Molecular Life

Proteomics of the post-translational modifications: the knowns and the unknowns

Roshanak Aslebagh[1], Kelly L. Wormwood[1], Devika Channaveerappa[1], Emmalyn J. Dupree[1], Izabela Sokolowska[1], Armand G. Ngounou Wetie[1], Sapan Patel[1], Marius Mihasan[1, 2], Alisa G. Woods[1], Costel C. Darie[1, *]

[1]*Biochemistry & Proteomics Group, Department of Chemistry & Biomolecular Science, Clarkson University, 8 Clarkson Avenue, Potsdam, NY, 13699-5810, USA*
[2]*Department of Biology, Alexandru Ioan Cuza University, Bd. Carol I, No. 11, 700506 Iasi, Romania*

(Received 16 March, 2017; accepted 17 April, 2017)

Abstract

Great progress has been made in identification and characterization of proteins by mass spectrometry-based proteomics, but there is still a great challenge to identify and characterize post-translational modifications (PTMs) within proteins. This challenge is even greater when one discovers that not all protein and peptide modifications are important and the experimental-borne modifications not only complicate the data, but prevents us from focusing on and identifying the true natural and physiological relevant PTMs. Here, we demonstrate the power of mass spectrometry (MS) in identification and characterization of PTMs. We also describe the most important natural and experimental-borne PTMs, along with several strategies to identify PTMs, including specific examples. Finally, we discuss the strengths and weaknesses within this field.

Keywords: post-translational modifications, mass spectrometry-based proteomics

Abbreviations: PTMs=post-translational modifications; PPIs=protein-protein interactions; SDS-PAGE=sodium dodecyl sulfate-polyacrylamide gel electrophoresis; 2D-PAGE=two-dimensional polyacrylamide gel electrophoresis; MS=mass spectrometry; ESI=electrospray ionization; MALDI=Matrix Assisted Laser Desorption Ionization; ESI-MS=electrospray ionization mass spectrometry; MALDI-MS=MALDI mass spectrometry; LC-MS/MS=liquid chromatography tandem mass spectrometry; ESI-MS/MS=electrospray ionization tandem mass spectrometry; m/z=mass/charge; PNGaseF=peptide-N-glycosidase F; HNE=4-hydroxy-2-nonenal; IMAC=immobilized metal affinity chromatography; TiMAC=combination of TiO2 and IMAC; MRM=multiple reaction monitoring; DDA=data dependent acquisition; PID=product ion discovery; BSA=bovine serum albumin; DTT=dithiothreitrol; IAA=iodoacetamide.

INTRODUCTION TO PROTEOMICS

Proteomics is the study of the protein complement or proteome, done through mass spectrometry (MS) (1-8). Proteomics focuses on analysis of proteins and protein derivatives (*i.e.*, glycoproteins), peptides, protein-protein interactions (PPIs) or post-translational modifications (PTMs) within proteins. A proteomics experiment starts with sample fractionation prior to analysis by MS (9), which involves one or more biochemical separation methods such as electrophoresis (SDS-PAGE, 2D-PAGE), centrifugation or chromatography (affinity purification). In some cases, fractionation can also be bypassed, provided the instrumentation used can handle complex protein samples (*i.e.*, cell lysates). However, in most cases, fractionation is followed by further separation methods such as additional chromatography (*i.e.*, size exclusion chromatography) or centrifugation (*i.e.*, centrifugation-based affinity chromatography columns). If the protein sample is not very complex, fractionation is directly followed by enzymatic digestion (*i.e.*, trypsin) and peptide extraction. The peptide mixture is then further fractionated by HPLC or UPLC and then analyzed by MS (1). Data analysis usually leads to identification of one or more proteins. This is the end goal in most proteomics experiments, but in some cases this is just a start. Additional investigation of the MS data can extract information that leads to identification of potential interaction partners of some proteins (protein-protein

*Corresponding author (Tel: (315) 268-7763; Fax: (315) 268-6610; E-mail: cdarie@clarkson.edu)

interactions), or of modifications of that protein (natural or unnatural PTMs (8, 10-12). A schematic workflow of a proteomics experiment is shown in Fig. 1.

PROTEIN IDENTIFICATION AND CHARACTERIZATION: IDENTIFICATION OF POST-TRANSLATIONAL MODIFICATIONS (PTMS)

MS-based identification of one peptide that contains the amino acid sequence specific to a protein is usually enough for identification of that protein. Once the protein is identified, then it is characterized. However, identification of two peptides that are part of the same protein is better and, as a general rule, the more peptides are identified per protein, the higher the probability (and confidence) that the protein identified is the correct one. Assuming that a 50 kDa protein can in theory produce at least 50 peptides, identifying a protein based on only one or two peptides seems easy. This is true, up to a point: good MS instrumentation, low flow rate of the HPLC/UPLC, good software for data analysis, good technical expertise for protein digestion and peptide extraction, etc. However, when it comes to protein characterization, a great challenge is encountered by most biochemists, protein chemists and mass spectrometrists. In addition, to succeed in protein identification and characterization, a combination of bottom-up and top-down proteomics should be used (13, 14).

TOP-DOWN MS AND PTM ANALYSIS

A top down approach allows for the identification of protein isoforms and/or any potential post-translational modifications within these proteins (15). Once the protein's identity is known, the protein characterization using MS is usually completed through an interconnected, two-step approach: 1) a top-down MS approach where intact proteins are investigated and their molecular mass is determined and 2) a bottom-up approach where proteins are digested and the peptides mixture is analyzed, as described in Fig. 1. While the bottom up approach is experimentally similar to top-down approach for both 1) protein identification based on minimum one peptide and 2) full characterization of a protein, the differences between these two approaches are rather great and will be listed below.

Top down MS can be used to investigate the molecular mass of the protein of interest, usually the first step in its characterization. While the mass of a protein can be determined with various native (*i.e.*, size exclusion chromatography) or non-native, denaturing (electrophoresis/SDS-PAGE) biochemical methods, the error produced by these methods is rather large (up

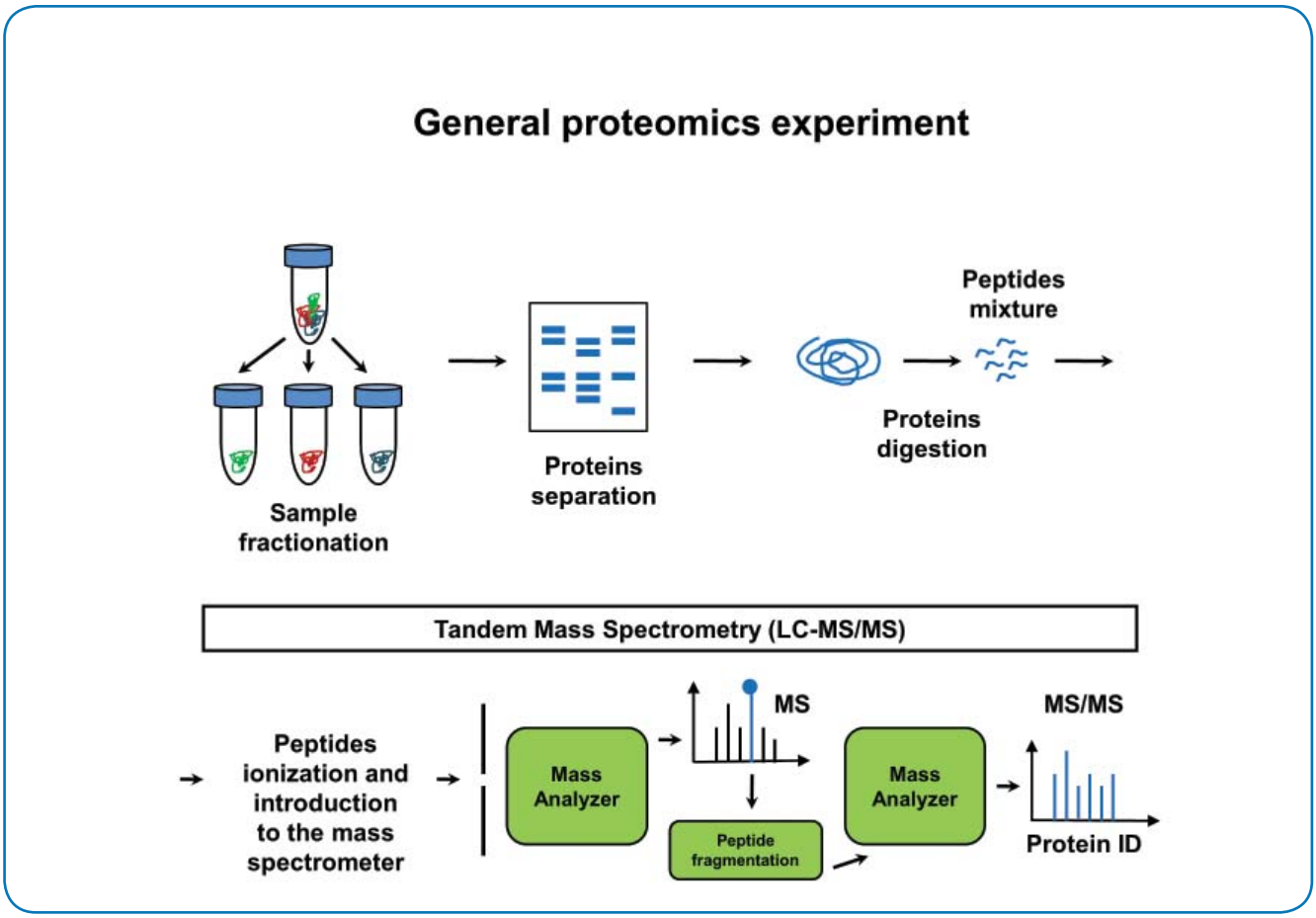

Fig. 1. – General proteomics experiment workflow schematic. Reprinted and adapted with permission from (9, 47)

to 10% of the mass of the protein). Therefore, these methods for determination of a protein's molecular mass are merely a gross estimation, far from their exact or nearly-exact mass. As such, the only accurate method used for determination of a protein's mass and then amino acid sequence information is by top down MS (using a MS with either a MALDI or ESI source followed by MS/MS).

Determination of the mass of a protein is important in cases like investigating the modifications and degradations to which hormones are subjected when they are released under natural and close-to-natural "sense and release" conditions. In one of our recent publications, we investigated whether insulin released electrochemically is 1) indeed released and 2) not modified or degraded (in a system that can sense high concentrations of glucose, then releases insulin to reduce the glucose levels) (16). While the insulin release could easily be detected by other methods (*i.e.*, insulin labeled with a fluorophore), any potential modifications of insulin (*i.e.*, oxidations) and degradations could only be demonstrated by MS. An example of such an experiment is shown in Fig. 2.

Another example where determination of the molecular mass of a protein by top-down proteomics is important is in identification and characterization of antibodies that

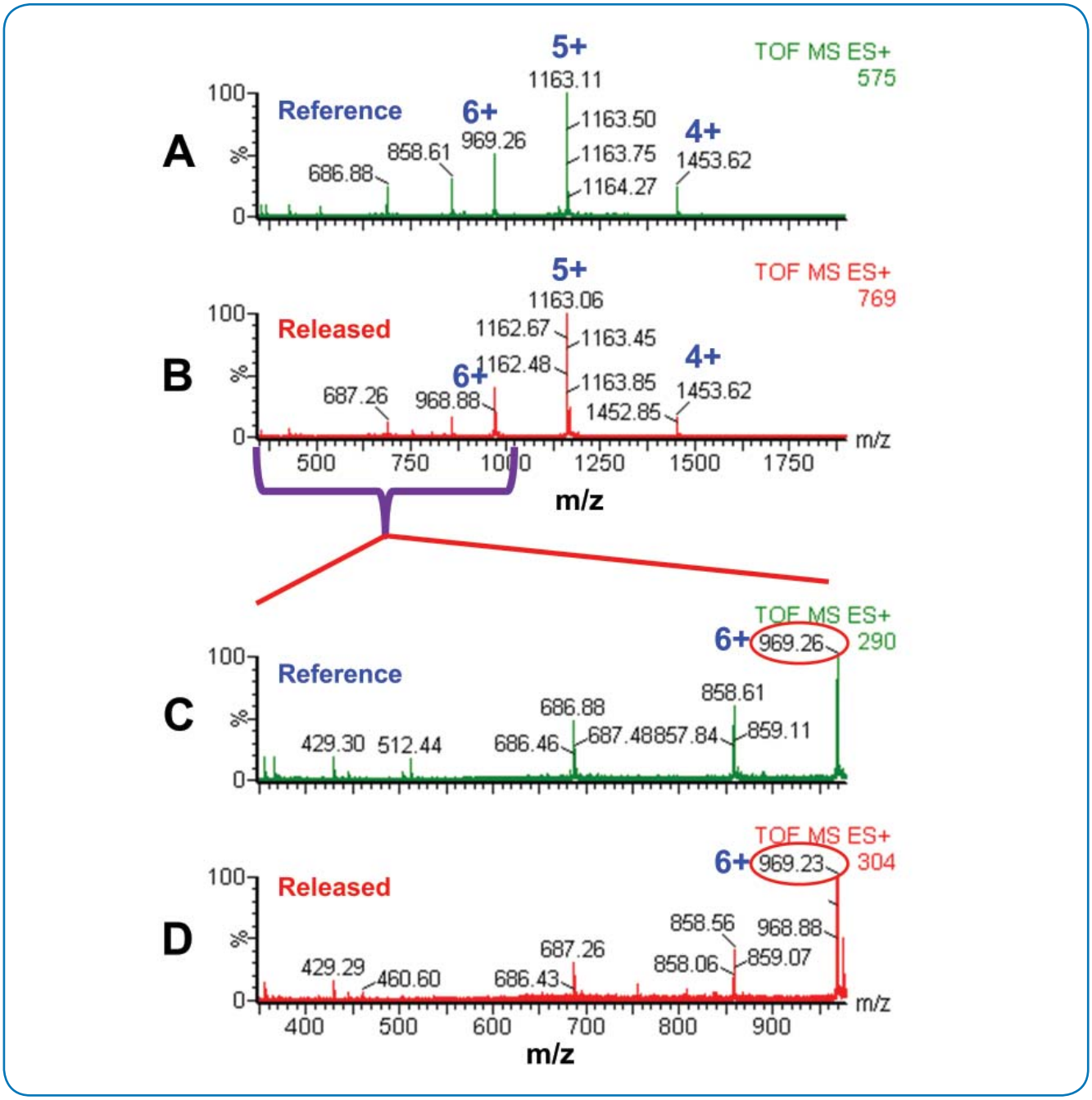

Fig. 2. – Analysis of insulin by MS. Full mass spectra of the reference (A) and released (B) insulin, which shows that the insulin does not degrade and no major peaks are produced. The 4+, 5+ and 6+ peaks of insulin are indicated. The low m/z range of these spectra (m/z of 350-1,000) from (A) and (B) are enhanced in (C) and (D). The (6+) peak of insulin is also shown. Again, no major degradation products of insulin were observed. Reprinted and adapted with permission from (16).

are used as either therapeutics/immunotherapeutics or as drug carriers (antibody-drug conjugates). Analysis of the antibody in the non-reducing conditions (disulfide-linked), then reducing conditions (individual light and heavy chains) and then de-glycosylated heavy chains (treated with peptide-N-glycosidase F (PNGaseF), which removes N-linked oligosaccharides from the NXS/T glycosylation sites) allows simultaneous identification of the differences between the glycosylated and non-glycosylated heavy chain as well as comparison with the theoretical molecular mass (30). This also leads to identification of potential mutations or modifications of both heavy and light chains.

Top down proteomics also allows for identification of the potential natural and artificial modifications. Furthermore, simple MS analysis of the intact proteins allows for identification of the type and number of these modifications. One such example is 4-hydroxy-2-nonenal (HNE) modification, shown in Fig. 3. As observed, MS analysis allows not only the identification of HNE molecules, but also the number of HNE molecules that are covalently bound to lysozyme,

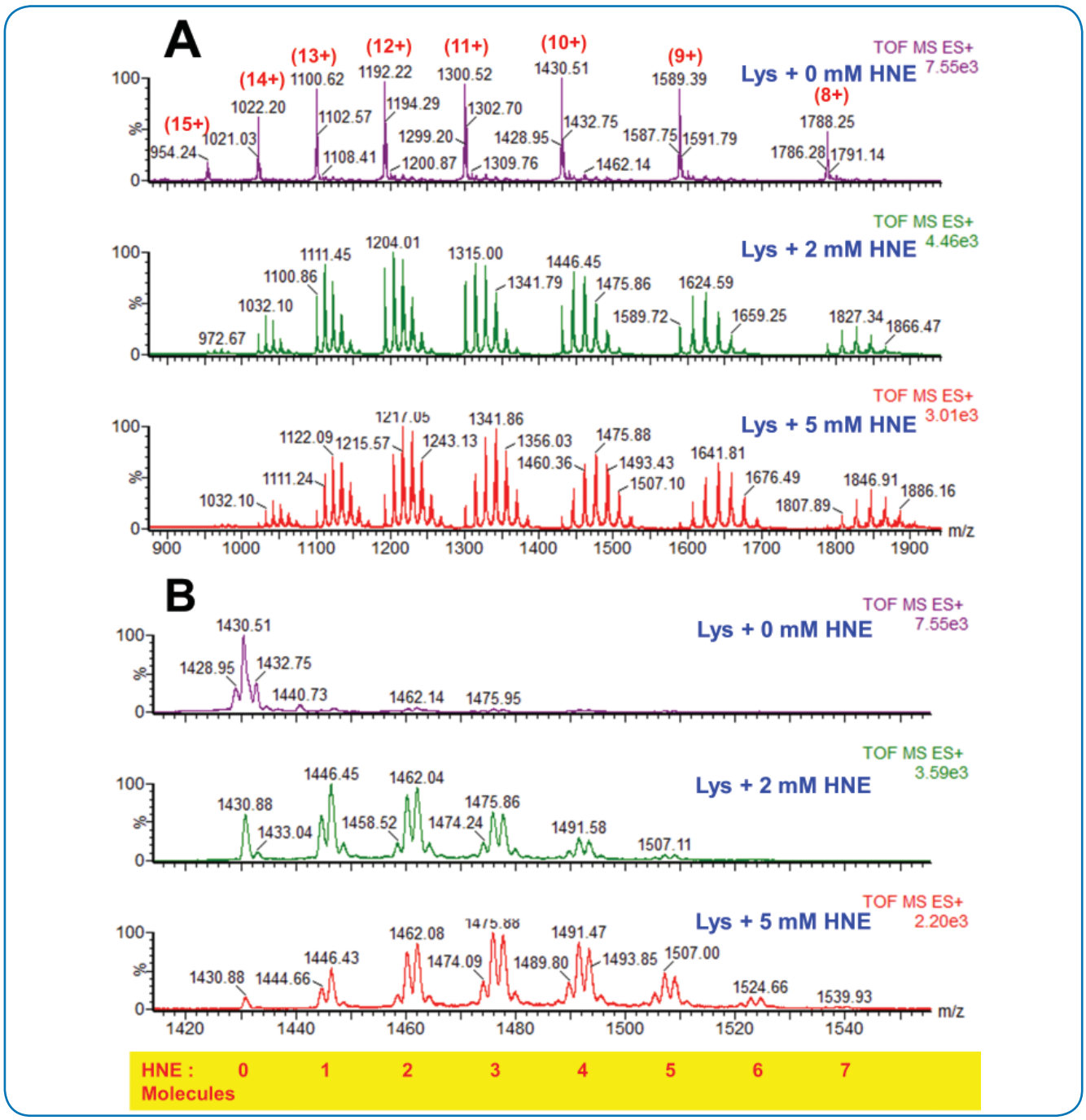

Fig. 3. – Direct infusion analysis by ESI-MS of untreated and HNE-treated lysozyme. (A) Different charge states of untreated and HNE-treated lysozyme. (B) Magnification of (10+) peak to determine the number of HNE molecules *per* protein and the type of HNE-induced modification. Reprinted and adapted with permission from (17).

a model protein, as well as the type of modification (Michael-based addition or Schiff-base formation) (17).

BOTTOM UP PROTEOMICS AND PTM ANALYSIS

In a bottom-up approach, after the proteins are fractionated by a biochemical approach, they are digested and subjected to MS analysis by tandem mass spectrometry (MS/MS or ESI-MS/MS) or by peptide mass fingerprinting for protein identification (in MALDI-MS analysis). Since bottom up approaches allow for identification of parts of a protein (peptides), their fragmentation by MS/MS also allows for identification of the amino acid sequence of that peptide. This leads to identification of the protein that has the identified amino acid sequence. Furthermore, MS/MS analysis of peptides also allows us to identify all potential PTMs on that peptide (and to pinpoint towards the amino acid that is modified), regardless of whether they are natural PTMs or artificial, experiment-borne PTMs.

While identification of a protein based on the amino acid sequence of one peptide is relatively straightforward, the challenge comes when the full coverage of that particular protein is required. For example, if a trypsin digestion of a protein will lead to identification of ~20-30% coverage of that particular protein, increase of that coverage can easily be extended to 60-70% by either increasing the amount of protein analyzed, or lowering the flow rate of the HPLC/UPLC or increasing the length of the UPLC gradient. However, rarely the protein coverage is higher than 70%, simply because some parts of the protein are hydrophobic or have not enough or too many tryptic cleavage sites (and as consequence, the peptide fragments are either too big or too small for MS analysis). Therefore, a complementary digestion approach by using a different enzyme, such as chymotrypsin or AspN will be needed. In addition, success in protein identification and characterization is fully dependent on a combination of bottom-up and top-down proteomics (13, 14). Once the protein is identified and has full/maximum peptide coverage, the PTM analysis can proceed, which is different, depending on the type of PTM (glycosylation, phosphorylation or acetylation), its status (stable or transient) and its origin (natural or experimental PTMs).

COMMON, KNOWN, NATURAL PTMS IN PROTEINS

MS-based characterization of proteins is difficult, even more complicated due to the PTMs of proteins (18). This is also dependent on what type of PTM is investigated. The stable modifications such as glycosylation or disulfide bridges are relatively easy to detect, but difficult to interpret. Other PTMs such as phosphorylation, acetylation or methylation are transient, and also difficult to identify. It is even more difficult to identify PTMs and give them biological significance, *i.e.* to pinpoint the role of that PTM in a cell or a tissue.

Two major PTMs in proteins are glycosylation and phosphorylation. Glycosylation is usually found in extracellular proteins and is responsible for protein-glycoprotein or ligand-lectin interaction (19, 20). In the pharmaceutical and biotechnology industries, glycosylation is an important modification/PTM of recombinant proteins, which has dramatic effects on the stability and solubility of those proteins (21, 22).

Glycoprotein analysis is usually accomplished by LC-MS/MS of the glycopeptides produced by enzymatic (*i.e.*, trypsin) digestion. Glycopeptide analysis allows for identification of hexose diagnostic fragments (23-25). However, this method is not very effective, because the glycoproteins are not detected very efficiently (poor signal in MS analysis). Therefore, affinity chromatography-based glycoprotein enrichment (*i.e.*, by lectins), chip-based glycoprotein analysis (26, 27) or customized fragmentation methods such as collision induced dissociation (CID) for producing ions mainly of the glycan chain or electron transfer dissociation (ETD) for the specific fragmentation of the peptide backbone greatly improves the outcome of the MS analysis of glycoproteins (28, 29).

A complementary analysis of glycoproteins involves enzymatic digestion of the glycans upon their enzymatic cleavage from the proteins. As such, the glycans can be analyzed in negative mode and the location of the glycan on peptide can be identified by analysis of the peptide that releases the glycan. Negative or negative ionization refers to polarity of the current; usually peptides are analyzed in positive ionization mode while glycans in negative ionization mode. In the simplest example, N-linked glycans are released by enzymatic treatment with peptide-N-glycosidase F (PNGaseF), while O-linked glycans can be released by β-elimination. Advantages and disadvantages of using the glycan digestion include among others increase of ionization efficiency and hydrophobicity of the digested peptide and separate characterization of the glycan and peptide sequence (pros) and potential false positive because of the spontaneous deamidation of asparagine residues and the loss of information regarding site-specific characterization of glycan heterogeneity (cons).

PNGaseF treatment leads to identification of the N-linked glycans, in which the asparagine that contains the glycan is converted to aspartate, which leads to a 1 mass unit increase within the peptide (30). This change is more than enough for MS-based identification. The PNGaseF treatment can also be performed in presence of $H_2{}^{18}O$, which allows an easier identification of the glycosylation site (31). In the case of O-linked glycans,

which are linked through an O-glycosidic bond to a serine or threonine residue, their enzymatic removal converts serine to alanine and threonine to aminobutyric acid; in both cases, the outcome is a loss of 16 mass units/Da within the peptide bearing the O-linked glycan (32, 33).

Glycan analysis can also be used to investigate the N-glycosylation site occupancy. Figure 4A shows an example of a MS/MS analysis where the increase of a peptide molecular mass is due to PNGaseF-based release of glycan and conversion of asparagine to aspartate, suggesting that the peptide is glycosylated. Figure 4B shows an example of a MS/MS analysis where the same asparagine from the same NST glycosylation site, from the same peptide as the one shown in Fig. 4A, is not glycosylated and therefore not converted to aspartate. This data suggests that the NST glycosylation site can be investigated not only for the transition of asparagine to aspartate in NST site due to glycosylation but also of the asparagine site in NST site due to non-glycosylation.

Phosphorylation is a different PTM that is reversible and transient and is important in regulation of many cellular processes (34). It is well-demonstrated that abnormal protein phosphorylation leads to the onset of various diseases (35, 36) and identification of protein phosphorylation allows understanding of many physiological processes such as the signal transduction pathways, which in turn can help in developing new therapeutic targets (37-39).

MS-based identification and characterization of phosphorylation on peptides is usually achieved by

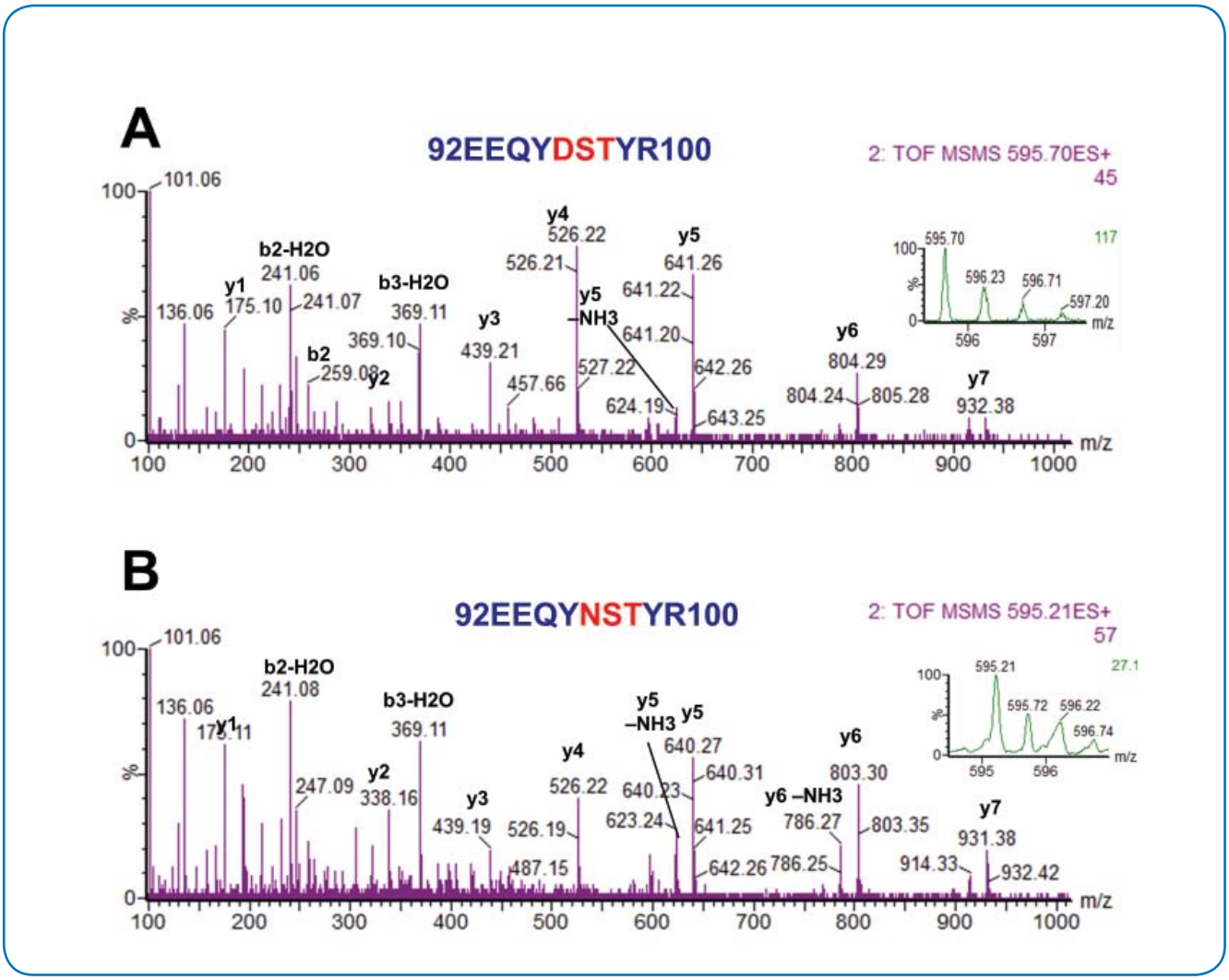

Fig. 4. – Analysis of N-linked glycosylation sites. (A) LC-MS/MS analysis of the PNGaseF-treated (deglycosylated) IgG heavy chain (IgG-HC) for identification of the potential glycosylation sites for N-linked oligosaccharides. PNGaseF-treated IgG was digested by trypsin and then by AspN (trypsin-AspN double digestion) and the resulting peptide mixture was analyzed by LC-MS/MS. A doubly charged peak with m/z of 595.70 (2+) was fragmented by MS/MS and produced a series of peaks (product b and y ions) that led to identification of the peptide with the sequence 92EEQYDSTYR100. The MS precursor peak with m/z 595.70 (2+) is also shown (expanded). (B) In a different experiment, LC-MS/MS analysis of the same peptide as shown in (A) but not treated with PNGaseF indicates that this IgG is not glycosylated: the MS/MS analysis of the doubly charged peak with m/z of 595.21 (2+) corresponds to peptide with the sequence 92EEQYNSTYR100, with N from NST glycosylation site unmodified. The MS precursor peak with m/z 595.21 (2+) is also shown (expanded). Reprinted and adapted with permission from (27).

scanning for neutral loss of HPO_3 (80 Da for phosphotyrosine) and H_3PO_4 (98 Da for phosphoserine and phosphothreonine residues), such as neutral loss triggered MS3 or neutral loss electron transfer dissociation/electron capture dissociation (40, 41). It is worth noting that phosphotyrosine peptides are not very often observed as neutral loss in positive ionization mode (42), but can easily be observed in negative ionization mode if the peptides are methyl esterified (43). These experiments usually lead to identification of the phosphopeptides and of the amino acid that is phosphorylated (serine, threonine or tyrosine) (44-46). While this approach is straightforward, many times it is not good enough for identification of the full phosphoproteome, simply because the phosphopeptides may not be abundant enough and as such, enrichment of phosphopeptides using TiO_2, immobilized metal affinity chromatography (IMAC), or a combination of TiO_2 and IMAC (TiMAC) is needed (47, 48). Affinity-based enrichment using antibodies (anti-phospho-serine, anti-phospho-threonine or anti-phospho-tyrosine) can also be (and are) used.

The approaches described for analysis of PTMs like phosphorylation or glycosylation at the MS level are not limited to these two PTMs and can be adapted and then applied to other PTMs. For example, methods such as multiple reaction monitoring (MRM) for quantitative analysis of phosphopeptides or customized data-dependent analysis (DDA using an inclusion list, also known as information-dependent analysis or IDA) for N-glycan occupancy analysis, product ion discovery (PID) for glycan analysis (and detection of the hexose or N-acetyl-neuraminic acid product ions) or neutral loss for phosphorylation analysis are commonly used for identification of all kinds of PTMs (8, 12, 17, 30, 49, 50).

EXPERIMENT-BORNE PTMS IN PROTEINS: THE KNOWNS AND THE UNKNOWNS

During protein digestion, two main experiment-borne procedures modify the proteins: reduction and alkylation of cysteine (acrylamide- or iodoacetamide (IAA)-based alkylation of cysteine residues) and methionine oxidation. The modifications on these two amino acids can easily be controlled at the bioinformatics level: during database search one will use fixed modifications for acrylamide- or IAA-based alkylation of cysteine (propionamide- and carbamidomethyl-cysteine) and variable modification for methionine (methionine oxidation). While controlling modifications of these amino acids in proteins is relatively straightforward, controlling of most experiment-borne PTMs on amino acids other than cysteine or methionine is practically impossible. In the simplest example, IAA usually alkylates/should alkylate only cysteine. However, in a previous study, our lab demonstrated that many cysteine-less peptides are still modified by IAA (and many times by more than one molecule of iodoacetamide) (50). One of the worst consequences for this unexpected modification is not only the inability to identify a modified peptide, but also the influence on the quantitative analysis when such peptide is used. An example of such a modification is shown in Fig. 5, where a cysteine-less peptide has no modifications, but it is also modified by one and two molecules of IAA (50). Interesting to note, when two molecules of IAA (2 x 57 Da) modify a peptide, then IAA-modified peptide artifacts mimic protein ubiquitination, which is a di-glycine (2 x 57 Da) (51).

A different example where unwanted modifications of amino acids in proteins happen, is in oxidative stress, particularly lipid peroxidation and subsequent modification of proteins. Such a phenomenon happens in the retina. In one of our recent studies, we used two model proteins (BSA and lysozyme) to investigate the amino acids that are modified by HNE (17). The amino acids that were expected to be modified were histidine, lysine and cysteine. Our proteomics experiments on our HNE-modified model proteins confirmed that these amino acids are indeed modified. However, we also demonstrated that many additional amino acids, previously not known to be modified by HNE were unexpectedly modified (17). Therefore, our lab showed again that the most dangerous modifications are the ones that are unknown and unpredictable. An example of such a modification is shown in Fig. 6.

FALSE POSITIVES, FALSE NEGATIVES AND GOOD QUALITY SPECTRA WITH NO MATCH

Many MS/MS spectra from a proteomics experiment do not lead to identification of any peptides, simply because the quality of the MS/MS spectra is not acceptable. Therefore, when one does a search against a database, *i.e.*, Mascot search, false positives (peptide with a low score) can be identified. While this is not of a concern, simply because the human factor can play a role in making a decision, the opposite can happen, as well. For example, a MS/MS with a low Mascot score can lead to identification of a peptide based on a good series of y ions (and sometimes some b ions, as well). However, only inspection of the original raw MS/MS spectrum can confirm whether it is a good match to the peptide that was identified or not. In addition, the quality of the MS/MS spectra that are acquired in DDA mode depends equally on both the peptides and the type of the instrument. For example, the quadrupole instruments produce MS/MS spectra mostly with y ions, while trap-based instruments produce both b and y ions. Conversely, the same happens the opposite way: false negative identifications. While false positive identifications are relatively easy to identify, it is very difficult and many times almost impossible to

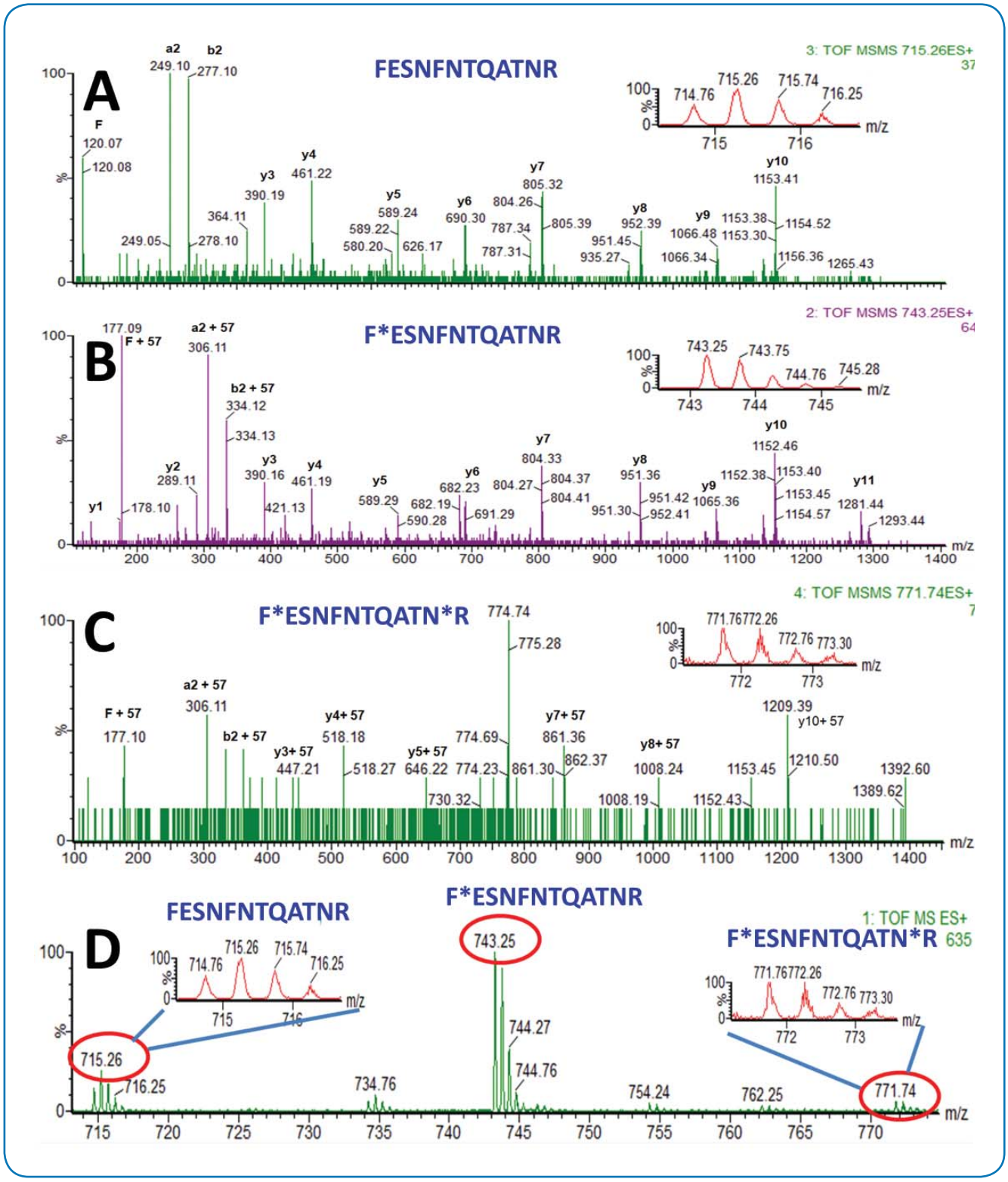

Fig. 5. – LC-MS/MS analysis of a SDS-PAGE gel band that was reduced by DTT, alkylated by IAA and digested by trypsin. (A) MS/MS of the doubly-charged precursor ion with m/z of 714.76 produced a series of b and y product ions that led to identification of the peptide with sequence FESNFNTQATNR that was part of the lysozyme protein. This peptide was identified during the Mascot database search.
(B) MS/MS of the doubly-charged precursor ion with m/z of 743.25 produced a series of b and y product ions that led to identification of the mono-alkylated peptide with sequence F*ESNFNTQATNR.
(C) MS/MS of the doubly-charged precursor ion with m/z of 771.76 produced a series of b and y product ions that led to identification of the di-alkylated peptide with sequence F*ESNFNTQATN*R. (D) MS spectrum showing the doubly-charged precursor ions with m/z of 714.76, m/z of 743.25 and m/z of 771.76 that corresponds to un-modified peptide N*TDGSTDYGILQINSR, mono-alkylated peptide N*TDGSTDYGILQINSR and di-alkylated peptide N*TDGSTDYGILQIN*SR. Reprinted and adapted with permission from (44).

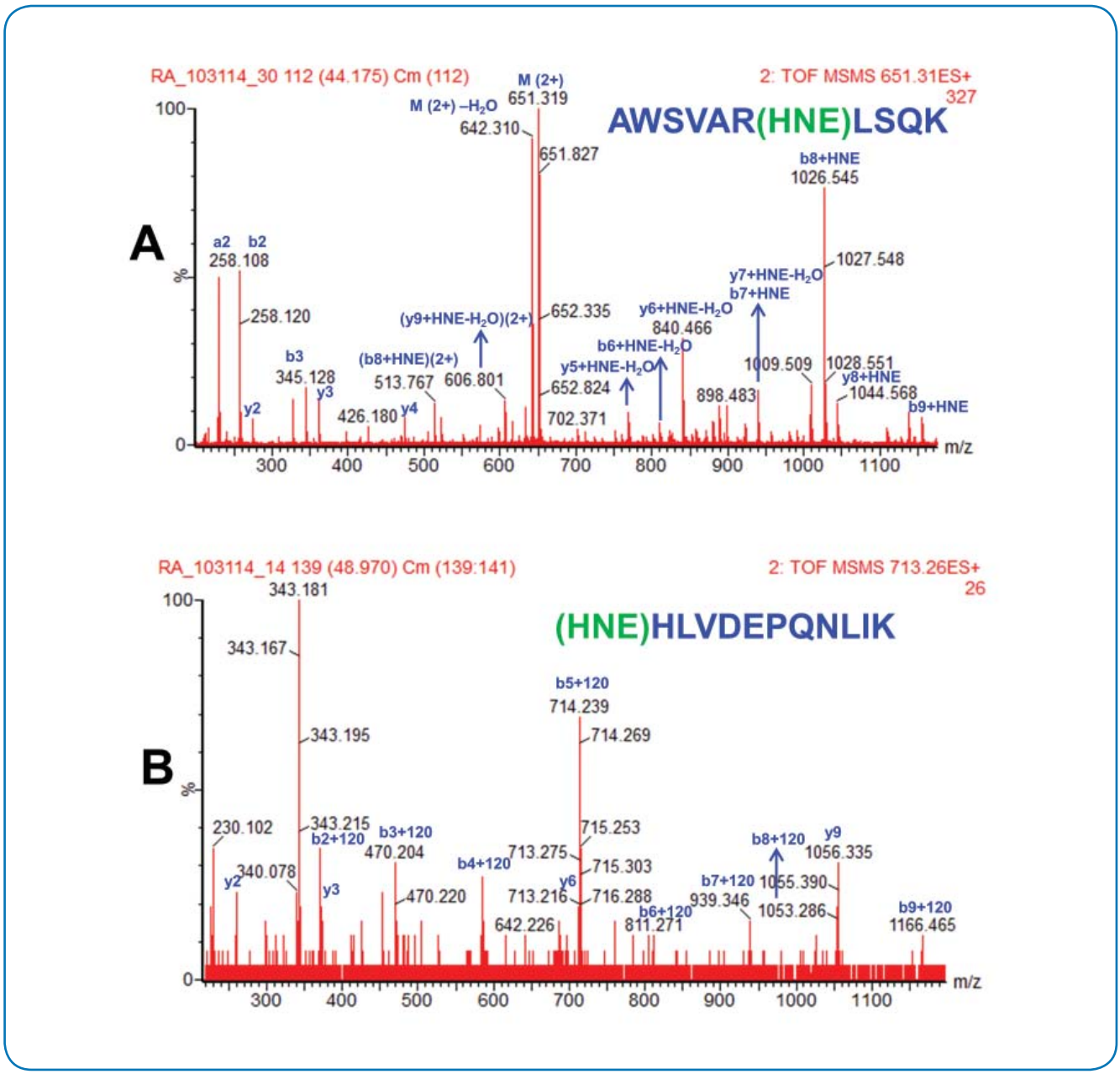

Fig. 6. – Example of HNE modification in BSA at the peptide and amino acid levels. (A) The precursor ion with m/z 651.31 (2+) was fragmented by MS/MS. Data analysis led to identification of the peptide with the amino acid sequence AWSVAR(HNE)LSQK with arginine modified by HNE; (B) In a similar experiment, a precursor ion with m/z 713.26(2+) was fragmented and data analysis led to identification of the peptide with the amino acid sequence (HNE)HLVDEPQNLIK, with N-terminal histidine modified by HNE. Reprinted and adapted with permission from (17).

identify false negatives. Therefore, many MS/MS spectra in a proteomics experiments are simply not assigned to any peptide or protein. Some of these MS/MS spectra, although of good quality, do not have any match in any database search, because of unknown PTMs. When they do have a match, these MS/MS spectra can be attributed to substoichiometric modifications like phosphorylation or glycosylation (52).

Another problem that is often encountered in MS analysis is redundant MS/MS spectra. These spectra are of good quality, but too many of them correspond to the same peptide, either unmodified or modified with experimental-borne PTMs. Such an example is presented in Fig. 7. Here, the MS/MS of the precursor ions with m/z of 890.70(3+) and m/z of 896.45(3+) (Fig. 7A-B) are almost identical and lead to the identification of the same peptide: ITAENTPLPIAGVLLPTIPGKLDVNK (Fig. 7C-D). However, after a closer inspection of the precursor ions shown in Fig. 7A, along with their MS/MS spectra shown in Fig. 8, we can conclude that in fact, all precursor ions shown in Fig. 7A are related and originate from the same peptide. For example, the peak with m/z of 890.70(3+) is in fact the peak with m/z 896.45(3+), but deaminated (a loss of an amino group of 17 Da). The peak with m/z 901.12(3+) is a modified version of the peptide that corresponds to peak with m/z of 896.45(3+),

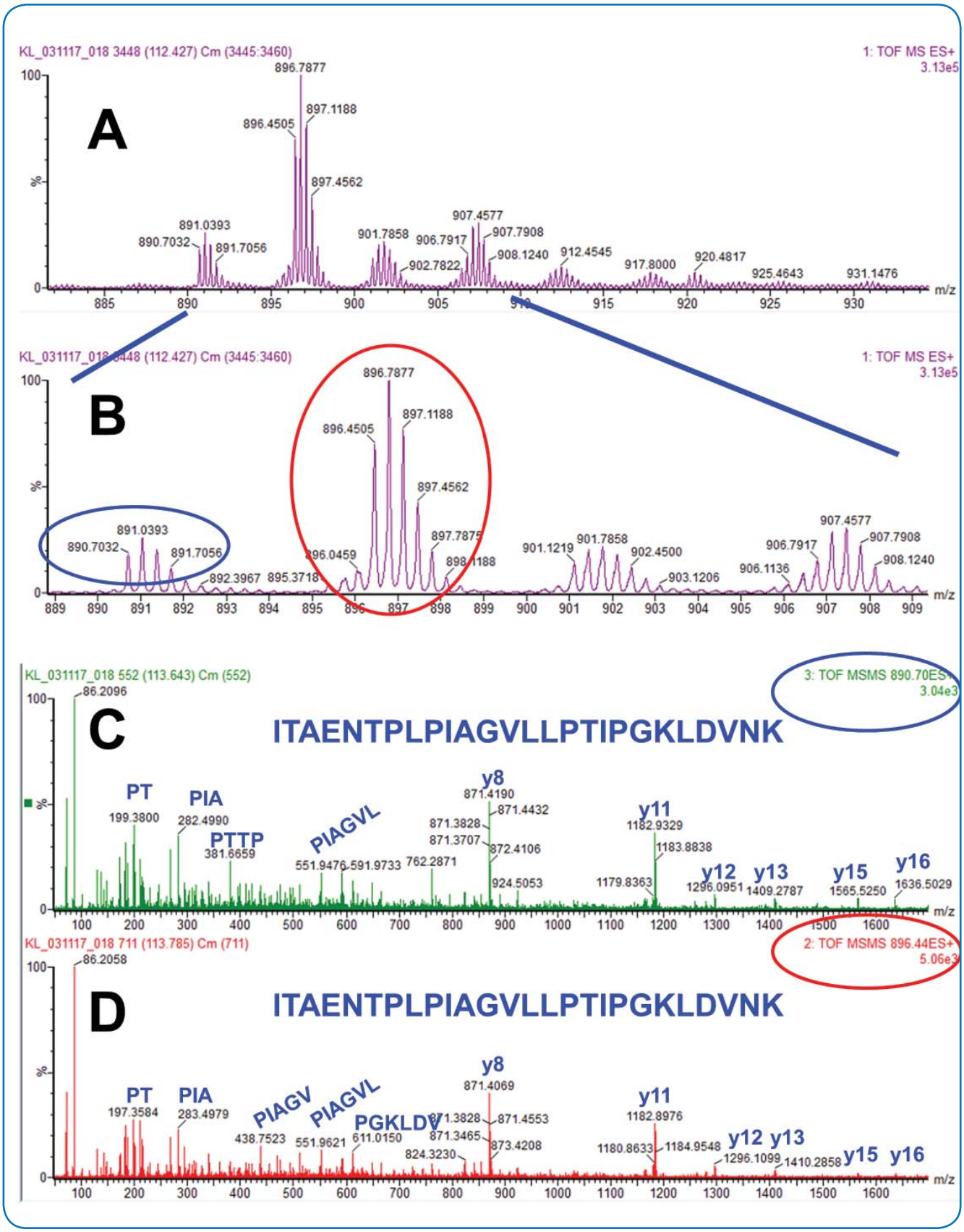

Fig. 7. – LC-MS/MS analysis of the experiment-borne PTMs. (A) The major peak with m/z of 896.45(3+) is experimentally modified by deamination or oxidation which produced peaks with m/z of 890.70(3+) and 901.12(3+), also shown in (B). Additional peaks in (A) with m/z of 906.11(3+), 911.79(3+) and 917.47(3+) may also be related to the 896.45(3+) peak. The MS/MS of peaks with m/z of 896.45(3+) and m/z of 890.70(3+) are interrelated, as demonstrated by their fragmentation in MS/MS spectra shown in (C) and (D). Note that the peak with m/z of 896.45(3+) corresponds to the unmodified peptide with the amino acid sequence ITAENTPLPIAGVLLPTIPGKLDVNK, while the peak with m/z of 890.70(3+) corresponds to the same peptide, but deaminated.

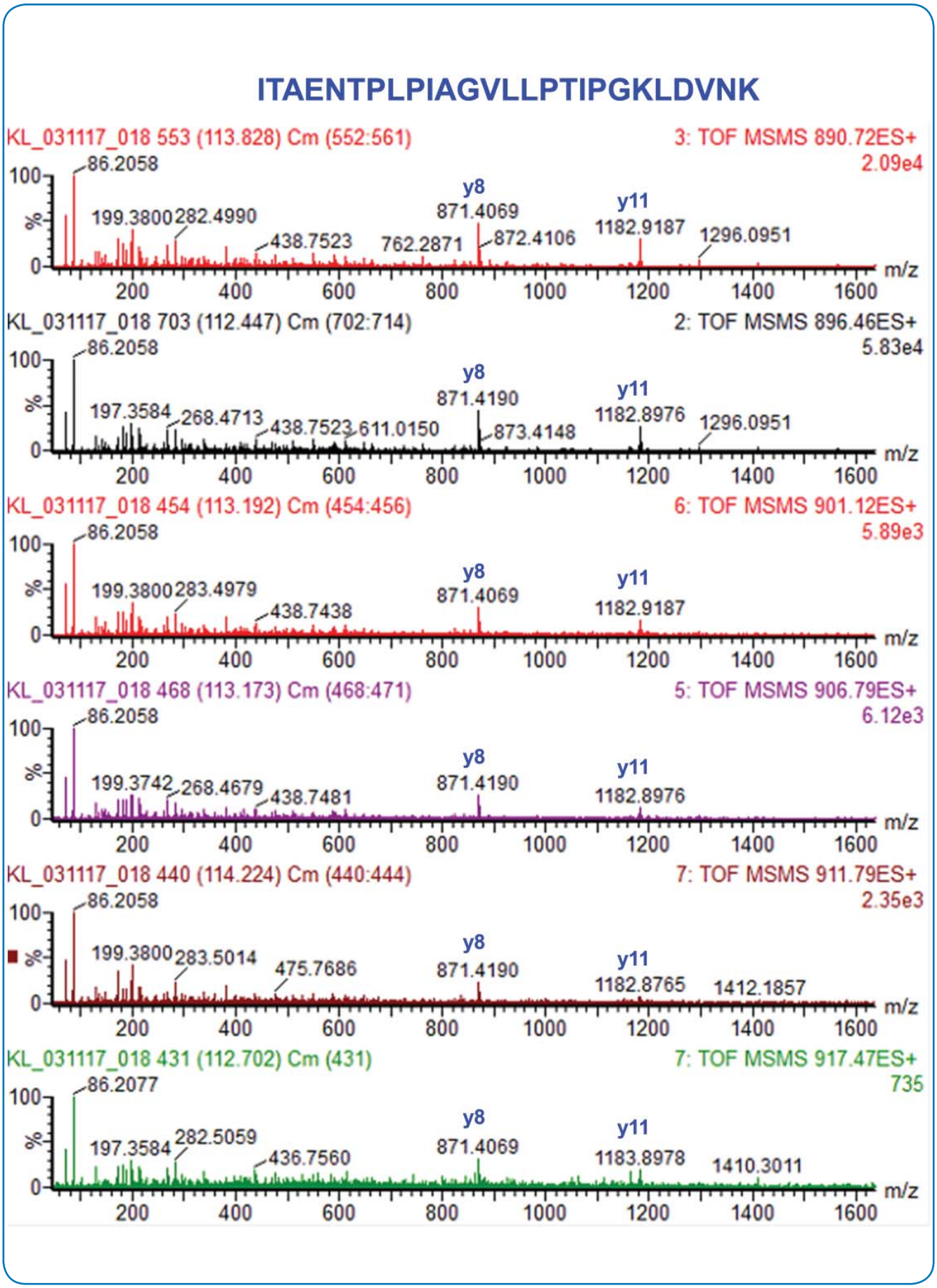

Fig. 8. – MS/MS spectra of the precursor ions 890.70(3+), 896.45(3+), 901.12(3+), 906.11(3+), 911.79(3+) and 917.47(3+), shown in Fig. 7A. The fragmentation patterns of these peaks demonstrate that they all correspond to the unmodified and experiment-based modified peptide with the amino acid sequence ITAENTPLPIAGVLLPTIPGKLDVNK.

due to the modification of a lysine residue (either delta-hydroxy-allysine (from Lysine) or oxidation of lysine (to aminoadipic acid). The same modification is also observed for peak with m/z of 906.11(3+), which derives from peak with m/z of 901.12(3+). A closer look at the precursor peaks from Fig. 7A and their corresponding MS/MS from Fig. 8 indicates that these peaks are related, but all these modifications are experimental-borne (deamination, oxidations, etc). These modifications, while apparently at a first look are not useful (do not produce any positive identification), may be helpful when full understanding of all modifications of a peptide is required (*i.e.*, quality control).

CONCLUSIONS

We have demonstrated that there are many options in identification and characterization of protein PTMs (8, 11, 12, 17, 25, 30, 49, 50, 53) and from a biochemical and proteomic perspective, characterization of common protein PTMs such as phosphorylation or acetylation is straightforward, but it becomes difficult when one attempts to identify transient PTMs such as signal transduction pathway-activated phosphorylation or uncommon protein PTMs. For example, there are challenges in identification of transient phosphorylations as well as in identification of PTMs such as nitrosylation, glycosylation, myristoylation, farnesylation, and identification of disulfide bridges (8, 11, 25, 30, 49, 50, 53). Furthermore, it is also critical to identify the correct disulfide linkage in proteins, as a particular connectivity may have one of two different protein configurations, with potentially two different activities. The simplest example is illustrated by antibodies with identical amino acid sequences, which have similar conformations, but different configurations, because their disulfide linkage is different (54).

Protein and protein PTMs characterization is always a great challenge, because the PTMs that one expects may not be the same as the ones produced by cells. The disulfide bridges are one example; the glycosylation is another example, as we discovered in our lab (55), where we demonstrated that introduction of new glycosylation sites in proteins may change the conformation and the location of the glycans on proteins.

The greatest challenge that one biochemist could have is in identification and characterization of experimental-borne PTMs. It is known that methionine, cysteine and tryptophan can be oxidized relatively easily. It is also known that hydroxyl-bearing amino acids such as serine, threonine, glutamate and aspartate can easily be dehydrated (water loss) while the amino-bearing amino acids such as arginine, lysine, asparagine and glutamine can easily be deaminated (amino loss). In addition, glutamate and glutamine can also cyclize to pyroglutamate. Furthermore, modifications such as alkylation that are expected to target only cysteine, can also modify amino acids other than cysteine and modifications such as the HNE, expected to happen at the cysteine and the amino group-bearing amino acids also target other amino acids and therefore, such modifications will not be identified by a mass spectrometer (8, 11, 12, 17, 25, 30, 49, 50, 53).

PERSPECTIVES

The human genome is known and from that we came to realize that, genetically, we are not very complicated. We can predict that from the about 30,000 genes that our genome has, about 100,000-300,000 protein isoforms are produced. However, the PTMs like glycosylation, phosphorylation, acetylation, methylation, truncation, etc., the number of proteins, protein isoforms and proteins with one or more PTMs greatly increases the number of different types of proteins that a cell, a tissue, an organ and an organism can produce. To characterize these proteins, we have the MS instrumentation that could accomplish the task of characterization of a whole proteome, however, we neither yet have the software to fully analyze the data produced by these instruments. We also lack the biological expertise to fully understand all PTMs that can occur in a cell at a particular developmental or divisional stage. However, the future of the PTM studies looks bright: PTM-omics.

Acknowledgements: We thank to our past and current lab members for their individual contribution to the success of our Biochemistry and Proteomics Lab. KLW was supported by an ASPIRE Graduate student fellowship (2014-2016). EJD was partially supported by the McNair program and an ASPIRE undergraduate student fellowship (2016-2016). MM was supported by a US-Romania Fulbright Fellowship (February-July 2017).

References

1. Aebersold R, Mann M, *Mass spectrometry-based proteomics*, ***Nature***, **422**, 198-207 (2003).
2. Aivaliotis M, Karas M, Tsiotis G, *High throughput two-dimensional blue-native electrophoresis: a tool for functional proteomics of cytoplasmatic protein complexes from Chlorobium tepidum*, ***Photosynth. Res.***, **88**, 143-157 (2006).
3. Blagoev B, Kratchmarova I, Ong SE, Nielsen M, Foster LJ, Mann M, *A proteomics strategy to elucidate functional protein-protein interactions applied to EGF signaling*, ***Nat. Biotechnol.***, **21**, 315-318 (2003).

4. Camacho-Carvajal MM., Wollscheid B, Aebersold R, Steimle V, Schamel WW, *Two-dimensional Blue native/SDS gel electrophoresis of multi-protein complexes from whole cellular lysates: a proteomics approach*, ***Mol. Cell Proteomics***, **3**, 176-182 (2004).
5. Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R, *Quantitative analysis of complex protein mixtures using isotope-coded affinity tags*, ***Nat Biotechnol.***, **17**, 994-999 (1999).
6. Ong SE, Foster LJ, Mann M, *Mass spectrometric-based approaches in quantitative proteomics*, ***Methods***, 29, 124-130 (2003).
7. Shevchenko A, Wilm M, Vorm O, Mann M, *Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels*, ***Anal. Chem.***, **68**, 850-858 (1996).
8. Zhang G, Spellman DS, Skolnik EY, Neubert TA, *Quantitative phosphotyrosine proteomics of EphB2 signaling by stable isotope labeling with amino acids in cell culture (SILAC)*, ***J. Proteome Res.***, **5**, 581-588 (2006).
9. Woods AG, Sokolowska I, Ngounou Wetie AG, Wormwood K, Aslebagh R, Patel S, Darie CC, *Mass spectrometry for proteomics-based investigation*, ***Advances in Experimental Medicine and Biology***, **806**, 1-32 (2014).
10. Darie CC, Biniossek ML, Gawinowicz MA, Milgrom Y, Thumfart JO, Jovine L, Litscher ES, Wassarman PM, *Mass spectrometric evidence that proteolytic processing of rainbow trout egg vitelline envelope proteins takes place on the egg*, ***The Journal of Biological Chemistry***, **280**, 37585-37598 (2005).
11. Darie CC, Biniossek ML, Jovine L, Litscher ES, Wassarman PM, *Structural characterization of fish egg vitelline envelope proteins by mass spectrometry*, ***Biochemistry***, **43**, 7459-7478 (2004).
12. Spellman DS, Deinhardt K, Darie CC, Chao MV, Neubert TA, *Stable isotopic labeling by amino acids in cultured primary neurons: application to brain-derived neurotrophic factor-dependent phosphotyrosine-associated signaling*, ***Mol. Cell Proteomics***, **7**, 1067-1076 (2008).
13. Wu S, Lourette NM, Tolic N, Zhao R, Robinson EW, Tolmachev AV, Smith RD, Pasa-Tolic L, *An integrated top-down and bottom-up strategy for broadly characterizing protein isoforms and modifications*, ***J. Proteome Res.***, **8**, 1347-1357 (2009).
14. Han X, Aslanian A, Yates JR, 3rd, *Mass spectrometry for proteomics*, ***Curr. Opin. Chem. Biol.***, **12**, 483-490 (2008).
15. McLafferty FW, Breuker K, Jin M, Han X, Infusini G, Jiang H, Kong X, Begley TP, *Top-down MS, a powerful complement to the high capabilities of proteolysis proteomics*, ***The FEBS Journal***, **274**, 6256-6268 (2007).
16. Honarvarfard E, Gamella M, Channaveerappa D, Darie CC, Poghossian A, Schöning MJ, Katz E, *Electrochemically Stimulated Insulin Release from a Modified Graphene–functionalized Carbon Fiber Electrode*, ***Electroanalysis***, n/a-n/a 2017).
17. Aslebagh R, Pfeffer BA, Fliesler SJ, Darie CC, *Mass spectrometry-based proteomics of oxidative stress: Identification of 4-hydroxy-2-nonenal (HNE) adducts of amino acids using lysozyme and bovine serum albumin as model proteins*, ***Electrophoresis***, **37**, 2615-2623 (2016).
18. Savitski MF, Savitski MM, *Unbiased detection of posttranslational modifications using mass spectrometry*, ***Methods Mol. Biol.***, **673**, 203-210 (2010).
19. Spiro RG, P*rotein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds*, ***Glycobiology***, **12**, 43R-56R (2002).
20. Marino K, Bones J, Kattla JJ, Rudd PM, *A systematic approach to protein glycosylation analysis: a path through the maze*, ***Nat. Chem. Biol.***, **6**, 713-723 (2010).
21. Read EK, Park JT, Brorson KA, *Industry and regulatory experience of the glycosylation of monoclonal antibodies*, B***iotechnol. Appl. Biochem.***, **58**, 213-219 (2011).
22. Kamoda S, Kakehi K, *Evaluation of glycosylation for quality assurance of antibody pharmaceuticals by capillary electrophoresis*, ***Electrophoresis***, 29, 3595-3604 (2008).
23. Pan S, Chen R, Aebersold R, Brentnall TA, *Mass spectrometry based glycoproteomics--from a proteomics perspective*, ***Mol. Cell Proteomics***, **10**, R110 003251 (2011).
24. Morelle W, Michalski JC, *Analysis of protein glycosylation by mass spectrometry*, ***Nat. Protoc.***, **2**, 1585-1602 (2007).
25. Wuhrer M, Catalina MI, Deelder AM, Hokke CH, *Glycoproteomics based on tandem mass spectrometry of glycopeptides*, ***J. Chromatogr. B. Analyt. Technol. Biomed. Life Sci.***, **849**, 115-128 (2007).
26. Robu AC, Vukelic Z, Schiopu C, Capitan F, Zamfir AD, *Mass spectrometry of gangliosides in extracranial tumors: Application to adrenal neuroblastoma*, ***Anal. Biochem.***, **509**, 1-11 (2016).
27. Sarbu M, Robu A, Peter-Katalinic J, Zamfir AD, *Automated chip-nanoelectrospray mass spectrometry for glycourinomics in Schindler disease type I*, ***Carbohydrate Research***, **398**, 90-100 (2014).
28. Mechref Y, Madera M, Novotny MV, *Glycoprotein enrichment through lectin affinity techniques*, ***Methods Mol. Biol.***, **424**, 373-396 (2008).
29. Alley WR Jr, Mechref Y, Novotny MV, *Characterization of glycopeptides by combining collision-induced dissociation and electron-transfer dissociation mass spectrometry data*, ***Rapid communications in mass spectrometry: RCM***, **23**, 161-170 (2009).
30. Sokolowska I, Ngounou Wetie AG, Roy U, Woods AG, Darie CC, *Mass spectrometry investigation of glycosylation on the NXS/T sites in recombinant glycoproteins*, ***Biochim. Biophys. Acta***, **1834**, 1474-1483 (2013).
31. Gonzalez J, Takao T, Hori H, Besada V, Rodriguez R, Padron G, Shimonishi Y, *A method for determination of N-glycosylation sites in glycoproteins by collision-induced dissociation analysis in fast atom bombardment mass spectrometry: identification of the positions of carbohydrate-linked asparagine in recombinant alpha-amylase by treatment with peptide-N-glycosidase F in 18O-labeled water*, ***Anal. Biochem.***, **205**, 151-158 (1992).
32. Bond MR, Kohler JJ, *Chemical methods for glycoprotein discovery*, ***Curr. Opin. Chem. Biol.***, **11**, 52-58 (2007).
33. Kaji H, Saito H, Yamauchi Y, Shinkawa T, Taoka M, Hirabayashi J, Kasai K, Takahashi N, Isobe T, *Lectin affinity capture, isotope-coded tagging and mass spectrometry to identify N-linked glycoproteins*, ***Nat. Biotechnol.***, **21**, 667-672 (2003).
34. Tarrant MK, Cole PA, *The chemical biology of protein phosphorylation*, ***Annu. Rev. Biochem.***, **78**, 797-825 (2009).
35. Blume-Jensen P, Hunter T, *Oncogenic kinase signalling*, ***Nature***, **411**, 355-365 (2001).
36. Cohen P, *The role of protein phosphorylation in human health and disease. The Sir Hans Krebs Medal Lecture*, ***Eur. J. Biochem.***, **268**, 5001-5010 (2001).
37. Badiola N, Suarez-Calvet M, Lleo A, *Tau phosphorylation and aggregation as a therapeutic target in tauopathies*, ***CNS Neurol. Disord. Drug Targets***, **9**, 727-740 (2010).
38. Cohen P, *Protein kinases--the major drug targets of the twenty-first century?*, ***Nat. Rev. Drug Discov.***, **1**, 309-315 (2002).
39. Strebhardt K, *Multifaceted polo-like kinases: drug targets and antitargets for cancer therapy*, ***Nat. Rev. Drug Discov.***, **9**, 643-660 (2010).
40. Le Blanc JC, Hager JW, Ilisiu AM, Hunter C, Zhong F, Chu I, *Unique scanning capabilities of a new hybrid linear ion trap mass spectrometer (Q TRAP) used for high sensitivity proteomics applications*, ***Proteomics***, **3**, 859-869 (2003).
41. Unwin RD, Griffiths JR, Leverentz MK, Grallert A, Hagan IM, *Whetton AD, Multiple reaction monitoring to identify sites of protein phosphorylation with high sensitivity*, ***Mol. Cell Proteomics***, **4**, 1134-1144 (2005).
42. Boersema PJ, Mohammed S, Heck AJ, *Phosphopeptide fragmentation and analysis by mass spectrometry*,

Journal of mass spectrometry: JMS, **44**, 861-878 (2009).
43. Xu CF, Lu Y, Ma J, Mohammadi M, Neubert TA, *Identification of phosphopeptides by MALDI Q-TOF MS in positive and negative ion modes after methyl esterification*, ***Mol. Cell Proteomics***, **4**, 809-818 (2005).
44. Ngounou Wetie AG, Sokolowska I, Woods AG, Roy U, Deinhardt K, Darie CC, *Protein-protein interactions: switch from classical methods to proteomics and bioinformatics-based approaches*, ***Cellular and Molecular Life Sciences: CMLS*** (2013).
45. Ngounou Wetie AG, Sokolowska I, Woods AG, Roy U, Loo JA, Darie CC, *Investigation of stable and transient protein-protein interactions: Past, present, and future*, ***Proteomics***, **13**, 538-557 (2013).
46. Woods AG, Sokolowska I, Yakubu R, Butkiewicz M, LaFleur M, Talbot C, Darie CC, *Blue native page and mass spectrometry as an approach for the investigation of stable and transient protein-protein interactions*, in: ***Oxidative Stress: Diagnostics, Prevention, and Therapy***, Andreescu S, Hepel M, Eds., 2011, American Chemical Society, Washington, D.C.
47. Beltran L, Cutillas PR, *Advances in phosphopeptide enrichment techniques for phosphoproteomics*, ***Amino Acids***, **43**, 1009-1024 (2012).
48. Corthals GL, Aebersold R, Goodlett DR, *Identification of phosphorylation sites using microimmobilized metal affinity chromatography*, ***Methods Enzymol.***, **405**, 66-81 (2005).
49. Ngounou Wetie AG, Sokolowska I, Woods AG, Darie CC, *Identification of post-translational modifications by mass spectrometry*, ***Australian Journal of Chemistry***, **66**, 734-748 (2013).
50. Woods AG, Sokolowska I, Darie CC, *Identification of consistent alkylation of cysteine-less peptides in a proteomics experiment*, ***Biochemical and Biophysical Research Communications***, **419**, 305-308 (2012).
51. Nielsen ML, Vermeulen M, Bonaldi T, Cox J, Moroder L, Mann M, I*odoacetamide-induced artifact mimics ubiquitination in mass spectrometry*, ***Nature methods***, **5**, 459-460 (2008).
52. Chick JM, Kolippakkam D, Nusinow DP, Zhai B, Rad R, Huttlin EL, Gygi SP, *A mass-tolerant database search identifies a large proportion of unassigned spectra in shotgun proteomics as modified peptides*, ***Nat. Biotechnol.***, **33**, 743-749 (2015).
53. Sokolowska I, Ngounou Wetie AG, Woods AG, Darie CC, *Automatic determination of disulfide bridges in proteins*, ***Journal of Laboratory Automation***, **17**, 408-416 (2012).
54. Jones LM, Zhang H, Cui W, Kumar S, Sperry JB, Carroll JA, Gross ML, *Complementary MS Methods Assist Conformational Characterization of Antibodies with Altered S-S Bonding Networks*, ***Journal of the American Society for Mass Spectrometry***, **24**, 835-845 (2013).
55. Sokolowska I, Wetie AGN, Woods AG, Darie CC, *Applications of Mass Spectrometry in Proteomics*, ***Australian Journal of Chemistry***, **66**, 721-733 (2013).