tagged ubiquitin (Danielsen *et al.*, 2011). Recent studies have employed endogenous ubiquitin enrichment using ubiquitin binding motifs such as ubiquitin interacting motif (UIM) or ubiquitin associated (UBA) domain, rather than Ub-antibody (Manzano *et al.*, 2008). About 200 ubiquitinated proteins binding to UBA domain of p62 were identified in Arabidopsis.

Glycosylated proteins exist as heterogenous populations characterized by a range of molecular weight, and play important role in membrane surface localization and as receptors by raising the hydrophilicity and changing the surface charge. The O-linked β-N-acetylglucosamine (O-GlcNAcylation) also plays a vital role modifying Ser and Thr residues in many cellular processes, including signal transduction, protein degradation, and regulation of gene expression. Since the lectin, Wheat Germ Agglutinin (WGA), has affinity for terminal N-acetylglucosamine (GlcNAc) and sialic acid residues, lectin immobilized affinity chromatography has been recently used for enrichment O-GlcNAc modified peptides (Vosseller *et al.*, 2006). Biotinylated O-GlcNac peptides are captured by avidin chromatography (Wang *et al.*, 2010).

Oxidative modification has emerged as a major PTM involved in oxidative stress. 4-Hydroxy-2-nonenal (HNE), generated during lipid peroxidation, modifies proteins. 4-HNE adducts are commonly enriched by immunoaffinity chromatography or solid-phase hydrazide enrichment strategy (Mendez *et al.*, 2010; Roe *et al.*, 2007). Other oxidation adducts of reactive Cys residues play key roles in cellular regulations. Recently dimedone, and chemicals that specifically label sulfenic acid, have been used to enrich sulfenic acid using biotin tagged dimedone, and to identify ROS sensitive Cys residues (Giron *et al.*, 2011,

Seo YH *et al.,* 2009, Leonard *et al.,* 2009). Large scale proteomic identifications of proteins containing the same kind of PTM, e.g., phosphorylation or acetylation, can be interesting if the relationship between a specific modification and biological function can be established.
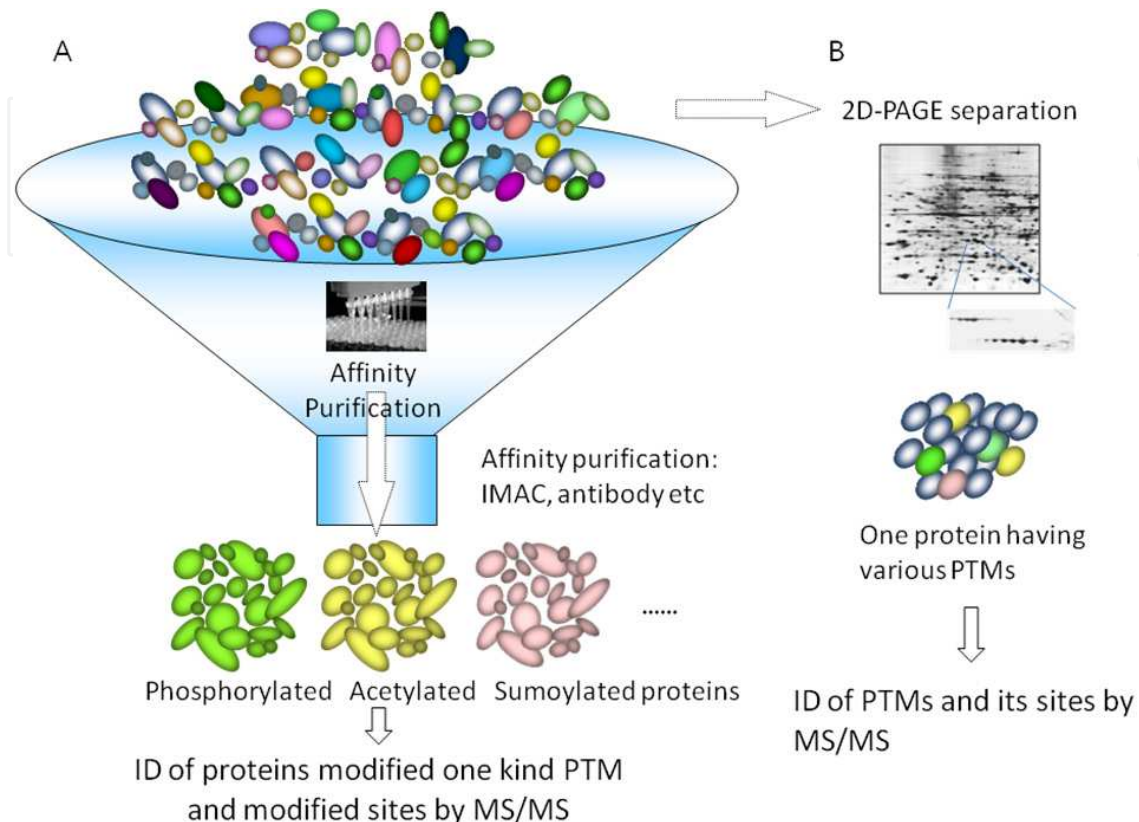

Fig. 2. Schematic diagram to identify one type modification in cellular proteins (A) and comprehensive modifications in one protein (B).

2.2.2 Comprehensive identification of multiple PTMs in the same protein by MS

If the same protein has multiple and diverse functions, one can hazard the assumption that it exists in several forms and contains different PTMs. Comprehensive identification of PTMs in a single protein can therefore help understand how the protein exerts multiple biological functions of multiply modified proteins as shown in Fig. 2B. However, it is not an easy task to clearly identify the PTMs in the same protein, because biological samples of proteins are mixtures of unmodified and modified populations, with unmodified molecules abundantly predominating and the much less abundant modified molecules. A 100% peptide coverage with MS/MS is required for identifying all modifications.

The low abundant PTMs can only be identified after adequately enriching their populations. When complex enrichment methods are needed, use of 2D-PAGE based separation in combination with mass spectrometry is beneficial. 2D-PAGE separates proteins based on their isoelectric points and molecular mass, and makes it possible to separate various modified proteins, spliced variants and proteolytic cleaved fragments. Proteomics has been traditionally exploited power of 2D-PAGE, to separate proteins, especially coupling it with MALDI-TOF MS, for the qualitative and quantitative analysis of proteins in complex extracts. However, the limitations of this approach in terms of throughput analysis of protein mixes have required the development of other proteomics approaches, based on

Proteomic Strategies for Comprehensive Identification
of Post-Translational Modifications of Cellular Proteins Including Low...

95

separation of peptides rather than of proteins, or on direct protein identification and selection on dedicated arrays (protein chips). However, 2D-PAGE based separation is adequate in the comprehensive identification of PTMs because it can show the range of complex PTMs.

PTM type (Mass Shift)	Enrichment methods	# of sites identified	Biological Association of the protein	References
Phosphorylation (79.9663)	IMAC, and TiO₂	20,443 sites on 6027 proteins	cell cycle	Olsen *et al.*, 2010
		1400 sites on >1000 proteins	Mitosis	Dephoure *et al.*, 2008
	IMAC	654 sites	cAMP/vasopressin-dependent signaling pathways	Gunaratne *et al.*,2010
	Immuno-affinitiy purification with	4551 sites on 2550 proteins	Lung cancer	Rikova *et al.*, 2007
	Anti-pY Ab	628 sites	Cancer cells	Rush *et al.*, 2005
	2D-PAGE, Anti-pY Ab	81 proteins	Heat shock	Kim *et al.*, 2002
		95 proteins	Endothelial cells VEGF-, ROS-signaling	Kim *et al.*, 2007a
		66 proteins	Angiopoietin-1 signaling	Kim *et al.*, 2007b
Acetylation (42.0105)	Immuno-affinitiy purification with	3600 sites on 1750 proteins	Human acute myeloid leukemia cell line	Choudhary *et al.*, 2009
	Anti-acetyl Lys Ab	1981 sites on 1013 proteins	Drosophila	Weinert *et al.*, 2011
		1000 sites	Human liver tissue	Guan *et al.*, 2010
Ubiquitination (114.0429)	Tagged ubiquitin	753 site 471 proteins	Human osteosarcoma cells	Danielsen *et al.*, 2011
		110 sites 72 proteins		Peng, J. *et al.*, 2003
Sumoylation	Tagged SUMO	339 putative polySUMO	polySUMO conjugates	Bruderer *et al.*, 2011
		17 sites on 205 proteins	HEK293 cells	Galisson *et al.*, 2010
		14 sites on 12 proteins		Blomster *et al.*, 2010
O-GlcNAcylation (203.0793)	Lectin-affinity purification	141 sites on 64 proteins	Cytokinesis	Wang *et al.*, 2010
		65 sites	Postsynaptic density preparation.	Vosseller *et al.*, 2006
4-HNE (156.1150)	Immunoaffinity	12 proteins	Human RBCs	Mendez *et al.*, 2010
	Solid-phase hydrazide enrichment	125 sites on 67 proteins		Roe *et al.*, 2007
	2D-PAGE, MS		Early Alzheimer's disease	Redd *et al.*, 2009
Sulfenic acid (15.9949)	Biotin-dimedone purification	39 proteins	HeLa cells	Leonard *et al.*, 2009
	2D-PAGE	35 proteins	Hypertensive Rat	Tyther *et al.*, 2010
biotinylated 15d-PGJ2 (316.2038)	Streptoavidin affinity	12 proteins	Neuroblastoma	Aldini *et al.*, 2007

Table 2. Large scale identification of each PTM in proteins from divergent sources and pathways.

For example, phosphorylated, acetylated, glycosylated and oxidized proteins move in acidic direction, ubiqutinated and sumoylated proteins move upward by increasing molecular weight and disulfide bonded proteins move either upward with intermolecular disulfide or downward with intramolecular disulfide bond. This information can allow the prediction of the type of PTM and overcomes the limitations due to the complexity of PTMs.

Following the separation of the heterogeneous populations of modified proteins on 2D-PAGE, we try to obtain comprehensive PTM information via replicate nanoLC-ESI MS/MS analysis by raising the modified peptide coverage (Seo *et al.*, 2008). To facilitate the characterization of PTMs as much as possible, we devised the strategy of selective exclusion acquisition in replicate run analysis. In data dependent acquisition (DDA) mode, most intense precursor ions are redundantly acquired in nanoLC-ESI MS/MS run. If the exclusion list is not used, identification of low-intensity ions in the presence of high-intensity ions would be far less successful in randomly repeated runs. The number of obtainable MS/MS spectra is limited in a single run analysis, because we select MS/MS spectra having appropriate quality and quantity by optimizing the experimental procedure including sensitivity, scan time, number of ion channel, time for return to MS scan from MS/MS scan, elution time in LC etc. This is the reason for selectively excluding unwanted high-intensity MS/MS data generation. Exclusion methodology is a way to separate wanted peptides from unwanted ones. The overall scheme of **S**electively **E**xcluded **M**ass **S**creening **A**nalysis (SEMSA) is shown in Fig. 3. An exclusive implementation using this unmodified peptide library, resulted in efficient identification of low abundant PTMs (Fig. 4). As the SEMSA progressed, exclusion list is cumulated and then separation of unwanted peptide can ameliorate the quality of MS/MS spectra. The LC-MS procedure is repeated three times to obtain more MS/MS data. The MS data of the first run are then processed by ProteinLynx v2.1 for peak deconvolution and peak list generation. The resulting MS/MS spectra are then generated and submitted to Mascot and ProteinLynx database searches to obtain peptide identifications. Only unmodified peptides now serve as candidates for a precursor exclusion list, in terms of m/z and LC run times in the subsequent run. After the peak list of the second separation is generated, the peaks are matched to peptides previously identified and included in the exclusion list are automatically blocked from the peak selection prior to MS/MS acquisition. The ranges of tolerance windows of excluding peak are typically determined by mass accuracy and resolution in the MS scan and peak widths in the chromatogram. This PTM specific exclusion strategy enables less intense PTM peptides to be identified, thereby enhancing confidence level of PTM identifications.

To determine whether the GAPDH spots on 2D-PAGE have differential modifications, we exhaustively examined the PTMs in each population using SEMSA (Seo *et al.*, 2008). Diverse PTM populations were identified in various peptides (Fig. 5). Especially, multiple modifications of Cys residue in the peptide containing active site ([152]CTTNC[156]) were clearly demonstrated: intra-disulfide between 152C and 156C, oxidation to cysteic acid (152C), and transformation of Cys to Ser (152C). Simultaneously, 247C was shown to transform to sulfinic acid, mainly dehydroalanine and cysteic acid. This indicates that reactive Cys residues in GAPDH can be oxidized to various oxidation states depending upon tertiary structural environments. Results of these studies clearly established the exact oxidation sites, oxidation species, and the levels of oxidation states. This strategy was applied for finding many low abundant modifications including phosphorylation, acetylation, glutathionylation and some novel modifications (Hwang *et al.*, 2009; Lee *et al.*, 2010; Jeong *et al.*, 2011).

Proteomic Strategies for Comprehensive Identification
of Post-Translational Modifications of Cellular Proteins Including Low...
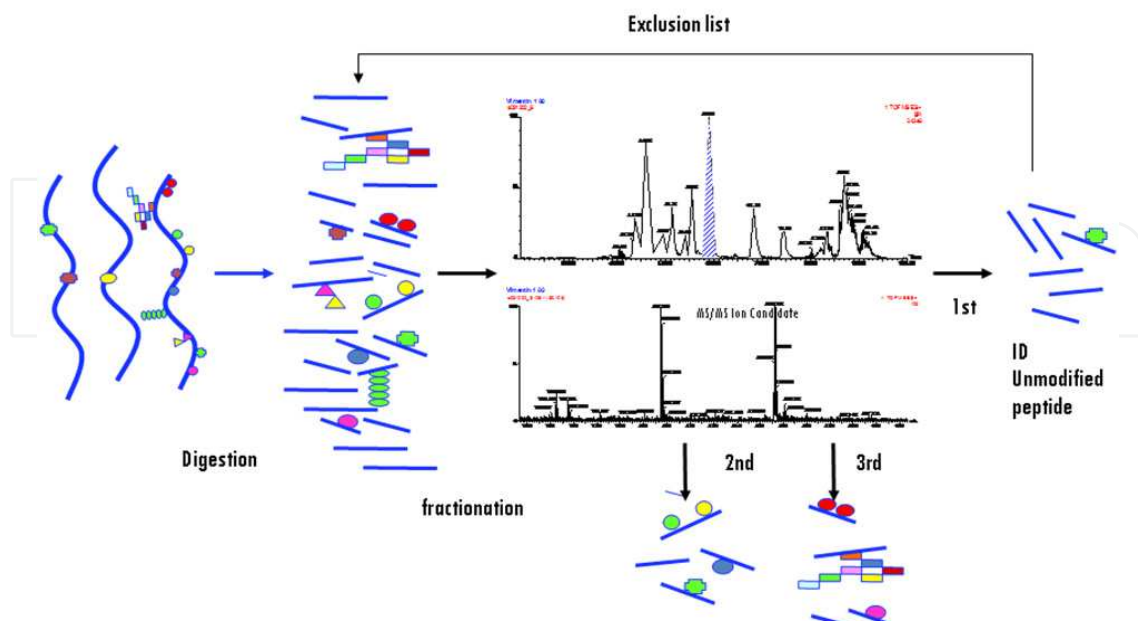
97

Fig. 3. Schematic diagram of SEMSA to enrich the low abundant modified peptides by excluding abundant unmodified peptides obtained in the first identification in the 2nd and 3rd runs.
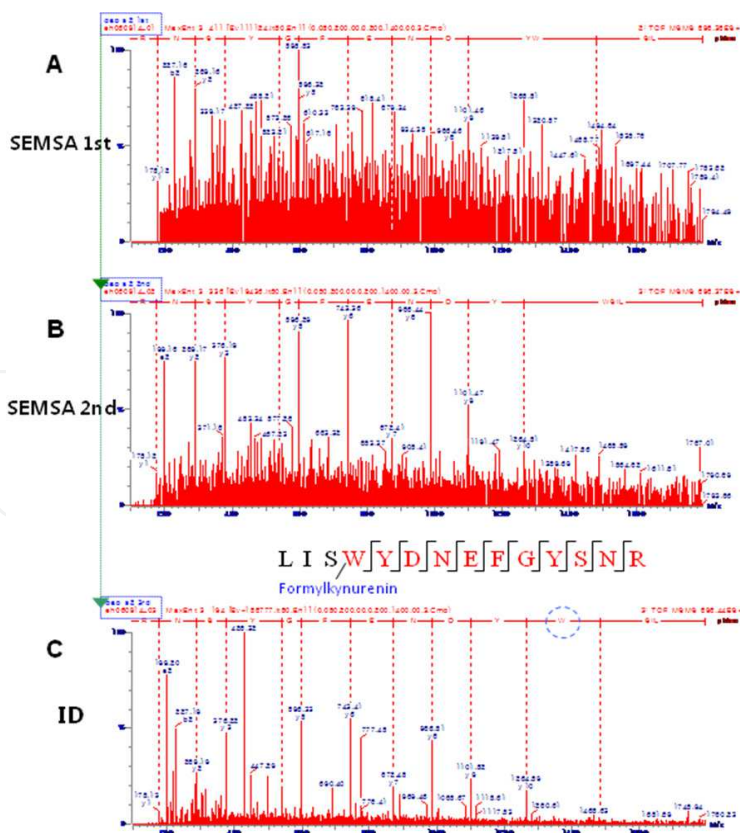

Fig. 4. Comparison of MS/MS spectra of chosen low-density peptides under without (A) and with (B,C) SEMSA strategy (cited from Seo *et al.*, 2008).

Combination of 2D-PAGE for separating modified populations and MS/MS analysis using SEMSA, makes it possible to identify low abundant modified peptides and to raise the identified peptide coverage nearly over 90%.

2.2.3 Bioinformatic tools for identifying multiple and novel modifications

The types and sites of PTMs in a protein vary widely. Although MS allows rapid identification of many types of PTMs, data analysis and interpretation of MS/MS spectra for identification of PTMs remain a major challenge. Most of the available search tools accept only a few types of PTMs as input. We have developed interpretative tools called MOD[i] (Na et al., 2008) and DBond (Choi et al., 2010) for rapidly interpreting tandem mass spectra of peptides with all known types of PTMs simultaneously without limiting a multitude of modified sites.

Early approaches to PTM identification using MS/MS involved exhaustive searches of all possible combinations of PTMs for each peptide from a protein database (Eng et al., 1994; Perkins et al., 1999). Because the search space grows exponentially as the number of PTMs increases, these early approaches performed a restrictive search that took into account only a few types of PTMs during data analysis, ignoring all others. Investigators were obliged to guess the PTMs expected to exist in a sample prior to a search, and many potentially important PTMs may have been overlooked. A few tools were recently introduced for blind PTM search. MS-Alignment (Tsur et al., 2005) predicts PTMs expected in a sample by spectral alignment between a database peptide and a spectrum followed by InsPecT search (Tanner et al., 2005). ModifiComb (Savitski et al., 2006) introduced a ΔM histogram between unassigned spectra and base peptides found in a database. These blind approaches predict PTMs based on the frequency of mass shifts (indicating potential PTMs) in a sample. Thus, they all have the intrinsic weakness of missing rare or infrequently observed PTMs that might provide important clues to understanding the function of a protein. Although many approaches have been developed to take into account several types of PTMs, most of them assume that there will be a single variable PTM per peptide and ignore peptides with multiple modifications. MOD[i] (pronounced "mod eye") is essentially a sequence tag approach (Mann et al., 1994; Tabb et al., 2003) (Fig. 6). It constructs a partial sequence of a peptide from an MS/MS spectrum using de novo sequencing. MOD[i] differs from previous approaches in that it simultaneously uses multiple sequence tags derived from a spectrum by introducing a notion of a tag chain, a combination structure of multiple sequence tags. A tag chain offers an effective localization of modified regions within a spectrum and thus allows rapid identification of multiple PTMs in a peptide, obviating search space explosion by inspecting PTMs only in the modified regions of a peptide.

The tag chain algorithm resists de novo sequencing errors, whereas most tag-based approaches depend critically on good de novo interpretations. This approach is scalable and performs well even when more than 900 types of modification are considered and the number of potential PTMs in a peptide increases. Compared with established tools, MOD[i] reliably identifies a greater variety of modification types in multiply modified peptides and even detects modifications of low abundance.

Another new algorithm called "DBond" analyzes disulfide linked peptides based on specific features of disulfide bonds (Choi et al., 2010). Identifying the sites of disulfide bonds in a protein is essential for thorough understanding of a protein's tertiary and quaternary structures and its biological functions. Disulfide linked peptides are usually identified indirectly by labeling free sulfhydryl groups with alkylating agents, followed by chemical

Proteomic Strategies for Comprehensive Identification
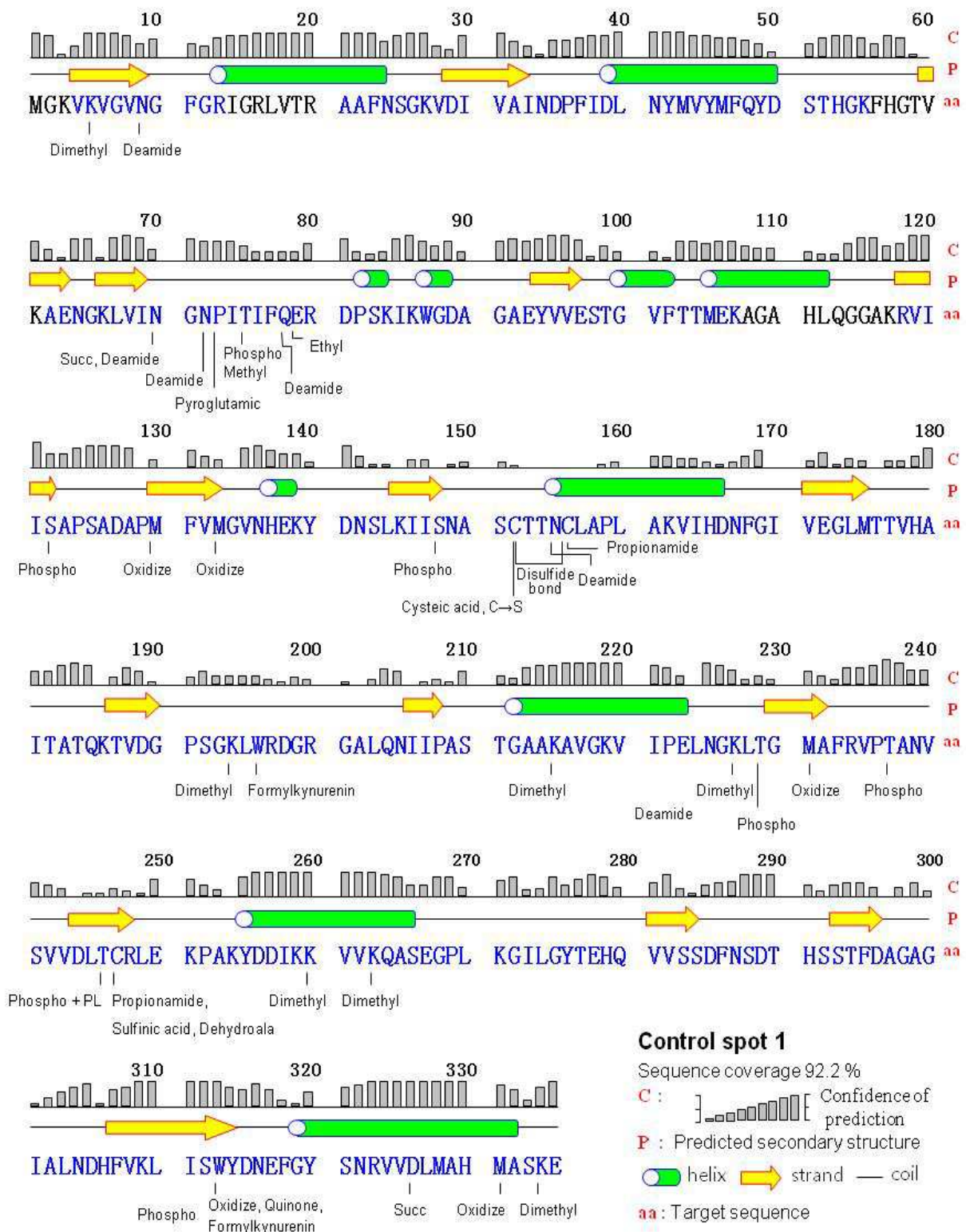of Post-Translational Modifications of Cellular Proteins Including Low...

99

Fig. 5. One example of comprehensive identifications of PTMs in cellular GAPDH (cited from Seo *et al.*, 2008).
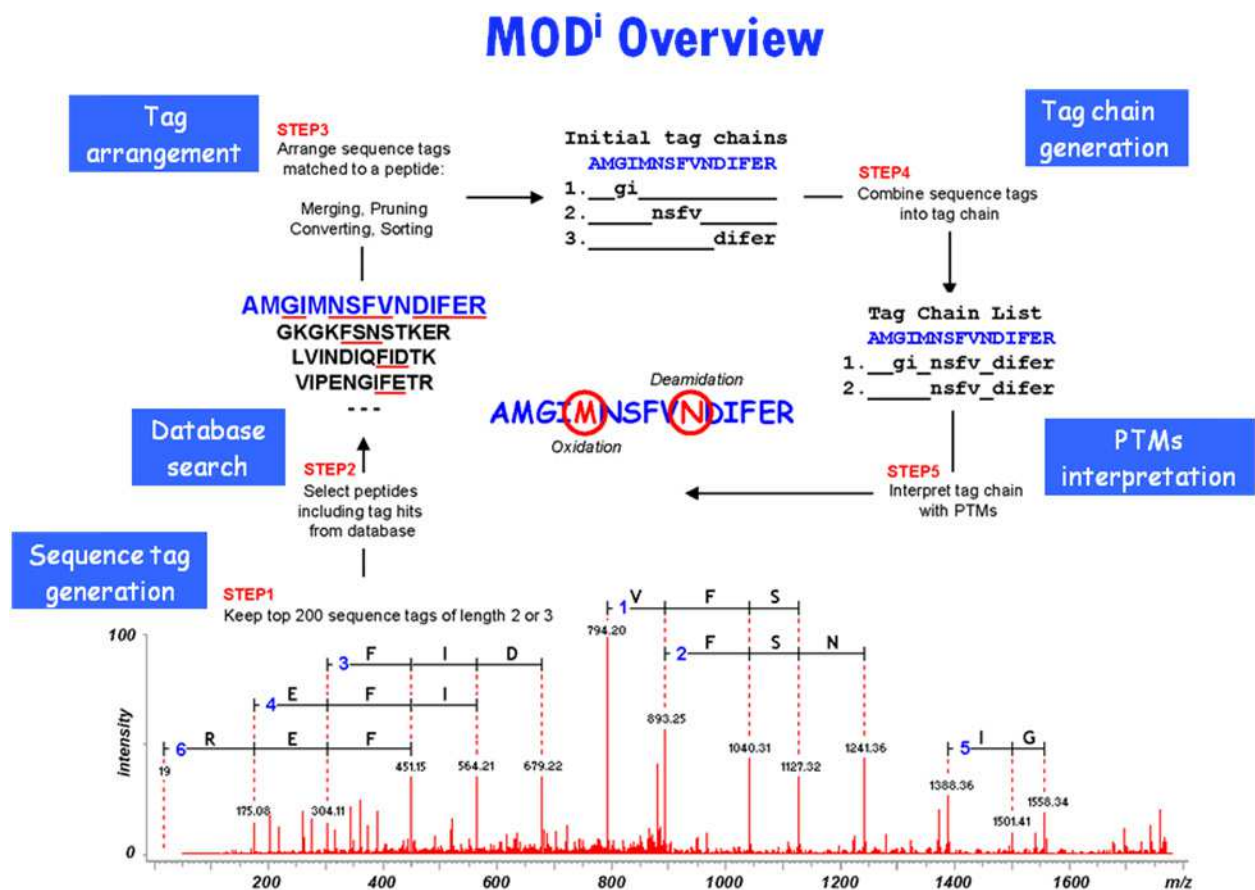
Fig. 6. Overview of MODⁱ algorithm (cited from Na *et al.*, 2008).

reduction and mass spectral comparison or by detecting the expected masses of disulfide linked peptides on mass scan level. However, these approaches for determination of disulfide bonds become ambiguous when the protein is highly bridged and modified. For accurate identification of disulfide linked peptides, we developed an algorithmic solution for the analysis of MS/MS spectra of disulfide bonded peptides under non-reducing condition. To determine disulfide linked sites, DBond takes into account fragmentation patterns of disulfide linked peptides in nucleoside diphosphate kinase (NDPK) as a model protein, considering fragment ions including cysteine, cysteine thioaldehyde (-2 Da), cysteine persulfide (+32 Da) and dehydroalanine (-34 Da). Using this algorithm, we successfully identified about a dozen novel disulfide bonds in a hexa EF-hand calcium binding protein secretagogin and in methionine sulfoxide reductase. We believe that DBond, which takes into account disulfide bond fragmentation characteristics and post-translational modifications, offers a novel approach for automatic identification of unknown disulfide bonds as well as their sites in proteins from MS/MS spectra (Choi *et al.*, 2010).

2.2.4 Examples of multiply modified proteins

Employing SEMSA, a sensitive mass spectrometric method for detecting low abundant protein modifications, and MODⁱ and DBond algorithm for searching for unknown modifications of separated protein on 2D-PAGE, we characterized the nature as well as the relative abundances of these hitherto unknown Cys modifications in cellular GAPDH purified on 2D-PAGE (Jeong *et al.*, 2011). We found unexpected mass shifts at active site Cys

residue (ΔM = -16, -34 and +64 Da) in addition to those of previously known oxidation products including sulfinic and sulfonic acids, and disulfide bonds. Similar changes were also found in other ROS-sensitive proteins including NDPK A, PRX6 and mitochondrial proteins. Mass differences of -16, -34 and +64 Da are presumed to reflect the conversion of Cys to Ser, DHA and Cys–SO_2-SH respectively. The plausible pathways leading to their formation from Cys were deduced from the distribution of the disulfide bonds and were confirmed in model systems by analyzing three dimensional protein structures and by employing model chemical reactions. Also sulfenic and sulfinic acids were detected as acrylamide adducts (ΔM = +87 and +103 Da) in samples on SDS-PAGE. These findings suggest that diverse modifications of redox-active Cys can be generated by ROS. These findings of unknown modifications are due to the sensitive mass spectrometric method, SEMSA, and unrestricted PTM search tool, MOD^i and DBond. These strategies should lead to identification of many unknown modifications which doubtless occur in various signaling pathways and in health and disease.

Combining 2D-PAGE for separating heterogenous population of one protein, SEMSA for analyzing low abundant modified peptide with MS/MS and MOD^i for searching algorithm to identify novel and multiple modifications, makes it possible to identify multiple and low abundant modifications of protein comprehensively.

3. Conclusion

Many protein modifications including amino acid substitutions, alternative splicing and post-translational modifications regulate the various biological functions of proteins by altering the protein-protein interactions, protein localization and protein activity. Understanding the relationship between protein modifications and their physiological activities will help define the principles on biological regulation. To demonstrate the exact protein function and its regulation, comprehensive identifications of low abundant and unexpected modification is essential. For the comprehensive identification of PTMs, peptide sequencing using tandem MS can give the most effective results. However, identification of PTMs does not simple as protein identification using MS. The enrichment of low abundant PTMs is necessary. Because most enrichment approaches of many type of PTMs are insufficient, we suggest the use of a combination of approaches: first separating the heterogeneous populations of a protein with multiple modifications on 2D-PAGE, then comprehensively identifying the PTMs employing SEMSA of MS/MS for detecting low abundant modifications and raising peptide coverage up to 100%, and searching MS/MS spectra using searching algorithm MOD^i for multiple and unexpected modifications. With these combined powerful proteomic tools, it will be possible to discover hitherto unknown PTMs in proteins, that influence various signaling pathways. This will hold the key for fully understanding the roles of proteins in biological and pathological processes in molecular level.

4. Acknowledgment

5. References

Aldini G, Carini M, Vistoli G, Shibata T, Kusano Y, Gamberoni L, Dalle-Donne I, Milzani A, Uchida K. (2007) Identification of actin as a 15-deoxy-Delta12,14-prostaglandin J2 target in neuroblastoma cells: mass spectrometric, computational, and functional approaches to investigate the effect on cytoskeletal derangement. *Biochemistry* 46(10), 2707-2718.

Blomster HA, Imanishi SY, Siimes J, Kastu J, Morrice NA, Eriksson JE, Sistonen L. (2010) In vivo identification of sumoylation sites by a signature tag and cysteine-targeted affinity purification. *J Biol Chem* 285(25), 19324-19329.

Bruderer R, Tatham MH, Plechanovova A, Matic I, Garg AK, Hay RT. (2011) Purification and identification of endogenous polySUMO conjugates. *EMBO Rep* 12(2), 142-148.

Choi S, Jeong J, Na S, Lee HS, Kim HY, Lee KJ, Paek E. (2010) A new algorithm for the identification of intact disulfide linkages based on fragmentation characteristics in tandem mass spectra. *J Proteome Res* 9(1), 626-635.

Choudhary C, Kumar C, Gnad F, Nielsen ML, Rehman M, Walther TC, Olsen JV, Mann M. (2009) Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science* 325(5942), 834-840.

Danielsen JM, Sylvestersen KB, Bekker-Jensen S, Szklarczyk D, Poulsen JW, Horn H, Jensen LJ, Mailand N, Nielsen ML. (2011) Mass spectrometric analysis of lysine ubiquitylation reveals promiscuity at site level. *Mol Cell Proteomics* 10(3), M110.003590.

Dephoure N, Zhou C, Villén J, Beausoleil SA, Bakalarski CE, Elledge SJ, Gygi SP. (2008) A quantitative atlas of mitotic phosphorylation. *Proc Natl Acad Sci* 105(31), 10762-10767.

Eng JK, McCormack AL, Yates JR. (1994) An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J Am Soc Mass Spectrom* 5, 976–989.

Galisson F, Mahrouche L, Courcelles M, Bonneil E, Meloche S, Chelbi-Alix MK, Thibault P. (2011) A novel proteomics approach to identify SUMOylated proteins and their modification sites in human cells. *Mol Cell Proteomics* 10(2), M110.004796.

Giron P, Dayon L, Sanchez JC. (2011) Cysteine tagging for MS-based proteomics. *Mass Spectrom Rev* 30(3), 366-395.

Gunaratne R, Braucht DW, Rinschen MM, Chou CL, Hoffert JD, Pisitkun T, Knepper MA. (2010) Quantitative phosphoproteomic analysis reveals cAMP/vasopressin-dependent signaling pathways in native renal thick ascending limb cells. *Proc Natl Acad Sci* 107(35), 15653-15658.

Guan KL, Yu W, Lin Y, Xiong Y, Zhao S. (2010) Generation of acetyllysine antibodies and affinity enrichment of acetylated peptides. *Nat Protoc* 5(9), 1583-1595.

Hwang NR, Yim SH, Kim YM, Jeong J, Song EJ, Lee Y, Lee JH, Choi S, Lee KJ. (2009) Oxidative Modifications of Glyceraldehyde-3-Phosphate Dehydrogenase Play a Key Role in Its Multiple Cellular Functions. *Biochem J* 423(2), 253-264.

Jamie D. Dunn, Gavin E. Reid, Merlin L. Bruening (2010) Techniques for phosphopeptide enrichment prior to analysis by mass spectrometry. *Mass Spectrometry Reviews* 29 (1), 29–54.

Jeong J, Jung Y, Na S, Jeong J, Lee E, Kim MS, Choi S, Shin DH, Paek E, Lee HY, Lee KJ. (2011) Novel oxidative modifications in redox-active cysteine residues. *Mol Cell Proteomics* 10(3), M110.000513.

Kim HJ, Song EJ, Lee KJ. (2002) Proteomic analysis of protein phosphorylations in heat shock response and thermotolerance. *J Biol Chem* 277(26), 23193-23207.

Kim S, Na S, Sim JW, Park H, Jeong J, Kim H, Seo Y, Seo J, Lee KJ, Paek E. (2006) MOD[i] : a powerful and convenient web server for identifying multiple post-translational peptide modifications from tandem mass spectra. *Nucleic Acids Research* 34, W258–W263.

Kim SC, Sprung R, Chen Y, Xu Y, Ball H, Pei J, Cheng T, Kho Y, Xiao H, Xiao L, Grishin NV, White M, Yang XJ, Zhao Y (2006). Substrate and functional diversity of lysine acetylation revealed by a proteomics survey. *Mol Cell* 23(4), 607–618.

Kim YM, Song EJ, Seo J, Kim HJ, Lee KJ. (2007) Proteomic analysis of tyrosine phosphorylations in vascular endothelial growth factor- and reactive oxygen species-mediated signaling pathway. *J Proteome Res* 6(2), 593-601.

Kim YM, Seo J, Kim YH, Jeong J, Joo HJ, Lee DH, Koh GY, Lee KJ. (2007) Systemic analysis of tyrosine phosphorylated proteins in angiopoietin-1 induced signaling pathway of endothelial cells. *J Proteome Res* 6(8), 3278-3290.

Lee E, Jeong J, Kim SE, Song EJ, Kang SW, Lee KJ. (2009) Multiple Functions of Nm23-H1 are Regulated by Oxido-Reduction System. *PLoS ONE* 4(11), e7949.

Leonard SE, Reddie KG, Carroll KS. (2009) Mining the thiol proteome for sulfenic acid modifications reveals new targets for oxidation in cells. *ACS Chem Biol* 4(9), 783-799.

Mann M, Wilm M. (1994) Error-tolerant identification of peptides in sequence databases by peptide sequence tags. *Anal Chem* 66, 4390–4399.

Manzano C, Abraham Z, López-Torrejón G, Del Pozo JC. (2008) Identification of ubiquitinated proteins in Arabidopsis. *Plant Mol Biol* 68(1-2), 145-158.

Mendez D, Hernaez ML, Diez A, Puyet A, Bautista JM. (2010) Combined proteomic approaches for the identification of specific amino acid residues modified by 4-hydroxy-2-nonenal under physiological conditions. *J Proteome Res* 9(11), 5770-5781.

Na S, Jeong J, Park H, Lee KJ, Paek E. (2008) Unrestrictive identification of multiple post-translational modifications from tandem mass spectrometry using an error-tolerant algorithm based on an extended sequence tag approach. *Mol Cell Proteomics* 7(12), 2452-2463.

Olsen JV, Vermeulen M, Santamaria A, Kumar C, Miller ML, Jensen LJ, Gnad F, Cox J, Jensen TS, Nigg EA, Brunak S, Mann M. (2010) Quantitative phosphoproteomics reveals widespread full phosphorylation site occupancy during mitosis. *Sci Signal* 3(104), ra3.

Peng J, Schwartz D, Elias JE, Thoreen CC, Cheng D, Marsischky G, Roelofs J, Finley D, Gygi SP. (2003) A proteomics approach to understanding protein ubiquitination. *Nat Biotechnol* 21(8), 921-926.

Perkin DN, Pappin DJC, Creasy DM, Cottrell JS. (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 20(18), 3551–3567.

Reed TT, Pierce WM, Markesbery WR, Butterfield DA. (2009) Proteomic identification of HNE-bound proteins in early Alzheimer disease: Insights into the role of lipid peroxidation in the progression of AD. *Brain Res* 1274, 66-76.

Rikova K, Guo A, Zeng Q, Possemato A, Yu J, Haack H, Nardone J, Lee K, Reeves C, Li Y, Hu Y, Tan Z, Stokes M, Sullivan L, Mitchell J, Wetzel R, Macneill J, Ren JM, Yuan J, Bakalarski CE, Villen J, Kornhauser JM, Smith B, Li D, Zhou X, Gygi SP, Gu TL,

Polakiewicz RD, Rush J, Comb MJ. (2007) Global survey of phosphotyrosine signaling identifies oncogenic kinases in lung cancer. *Cell* 131(6), 1190-1203.

Roe MR, Xie H, Bandhakavi S, Griffin TJ. (2007) Proteomic mapping of 4-hydroxynonenal protein modification sites by solid-phase hydrazide chemistry and mass spectrometry. *Anal Chem* 79(10), 3747-3756.

Rush J, Moritz A, Lee KA, Guo A, Goss VL, Spek EJ, Zhang H, Zha XM, Polakiewicz RD, Comb MJ. (2005) Immunoaffinity profiling of tyrosine phosphorylation in cancer cells. *Nat Biotechnol* 23(1), 94-101.

Savitski MM, Nielsen ML, Zubarev RA. (2006) ModifiComb, a new proteomic tool for mapping substoichiometric post-translational modifications, finding novel types of modifications, and fingerprinting complex protein mixtures. *Mol Cell Proteomics* 5(12), 935–948.

Seo J, Jeong J, Kim YM, Hwang N, Paek E, Lee KJ. (2008) A strategy for Comprehensive Identification of Post-Translational Modifications in Cellular Proteins, Including Low Abundance Modifications: Application to Glyceraldehyde-3-phosphate Dehydrogenase. *J Proteome Res* 7(2), 587-602.

Seo YH, Carroll KS. (2009) Profiling protein thiol oxidation in tumor cells using sulfenic acid-specific antibodies. *Proc Natl Acad Sci* 106(38), 16163-16168.

Tabb DL, Saraf A, Yates JR. (2003) GutenTag: high-throughput sequence tagging via an empirically derived fragmentation model. *Anal Chem* 75(23), 6415–6421.

Tanner S, Shu H, Frank A, Wang LC, Zandi E, Mumby M, Pevzner PA, Bafna V. (2005) InsPecT: identification of posttranslationally modified peptides from tandem mass spectra. *Anal Chem* 77(14), 4626–4639.

Thingholm TE, Jensen ON, Larsen MR. (2009) Analytical strategies for phosphoproteomics. *Proteomics* 9(6), 1451–1468.

Tsur D, Tanner S, Zandi E, Bafna V, Pevzner PA. (2005) Identification of post-translational modifications by blind search of mass spectra. *Nat Biotechnol* 23(12), 1562–1567.

Tyther R, Ahmeda A, Johns E, McDonagh B, Sheehan D. (2010) Proteomic profiling of perturbed protein sulfenation in renal medulla of the spontaneously hypertensive rat. *J Proteome Res* 9(5), 2678-2687.

Vosseller K, Trinidad JC, Chalkley RJ, Specht CG, Thalhammer A, Lynn AJ, Snedecor JO, Guan S, Medzihradszky KF, Maltby DA, Schoepfer R, Burlingame AL. (2006) O-linked N-acetylglucosamine proteomics of postsynaptic density preparations using lectin weak affinity chromatography and mass spectrometry. *Mol Cell Proteomics* 5(5), 923-934.

Wang Z, Udeshi ND, Slawson C, Compton PD, Sakabe K, Cheung WD, Shabanowitz J, Hunt DF, Hart GW. (2010) Extensive crosstalk between O-GlcNAcylation and phosphorylation regulates cytokinesis. *Sci Signal* 3(104), ra2.

Weinert BT, Wagner SA, Horn H, Henriksen P, Liu WR, Olsen JV, Jensen LJ, Choudhary C. (2011) Proteome-wide mapping of the Drosophila acetylome demonstrates a high degree of conservation of lysine acetylation. *Sci Signal* 4(183), ra48.
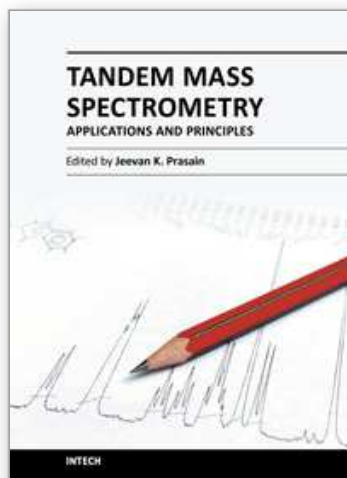
Tandem Mass Spectrometry - Applications and Principles

Edited by Dr Jeevan Prasain

Tandem Mass Spectrometry - Applications and Principles presents comprehensive coverage of theory, instrumentation and major applications of tandem mass spectrometry. The areas covered range from the analysis of drug metabolites, proteins and complex lipids to clinical diagnosis. This book serves multiple groups of audiences; professional (academic and industry), graduate students and general readers interested in the use of modern mass spectrometry in solving critical questions of chemical and biological sciences.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Jaeho Jeong, Jae-Jin Lee and Kong-Joo Lee (2012). Proteomic Strategies for Comprehensive Identification of Post-Translational Modifications of Cellular Proteins Including Low Abundant and Novel Modifications, Tandem Mass Spectrometry - Applications and Principles, Dr Jeevan Prasain (Ed.), ISBN: 978-953-51-0141-3, InTech, Available from: http://www.intechopen.com/books/tandem-mass-spectrometry-applications-and-principles/proteomic-strategies-for-comprehensive-identification-of-post-translational-modifications-of-cellula

INTECH
open science | open minds