

Chapter 29

Bottlenecks in Proteomics

Armand G. Ngounou Wetie, Devon A. Shipp, and Costel C. Darie

Abstract Mass spectrometry (MS) is the core for advanced methods in proteomic experiments. When effectively used, proteomics may provide extensive information about proteins and their post-translational modifications, as well as their interaction partners. However, there are also many problems that one can encounter during a proteomic experiment, including, but not limited to sample preparation, sample fractionation, sample analysis, data analysis & interpretation, and biological significance. Here we discuss some of the problems that researchers should be aware of when performing a proteomic experiment.

29.1 Introduction

Mass spectrometry (MS) is the core for advanced methods in proteomic experiments. Integration of MS with a variety of other analytical methods has made it possible to examine virtually all types of samples derived from tissues, organs, and organisms and has led to the identification and quantification of thousands of proteins and peptides from complex biological samples. Unfortunately, due to expense and time requirements, MS-based proteomics remain prohibitive for many labs that could greatly benefit from this technology.

A.G. Ngounou Wetie • C.C. Darie (✉)
Biochemistry & Proteomics Group, Department of Chemistry & Biomolecular Science,
Clarkson University, 8 Clarkson Avenue, Potsdam, NY 13699-5810, USA
e-mail: cdarie@clarkson.edu

D.A. Shipp
Polymer Science & Materials Group, Department of Chemistry & Biomolecular Science,
Clarkson University, 8 Clarkson Avenue, Potsdam, NY 13699-5810, USA

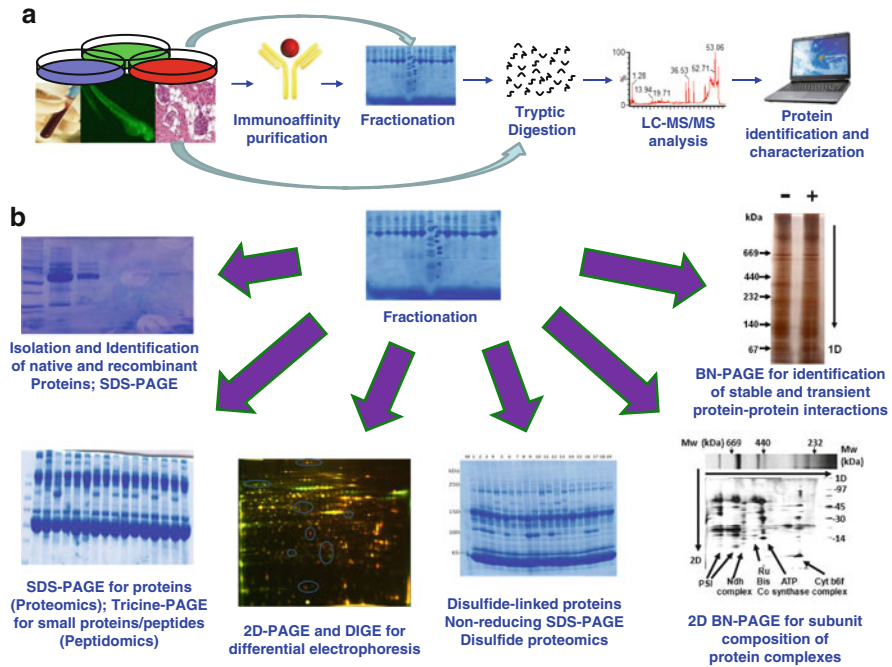


Fig. 29.1 (a) Common steps of a proteomic experiment. In many experiments, the samples are fractionated by electrophoresis and then digested in gel. When the biological materials are fluids (e.g., blood, saliva, etc), the electrophoresis-based fractionation step may be bypassed. (b) The types of protein samples and the types of electrophoresis that could benefit from a degradable gel

Proteomics is the study of proteome which is the whole protein complement in a cell or organism at any given time [1–8]. The proteome of an organism or even a single type of cell is much more complex than its corresponding genome. This is mostly due to alternation that can be introduced via alternate splicing and post-translational modifications (PTMs) which affect virtually all proteins. A proteome differs from cell to cell and from time to time with composition depending on the physiological or pathological state of cells or organisms. Understanding that proteins are the actual effectors of biological function has been an essential part of biochemistry for over a 100 years [9]. Due to complexity, the analysis of a proteome is extremely challenging. Therefore, modern biochemical technologies with improved separation and identification methods need to be introduced.

The workflow in a proteomic experiment involves sample fractionation by a biochemical approach, followed by enzymatic digestion (usually trypsin), peptide extraction, and MS analysis (Fig. 29.1a) [10–17]. When the peptide mixture is analyzed by MALDI-MS, the proteins of interest are identified using a procedure named peptide mass fingerprinting. Alternatively, the peptide mixture is further fractionated by HPLC on different columns (usually reverse-phase-HPLC or -LC), followed by ESI-MS analysis. The combination of LC and ESI-MS is usually named LC-MS/MS, and analysis of a protein using this approach provides not only the

protein identity, but also sequence information for that particular protein. In addition to qualitative information provided by MALDI-MS or LC-MS/MS analysis, MS may also provide quantitative information about a particular protein. Methods such as differential gel electrophoresis (DIGE) [18], isotope-coded affinity tag (ICAT) [19], stable isotope labeling by amino acids in cell culture (SILAC) [20], absolute quantitation (AQUA) [21], multiple reaction monitoring (MRM) [22], or spectral counting [23] allow detection, identification, and quantification of proteins or peptides. These methods are currently known as functional proteomics and are widely used in basic research.

Another dimension to identification and characterization of proteins is added due to the intensive PTMs of proteins. It is a great challenge to fully identify a PTM pattern at any given time in cells, tissues, and organisms and MS-based proteomics became the method of choice for their detection and characterization [24]. PTMs are important to virtually all biological processes. For example, glycosylation enhances many biological processes such as cell–cell recognition and influences the protein’s biological activity [25, 26], while phosphorylation is a reversible and common PTM that plays a role in controlling and modifying the majority of cellular processes [27]. Another important modification of proteins is the formation of disulfide bridges, which plays a significant role in maintaining correct protein function [28, 29]. Additional stable or transient modifications in proteins include acetylation, ubiquitination, and methylation, which when combined with stable or transient protein–protein interactions (PPIs) [30], add an additional level of complexity to proteomic approaches [31].

29.2 Bottleneck # 1: Sample Preparation and Separation

Purification and fractionation steps are to be carefully selected to avoid loss of samples and introduction of errors, which would be reflected later in the MS analysis. Usually, the first step in proteomic workflows is the extraction of proteins from tissues, cells, or biological fluids. At this stage of sample preparation, a loss of material is usually encountered as the homogenization or alternative extraction procedures (e.g., centrifugation) are not very effective. Coevally, one should attempt to reduce sample complexity and matrix effects by isolating and enriching for the fractions of interest (subproteomes, organelle proteomes). To this end, subcellular fractionation is sometimes undertaken. However, any additional sample preparation steps that are performed increase the risk of introducing technical variability and contamination in the analysis. Also, the physiological state/time point of protein extraction from the source plays an important role in the outcome of the study and the biological significance that can be inferred from it. Therefore, the design of proteomic studies should be thoroughly thought through in order to draw meaningful conclusions from the experiments. Further, it should be noted that most sample preparation and purification techniques discriminate against “extreme” proteins (very hydrophobic/hydrophilic proteins, too low/high pI, low abundant, membrane proteins).

29.3 Bottleneck # 2: Gel Fractionation

Key parameters of MS-based proteomic experiments are sensitivity, resolution, dynamic range, and mass accuracy. Therefore, various elements need to be taken into consideration during a typical MS-based proteomic experiment (Fig. 29.1). One of the most important bottlenecks in these experiments is always the sample fractionation and processing, which happens prior to MS analysis. Sample fractionation is primarily performed by some kind of electrophoresis. Sample processing involves in-gel protein digestion, followed by peptide extraction and concentration, and MS analysis. This is a very time-consuming procedure. An alternative to electrophoresis and in-gel digestion steps is the in-solution digestion of proteins. The difference between the two methods is time and labor, as well as the scientific outcome. For example, protein fractionation by electrophoresis and in-gel digestion of proteins, up to the MS analysis step can take up to 3 days, while the in-solution digestion of proteins up to the MS analysis can take only about 5 h. The labor required when the electrophoresis step and in-gel digestion is used can be as much as 24 h of labor, while the in-solution digestion step typically requires a maximum of 3 h. The in-solution digestion step is advantageous over the electrophoresis and in-gel digestion step because of the time and labor savings. The reason for which many scientists do not use in-solution digestion is because of the scientific outcome. For example, without the SDS-PAGE electrophoretic fractionation step, one cannot create molecular mass constraints to monitor particular proteins (i.e., the focus is only on low abundant 50 kDa protein), simply because the samples without fractionation are analyzed as a whole and the low abundant proteins may not be identified by MS analysis. In addition, when the proteins are investigated for identification of their amino acid sequence information or for identification of their PTMs such as phosphorylation, acetylation, or glycosylation, and the proteins that bear these PTMs are not abundant, the SDS-PAGE step is almost mandatory. Furthermore, when the stable and transient PPIs are investigated, native gel electrophoresis step is a must for biochemical fractionation.

Native gel electrophoresis (Blue Native-PAGE or BN-PAGE and Colorless Native-PAGE or CN-PAGE) allows protein complexes from various sources to be separated according to their molecular weight (MW) and external charge (BN-PAGE) or based on the internal charge of subunits in protein complexes (CN-PAGE) [32, 33]. Therefore, to solve one of the most critical bottlenecks in proteomics, it would be ideal to combine the power of protein fractionation (electrophoresis) with the speed (low time and labor) of in-solution digestion, and to obtain the optimal scientific outcome, similar to the electrophoresis and in-gel digestion procedure. In other words, to use electrophoresis step for protein fractionation but not use the in-gel digestion; but to rather dissolve the gel followed by the low-time, low-labor in-solution digestion step and MS analysis.

One solution to this could be to use degradable gels, thus addressing the bottleneck issue and bypassing the in-gel protein digestion and peptide fragmentation by simply replacing the current electrophoresis gel crosslinker (*N,N'*-methylenebisacrylamide (MBA)) with a cleavable one. There are various types of

gel monomers used to build electrophoresis gels. The most frequently used are an acrylamide-MBA mixture, with MBA used as a crosslinker. Alternatives to MBA as a crosslinker are cleavable crosslinkers, which can be used to generate reversible polyacrylamide gels. They can be divided into those cleavable by oxidation (e.g., *N,N'*-1,2-dihydroxyethylenebisacrylamide or *N,N'*-diallyl-tartar-diamide) [34, 35] and those which undergo reductive cleavage (e.g., *N,N'*-bisacryloyl-cystamine) [36]. Recently, a modified gel system based on co-polymerization of acrylamide with MBA and ethylene-glycol-diacrylate to offer controllable pore size was also communicated [37]. However, there are numerous problems with these crosslinkers, which are far from being optimal for proteomic studies. For example, one of these crosslinkers (*N,N'*-bis(acryloyl)cystamine) is a disulfide-linked (commercially available) crosslinker. This crosslinker can be reduced by incubating the gel or gel piece in a reducing agent such as dithiothreitol or beta-mercaptoethanol. The end-effect is disintegration of the gel piece and release of the protein. However, experiments in our lab identified several problems with this crosslinker: (1) since this is a disulfide linker, the gel has to be run under non-reducing conditions, thus preventing one to investigate the cysteine-containing proteins, (2) the protein recovery efficiency is very low (we recovered about 10–20 % of the initially loaded protein), (3) the proteins that are extracted from the gel are over-alkylated, thus complicating the protein identification due to artificial alkylation, (4) the method is environmentally unfriendly, since it creates a large amount of reducing agent (e.g., beta-mercaptoethanol), (5) the protein is very diluted and requires concentration (the protein within the gel piece is usually incubated in 100 mM beta-mercaptoethanol). Therefore, designing new crosslinkers that are compatible with proteomic experiments, does not artificially modify proteins, allows the protein to be recovered with a high efficiency, and is both time- and labor efficient is important and would solve one of the most critical bottlenecks in proteomics.

29.4 Bottleneck # 3: Ionization

Another important bottleneck in MS-based proteomics is related to the ionization method. It is imperative to choose the appropriate matrices (MALDI), solvents, and salt for a specific analysis under consideration (ESI). Of the two mostly used soft ionization techniques (MALDI and ESI), it has been shown that ESI is the more appropriate for fragile and/or labile biomacromolecules. With ESI, in-source fragmentation is limited and PTMs as well as non-covalent attachments are preserved, whereas MALDI is more prone to loss of information. Another issue with MALDI is the high background generated by matrix ions that increases the noise and could hamper identification of low-abundant biomolecules. It has been suggested that the organic matrix could be replaced with an inorganic nanomaterial surface to increase the signal-to-noise ratio [38].

29.5 Bottleneck # 4: Data Analysis and Interpretation

Post-MS analysis, a huge amount of data generated has to be processed and converted into biological meaning. This requires understanding of the fragmentation pattern of the biomacromolecules at hand so that useful structural information can be obtained from the data. This task can be undertaken manually (this is time-consuming and requires a lot of expertise and skills) or can be done with the help of a multitude of available software. The goal here will be to extract biological information from mass spectrometric data through a completely simple automated approach that requests little time and effort from the user. One particular issue encountered during data analysis is the different outcome resulting from the analysis of MS data using different software which makes the interpretation of results somehow uncertain. Though these software programs are mostly very sophisticated, they are not always easy to use and to understand [10, 39]. In the case of MS/MS, the availability of reference spectra is of critical importance. Identification of novel proteins or copolymers is rendered difficult or could even be missed if corresponding reference spectra are absent. In this case, one must manually investigate the data in order to identify these novel entities. This is, for example, observable during Mascot searches, where peptides with very good scores are not assigned by the software for reasons mentioned above or some other unknown factors (organism with incomplete or unavailable genome sequence, novel isoforms of a protein). Therefore, researchers have to perform at least two searches (e.g., Mascot and Sequest) for confident identification [40, 41]. This holds true also for software used for visualization/presentation or quantitation of MS raw data such as Scaffold or DTASElect [42, 43].

As for the analysis of synthetic or derivatives of biomacromolecules, tandem MS libraries should be created and made accessible to the whole scientific community. Moreover, an effort should be undertaken to gather and organize all the published MS-based proteomic data in an openly accessible database so that researchers can gain access to them for analysis, evaluation, reproducibility, interpretation, and extraction of information. The idea is to store proteomic data obtained with different MS instruments in a single global database in a format that is compatible with a free online tool. An example of such a format is “mzXML,” an open, generic XML (extensible markup language) for MS, MS/MS, and MSⁿ data output. We believe that this should be a requirement for all manuscript submissions to proteomic journals. The idea of a centralized, organized, structured, and openly accessible MS-data storage center will advance proteomic research tremendously [44–46].

29.6 Bottleneck # 5: False Positives, False Negatives, and Unassigned Spectra

Although this is really part of data analysis, this topic deserves special consideration. It is a great challenge to do a database search (i.e., using Mascot database search engine) and to realize that many, sometimes too many MS/MS spectra are unassigned to any peptide. In addition, some MS/MS spectra that correspond to a

peptide are not in the Mascot database search, due to the search parameters, i.e., an oxidized peptide will not be identified if we do not choose to search for it, or a peptide with two missed tryptic cleavage sites will not be identified if we use only one missed cleavage site during the search. Furthermore, searching in Mascot database with or without the decoy function eliminates some true positive spectra and only inspection of the raw data elucidates whether a MS/MS spectrum indeed corresponds to a peptide with a particular amino acid sequence or not. While all these “little bottlenecks” are not a problem and for most of them there is already a solution such as using the variable modification option for methionine oxidation or using two missed cleavage sites for trypsin, or submitting the MS/MS spectra as supplementary data for a manuscript submitted for publication, the largest problem for which there is still no solution is the natural and artificial modification of the peptides, which leads to two major problems: (1) a perfectly fine MS/MS has no peptide assigned because the peptide contains an artificial modification such as iodoacetamide modification of the N-terminal amino acids within peptides, leading to an unassigned MS/MS spectrum and an unidentifiable peptide and (2) artificial alteration of the peptides by experiment-induced peptide modification, which leads to errors in data quantitation. Again, in this case, the same example is given: when a peptide is alkylated at the non-cysteine amino acids, the amount of the precursor that corresponds to the unmodified peptide is sometimes very low. Such an example of artificial peptide modification was published by our lab with one example given in Fig. 29.2.

29.7 Bottleneck # 6: Instrumentation

MS instruments are very costly, delicate, and require a lot of maintenance and troubleshooting. This is one of the main reason MS is usually compared to NMR, which is thought to have the upper hand with regard to high-throughput, robustness, and the relative simplicity of the sample preparation steps. However, NMR cannot match the sensitivity and structural ability of MS [47, 48]. Additionally, one needs to possess a diverse range of instrument types in order to cover all possible applications, as one specific instrument cannot perform all kinds of experiments. Furthermore, there is still a need for mass spectrometers with better sensitivity, accuracy, and resolution. For example, dynamic ranges obtained by current MS instruments oscillates between 10^4 and 10^5 , while the concentration range of proteins in human blood plasma is about 10^{12} [49, 50]. The aforementioned limitation is also related to another challenge in proteomics, which is the bias towards high-abundant and soluble proteins leading to the non-detection of disease-relevant low-abundant proteins [51]. This is also true for low-abundant PTMs or PPIs in complex samples. To partly remedy this situation, it is often recommended to operate one or more fractionation steps (usually chromatography) or enrichment of selected species (antibody-based, IMAC, chemical methods) prior to MS as a means of reducing sample complexity.

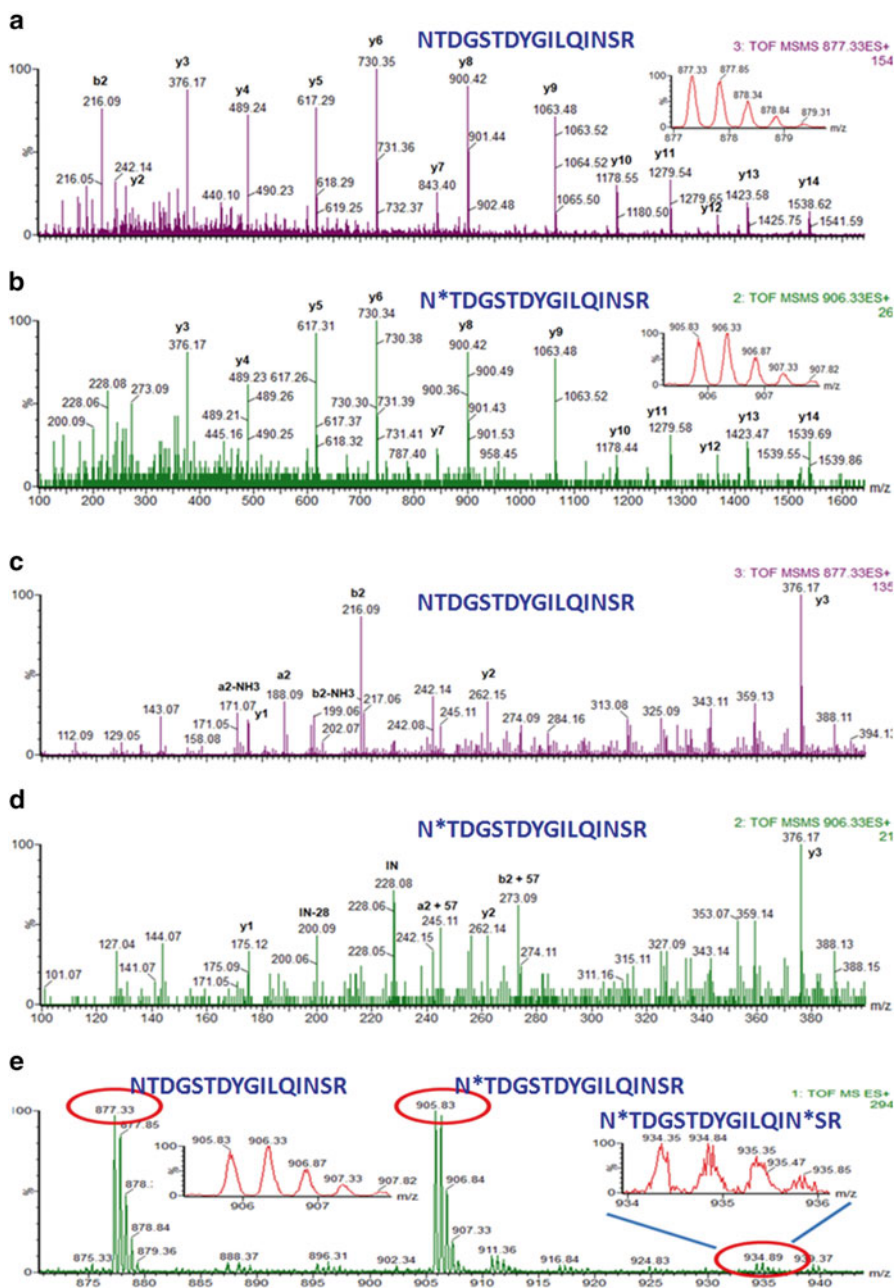


Fig. 29.2 NanoLC-MS/MS analysis of an SDS-PAGE gel band that was reduced by dithiothreitol (DTT), alkylated by iodoacetamide (IAA) and digested by trypsin. (a) MS/MS of the doubly-charged precursor ion with m/z of 877.33 produced a series of b and y product ions that led to the identification of the peptide with sequence NTDGSTDYGILQINSR that was part of the lysozyme protein. This peptide was identified during the Mascot database search. (b) MS/MS of the doubly-charged

29.8 Bottleneck # 7: Biological Significance

Perhaps one of the biggest bottlenecks in proteomics is the assignment of a biological role and of a biological, physiological, pathological, or clinical significance to a proteomic experiment. For example, when one performs a proteomics experiment, the outcome is a list of identified proteins or a list of ratios for a number of proteins. However, these proteins are not deeply investigated; having a list of proteins in a proteomic experiment is far from reflecting the physiological phenomenon that happens in an organelle, cell, tissue, organ, or organism. Information about the PTMs or interaction partners of these proteins is often not investigated. Instead, researchers focus on the investigation of several proteins (usually 3–5) by one (relative quantitation using the precursor ions or Western blotting) or more (Western blotting and immunofluorescence) methods, sometimes associated with functional experiments such as down-/up-regulation of these proteins, mutation of a protein, or transient/stable transfection of the DNA that encodes that protein. In addition, researchers rarely investigate more than 4–5 proteins in one proteomic experiment; in the proteomic world, these proteins are called the “validated proteins.” While it is good to focus on a few proteins that are identified as crucial in a proteomic experiment, the other proteins are almost never investigated. The outcome of a proteomic dataset is published and almost never re-investigated for other proteins. Therefore, the biological significance in a dataset is explored only with regard to a few proteins; the rest of the information, even if it is deposited in a proteomic database (i.e., PRIDE), is lost or is almost never investigated. As such, a large amount of proteomic data is produced, but a very little amount is truly investigated.

Although not a bottleneck, another problem with assignment of a biological significance is the enzymatic activity of a protein and the position of a modification such as phosphorylation, acetylation, or methylation within the three-dimensional structure of that protein. One of these modifications, phosphorylation, is not only an important functional part of proteomics, but also an established proteomic subfield (phosphoproteomics). Phosphoproteomics is an important functional part of proteomics, and a method that quantitatively compares the phosphorylation of proteins from two different conditions, e.g., unstimulated and stimulated *in vitro*-grown cells. However, it is now clear that phosphorylation of a protein does not have



Fig. 29.2 (continued) precursor ion with m/z of 905.83 produced a series of b and y product ions that led to the identification of the mono-alkylated peptide with sequence N*TDGSTDYGILQINSR. This modification led to a 57 Da increase in the m/z of the precursor ion, compared with the precursor ion shown in (a). (c) The low mass range of the MS/MS spectrum shown in (a). (d) The low mass range of the MS/MS spectrum shown in (b). The *star* on the amino acids such as N*TDGSTDYGILQINSR indicates that the marked amino acid is alkylated. (e) MS spectrum showing the doubly-charged precursor ions with m/z of 877.33, m/z of 905.83, and m/z of 934.35 that corresponds to the unmodified peptide NTDGSTDYGILQINSR, mono-alkylated peptide N*TDGSTDYGILQINSR, and di-alkylated peptide N**TDGSTDYGILQIN**SR. Reproduced with permission from [52]

biological significance in itself, but rather the amino acid residues that are phosphorylated, the number of amino acids phosphorylated in one protein at a particular physiological condition, and the role in inserting a negative charge within a particular region(s) of the protein, as well as the biological effect that is observed or monitored. This information is almost never taken into account by any proteomic researcher. Yet, the role of protein phosphorylation and the number of phosphorylations in a protein, as well as the position of the phosphate group in a three dimensional structure of a protein seriously affects its enzymatic activity, as reflected in the biochemistry textbooks for glycogen synthase. Additional information associated with protein phosphorylation is also not considered. For example, protein truncation is almost never investigated in a regular proteomic experiment. Similarly, association of the protein phosphorylation with PPIs (particularly transient ones), is also not taken into account. As such, protein phosphorylation may also be considered a bottleneck in proteomics with regard to the interpretation of the biological significance of a phosphorylation of a specific protein, at a specific amino acid residue within a protein, at a particular time.

29.9 Conclusions

MS-based techniques have been employed for the comprehensive analysis of proteins. These methods have enabled multiple discoveries and significant advances in the biomedical field. However, it is only in combination with other omics disciplines (lipidomics, glycomics, transcriptomics, metabonomics) that proteomics will be able to provide complete answers to pressing biomedical questions of our time. We even envision that MS-based techniques could be expanded to the analysis of synthetic biomolecules, polymers, and other chemical entities. This has been mainly possible due to the recent development of better performing instrumentation. Nonetheless, there is still plenty of room for improvements in terms of instrumentation. An example is the fact that currently available instruments can only sequence precursor ions up to 5,000 m/z . Instruments that could go beyond this present limit are highly desirable for better identifications and sequence coverage. Further, there are not globally acceptable standards for the submission of proteomics data. For instance, the number of required technical and biological replicates varies from journal to journal. In some cases, two is enough and in other cases three is the minimum requested. One thing that has been shown in numerous studies is the much larger biological variability compared to the rather less significant technical variability. Another controversial point is the mention of false discovery rates (FDR) and where the threshold for this parameter should lie. Moreover, MS results should always be validated, preferentially with a different method such as immunoblotting or ELISA. However, MS methods like single/multiple reaction monitoring (SRM/MRM) are becoming more and more employed for validation purposes.

Though not mentioned as a bottleneck, the lack of adequate training for graduate students in this field is an important aspect hampering the large-scale use and

understanding of MS data analysis. Many students that work with mass spectrometry are good users, but do not understand in depth the complexities and challenges associated with mass spectrometry. There should be a transfer of knowledge from established MS experts towards interested graduate students.

Acknowledgements We would like to thank Ms. Laura Mulderig, Scott Nichols, and their colleagues (Waters Corporation) for their generous support in setting up the Proteomics Center at Clarkson University. CCD thanks Drs. Thomas A. Neubert (New York University), Belinda Willard (Cleveland Clinic), and Gregory Wolber & David McLaughlin (Eastman Kodak Company) for donation of a ToFSpec2E MALDI-MS (each). C.C.D. thanks his advisors, Vlad Artenie, Wolfgang Haehnel, Paul M. Wassarman & Thomas A. Neubert for advice and support. This work was supported in part by the Keep a Breast Foundation (KEABF-375-35054), the Redcay Foundation (SUNY Plattsburgh), the Alexander von Humboldt Foundation, SciFund Challenge, the David A. Walsh fellowship, and by the U.S. Army research office (DURIP grant #W911NF-11-1-0304).

References

1. Darie CC, Shetty V, Spellman DS, Zhang G, Xu C, Cardasis HL, Blais S, Fenyo D, Neubert TA (2008) Blue native PAGE and mass spectrometry analysis of the ephrin stimulation-dependent protein-protein interactions in NG108-EphB2 cells. *Applications of Mass Spectrometry in Life Safety, NATO Science for Peace and Security Series*. Springer, Düsseldorf, Germany
2. Ngounou Wetie AG et al (2012) Investigation of stable and transient protein-protein interactions: past, present, and future. *Proteomics* 13(3–4):538–557
3. Darie CC (2013) Mass spectrometry and its application in life sciences. *Aust J Chem* 66:1–2
4. Ngounou Wetie AG et al (2013) Identification of post-translational modifications by mass spectrometry. *Aust J Chem* 66:734–748
5. Ngounou Wetie AG et al (2014) Protein-protein interactions: switch from classical methods to proteomics and bioinformatics-based approaches. *Cell Mol Life Sci* 71:205–228
6. Petrareanu C et al (2013) Comparative proteomics reveals novel components at the plasma membrane of differentiated HepaRG cells and different distribution in hepatocyte-and biliary-like cells. *PLoS One* 8(8):e71859
7. Sokolowska I et al (2012) Proteomic analysis of plasma membranes isolated from undifferentiated and differentiated HepaRG cells. *Proteome Sci* 10(1):47
8. Sokolowska I et al (2013) Applications of mass spectrometry in proteomics. *Aust J Chem* 66:721–733
9. James P (1997) Protein identification in the post-genome era: the rapid rise of proteomics. *Q Rev Biophys* 30(4):279–331
10. Aebersold R, Mann M (2003) Mass spectrometry-based proteomics. *Nature* 422(6928):198–207
11. Ngounou Wetie AG et al (2013) Automated mass spectrometry-based functional assay for the routine analysis of the secretome. *J Lab Autom* 18(1):19–29
12. Sokolowska I et al (2012) Disulfide proteomics for identification of extracellular or secreted proteins. *Electrophoresis* 33(16):2527–2536
13. Sokolowska I et al (2013) Mass spectrometry investigation of glycosylation on the NXS/T sites in recombinant glycoproteins. *Biochim Biophys Acta* 1834(8):1474–1483
14. Sokolowska I et al (2012) Automatic determination of disulfide bridges in proteins. *J Lab Autom* 17(6):408–416
15. Sokolowska I et al (2011) Mass spectrometry for proteomics-based investigation of oxidative stress and heat shock proteins. In: Andreescu S, Hepel M (eds) *Oxidative stress: diagnostics, prevention, and therapy*. American Chemical Society, Washington, DC

16. Woods AG et al (2012) Potential biomarkers in psychiatry: focus on the cholesterol system. *J Cell Mol Med* 16(6):1184–1195
17. Woods AG et al (2011) Blue native page and mass spectrometry as an approach for the investigation of stable and transient protein-protein interactions. In: Andreescu S, Hepel M (eds) *Oxidative stress: diagnostics, prevention, and therapy*. American Chemical Society, Washington, DC
18. Viswanathan S, Unlu M, Minden JS (2006) Two-dimensional difference gel electrophoresis. *Nat Protoc* 1(3):1351–1358
19. Gygi SP et al (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol* 17(10):994–999
20. Ong SE et al (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics* 1(5):376–386
21. Stemmann O et al (2001) Dual inhibition of sister chromatid separation at metaphase. *Cell* 107(6):715–726
22. Anderson L, Hunter CL (2006) Quantitative mass spectrometric multiple reaction monitoring assays for major plasma proteins. *Mol Cell Proteomics* 5(4):573–588
23. Liu H, Sadygov RG, Yates JR III (2004) A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal Chem* 76(14):4193–4201
24. Savitski MF, Savitski MM (2010) Unbiased detection of posttranslational modifications using mass spectrometry. *Methods Mol Biol* 673:203–210
25. Spiro RG (2002) Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. *Glycobiology* 12(4):43R–56R
26. Marino K et al (2010) A systematic approach to protein glycosylation analysis: a path through the maze. *Nat Chem Biol* 6(10):713–723
27. Tarrant MK, Cole PA (2009) The chemical biology of protein phosphorylation. *Annu Rev Biochem* 78:797–825
28. Gorman JJ, Wallis TP, Pitt JJ (2002) Protein disulfide bond determination by mass spectrometry. *Mass Spectrom Rev* 21(3):183–216
29. McAuley A et al (2008) Contributions of a disulfide bond to the structure, stability, and dimerization of human IgG1 antibody CH3 domain. *Protein Sci* 17(1):95–106
30. Koh GC et al (2012) Analyzing protein-protein interaction networks. *J Proteome Res* 11(4):2014–2031
31. De Las Rivas J, Fontanillo C (2010) Protein-protein interactions essentials: key concepts to building and analyzing interactome networks. *PLoS Comput Biol* 6(6):e1000807
32. Krause F (2006) Detection and analysis of protein-protein interactions in organellar and prokaryotic proteomes by native gel electrophoresis: (membrane) protein complexes and supercomplexes. *Electrophoresis* 27(13):2759–2781
33. Schagger H, von Jagow G (1991) Blue native electrophoresis for isolation of membrane protein complexes in enzymatically active form. *Anal Biochem* 199(2):223–231
34. Anderson LE, McClure WO (1973) An improved scintillation cocktail of high-solubilizing power. *Anal Biochem* 51(1):173–179
35. O'Connell PB, Brady CJ (1976) Polyacrylamide gels with modified cross-linkages. *Anal Biochem* 76(1):63–73
36. Hansen JN (1976) Electrophoresis of ribonucleic acid on a polyacrylamide gel which contains disulfide cross-linkages. *Anal Biochem* 76(1):37–44
37. Bornemann S et al (2010) A novel polyacrylamide gel system for proteomic use offering controllable pore expansion by crosslinker cleavage. *Electrophoresis* 31(4):585–592
38. Wei J, Buriak JM, Siuzdak G (1999) Desorption-ionization mass spectrometry on porous silicon. *Nature* 399(6733):243–246
39. Yates JR, Ruse CI, Nakorchevsky A (2009) Proteomics by mass spectrometry: approaches, advances, and applications. *Annu Rev Biomed Eng* 11:49–79
40. Perkins DN et al (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 20(18):3551–3567

41. 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(11):976–989
42. Tabb DL, McDonald WH, Yates JR III (2002) DTASelect and Contrast: tools for assembling and comparing protein identifications from shotgun proteomics. *J Proteome Res* 1(1):21–26
43. Craig R, Beavis RC (2003) A method for reducing the time required to match protein sequences with tandem mass spectra. *Rapid Commun Mass Spectrom* 17(20):2310–2316
44. Pedrioli PG et al (2004) A common open representation of mass spectrometry data and its application to proteomics research. *Nat Biotechnol* 22(11):1459–1466
45. Deutsch EW, Lam H, Aebersold R (2008) Data analysis and bioinformatics tools for tandem mass spectrometry in proteomics. *Physiol Genomics* 33(1):18–25
46. Deutsch E (2008) mzML: a single, unifying data format for mass spectrometer output. *Proteomics* 8(14):2776–2777
47. Lenz EM et al (2005) Metabonomics with ^1H NMR spectroscopy and liquid chromatography-mass spectrometry applied to the investigation of metabolic changes caused by gentamicin-induced nephrotoxicity in the rat. *Biomarkers* 10(2–3):173–187
48. Kussmann M, Rezzi S, Daniel H (2008) Profiling techniques in nutrition and health research. *Curr Opin Biotechnol* 19(2):83–99
49. Lescuyer P, Hochstrasser D, Rabilloud T (2007) How shall we use the proteomics toolbox for biomarker discovery? *J Proteome Res* 6(9):3371–3376
50. Surinova S et al (2011) On the development of plasma protein biomarkers. *J Proteome Res* 10(1):5–16
51. Drake RR et al (2011) Challenges to developing proteomic-based breast cancer diagnostics. *OMICS* 15(5):251–259
52. Woods AG, Sokolowska I, Darie CC (2012) Identification of consistent alkylation of cysteine-less peptides in a proteomics experiment. *Biochem Biophys Res Commun* 419(2):305–308