
Mass Spectrometry-Based Proteomics: Basic Principles and Emerging Technologies and Directions

1

Susan K. Van Riper, Ebbing P. de Jong, John V. Carlis,
and Timothy J. Griffin

Abstract

As the main catalytic and structural molecules within living systems, proteins are the most likely biomolecules to be affected by radiation exposure. Proteomics, the comprehensive characterization of proteins within complex biological samples, is therefore a research approach ideally suited to assess the effects of radiation exposure on cells and tissues. For comprehensive characterization of proteomes, an analytical platform capable of quantifying protein abundance, identifying post-translation modifications and revealing members of protein complexes on a system-wide level is necessary. Mass spectrometry (MS), coupled with technologies for sample fractionation and automated data analysis, provides such a versatile and powerful platform. In this chapter we offer a view on the current state of MS-proteomics, and focus on emerging technologies within three areas: (1) New instrumental methods; (2) New computational methods for peptide identification; and (3) Label-free quantification. These emerging technologies should be valuable for researchers seeking to better understand biological effects of radiation on living systems.

Keywords

Proteome • Mass spectrometry • Matrix-assisted laser desorption/ionization • MALDI • Electrospray ionization • ESI • Nanoscale reversed-phase liquid chromatography • NanoLC • Sequence database • SEQUEST • Mascot • 2-dimensional gel electrophoresis • Phosphorylation • Glycosylation • Isotope labeling • Peptide sequencing • Peptide identification

S.K. Van Riper (✉)
Department of Biomedical Informatics and
Computational Biology, University of Minnesota,
321 Church St SE/6-155 Jackson Hall, Minneapolis,
MN 55455, USA
e-mail: vanr0014@umn.edu

E.P. de Jong
Department of Biochemistry, Molecular Biology and
Biophysics, University of Minnesota, 321 Church St
SE/6-155 Jackson Hall, Minneapolis, MN 55455, USA
e-mail: dejon039@umn.edu

Abbreviations

| | |
|-----------------|--|
| 2DGE | Two-dimensional gel electrophoresis |
| APEX | Absolute protein expression |
| AUC | Area-under-curve |
| CAD | Collision activated dissociation |
| CID | Collision induced dissociation |
| ECD | Electron capture dissociation |
| ESI | Electrospray ionization |
| ETD | Electron transfer dissociation |
| FAIMS | Field-asymmetry ion mobility spectrometry |
| FDR | False discovery rate |
| HCD | High-energy collision dissociation |
| HUPO | Human proteome organization |
| ICAT | Isotope coded affinity tags |
| IMS | Ion mobility spectrometry |
| IRMPD | Infrared multiphoton dissociation |
| iTRAQ | Isotope tagging for relative and absolute quantification |
| LC | Liquid chromatography |
| m/z | Mass-to-charge |
| MALDI | Matrix-assisted laser desorption/ionization |
| MRM | Multiple reaction monitoring |
| MS | Mass spectrometry |
| MS ² | Tandem mass spectrometry |
| NIST | National institute of standards and testing |
| NSAF | Normalized spectral abundance factor |

| | |
|-------|--|
| PAI | Protein abundance index |
| PQD | Pulsed Q dissociation |
| PTM | Post-translational modification |
| SILAC | Stable isotope labeling of amino acids in cell culture |
| SRM | Selected reaction monitoring |
| TMT | Tandem mass tags |
| Xcorr | Correlation score |

1.1 Introduction

Genome sequencing efforts initiated in the 1980s fostered a new paradigm in biological research: the system-wide characterization of biomolecules. Within this new paradigm, the field of proteomics, which seeks to characterize proteins on a system-wide level, emerged. Proteins, the major catalytic and structural components within all living systems, are arguably the most informative biomolecules for understanding cellular function and response to systematic perturbations, such as radiation exposure. Unfortunately, proteins are also the most challenging of all biomolecules to study on a system-wide level. In addition to cataloging and quantifying proteins within a complex biological sample, information on their post-translational modification (PTM) state, subcellular localization and interactions with other biomolecules is necessary for full proteome characterization. Adding to the challenge, proteins are dynamic, changing their abundance, PTM state, localization and interactions in response to stimuli. Gene sequences or even mRNA expression levels cannot reveal or predict this protein-level information [1, 2]. Therefore technologies for direct analysis of proteins are necessary for proteome characterization.

Although no single technology can fully characterize all aspects of proteomes, mass spectrometry (MS) is the most powerful and flexible for proteomic analysis. The revolutionary discoveries in the late 1980s of Matrix-Assisted Laser Desorption/Ionization (MALDI) [3] and Electrospray Ionization (ESI) [4] made possible analysis of intact polypeptides and proteins by MS. Along with these ionization methods, three

J.V. Carlis

Department of Biomedical Informatics and Computational Biology, University of Minnesota, 321 Church St SE/6-155 Jackson Hall, Minneapolis, MN 55455, USA

Department of Computer Science and Engineering, University of Minnesota, 321 Church St SE/6-155 Jackson Hall, Minneapolis, MN 55455, USA
e-mail: carlis@umn.edu

T.J. Griffin

Department of Biomedical Informatics and Computational Biology, University of Minnesota, 321 Church St SE/6-155 Jackson Hall, Minneapolis, MN 55455, USA

Department of Biochemistry, Molecular Biology and Biophysics, University of Minnesota, 321 Church St SE/6-155 Jackson Hall, Minneapolis, MN 55455, USA
e-mail: tgriffin@umn.edu

technologies combined to provide an analytical platform underpinning the field of MS-based proteomics and enabling system-wide protein analysis. First, nanoscale reversed-phase liquid chromatography (nanoLC) coupled online with MS instruments came about for separating peptide digests from complex protein mixtures [5]. Second, tandem mass spectrometry, commonly referred to as MS/MS, arose for predictably fragmenting peptides, necessary for determining their amino acid sequence [6]. Tandem mass spectrometry initially scans all mass-to-charge (m/z) values of peptide ions as they elute from the nanoLC column, and records their signal intensities in an MS¹ spectrum. Detected peptide ions are then isolated, and fragmented, with the instrument undertaking another scan of all m/z values of fragment ions, recording their signal intensities in an MS² spectrum. Third, automated sequence database searching, led by the program SEQUEST [7] and followed by Mascot [8], was developed to match large amounts of MS² spectra to peptide sequences contained in databases, and in turn infer protein identities present within complex mixtures.

This basic platform for what has been termed “shotgun” or “bottom-up” proteomics, offered researchers a new way forward for identifying proteins within complex mixtures. However, two problems, the extreme chemical heterogeneity and large dynamic range of protein abundance within protein mixtures derived from cells, tissues or bodily fluids, required new methods for more sensitive identification of proteins. Multidimensional liquid chromatography-based methods for fractionating peptide digests upstream of MS analysis, helped to, at least in part, address these problems, by simplifying complex mixtures and minimizing signal suppression within the MS instrument [9–11]. These fractionation methods also overcame the limitations [12] of traditionally used two-dimensional gel electrophoresis (2DGE) for separating complex protein mixtures. Methods for enriching PTMs prior to MS analysis improved identification of proteins carrying important modifications, such

as phosphorylation [13–15] or glycosylation [16], on a large-scale. Stable isotope labeling and dilution, traditionally used in mass spectrometry analysis of small molecules, was adapted for quantitative measurements of proteins analyzed by MS [17].

Collectively, these components of the MS-based proteomics “toolbox” fostered a new and powerful means to study proteins on a system-wide level. This enhanced platform can now routinely identify and quantify thousands of proteins, including those carrying PTMs in complex protein mixtures. Because proteins are the ubiquitous molecular “effectors” within any organism, MS-based proteomics applies to all fields of biological research, including the effects of radiation on the cellular environment.

MS-based proteomics has always been and remains a collection of dynamic technologies, with new ones constantly emerging across all facets of the platform. Continuous improvements in technologies have moved the proteomics field closer to its ambitious goal to fully characterize proteins within complex biological samples with high throughput. Examples of such technologies include: improvements in MS instrument sensitivity increases identification of low-abundance proteins; more sophisticated software programs for peptide identification from MS² data enables detecting a higher proportion of the hundreds of known PTMs [18] of proteins; higher throughput and more quantitatively accurate methods makes possible quantification of protein targets of interest in a large number of individual samples, which is especially important for biomarker studies. However, despite continued technological improvements, the sheer complexity of biological systems greatly challenges the current platform in meeting the goal of full proteome characterization. To illustrate using a rough estimation, the human genome contains about ~25,000 genes that are processed by a variety of regulated steps (mRNA splicing, proteolysis, etc.) to produce ~250,000 distinct proteins. These are in turn covalently modified via phosphorylation, acetylation, ubiquitination, oxidation, sumoylation, etc., to generate a

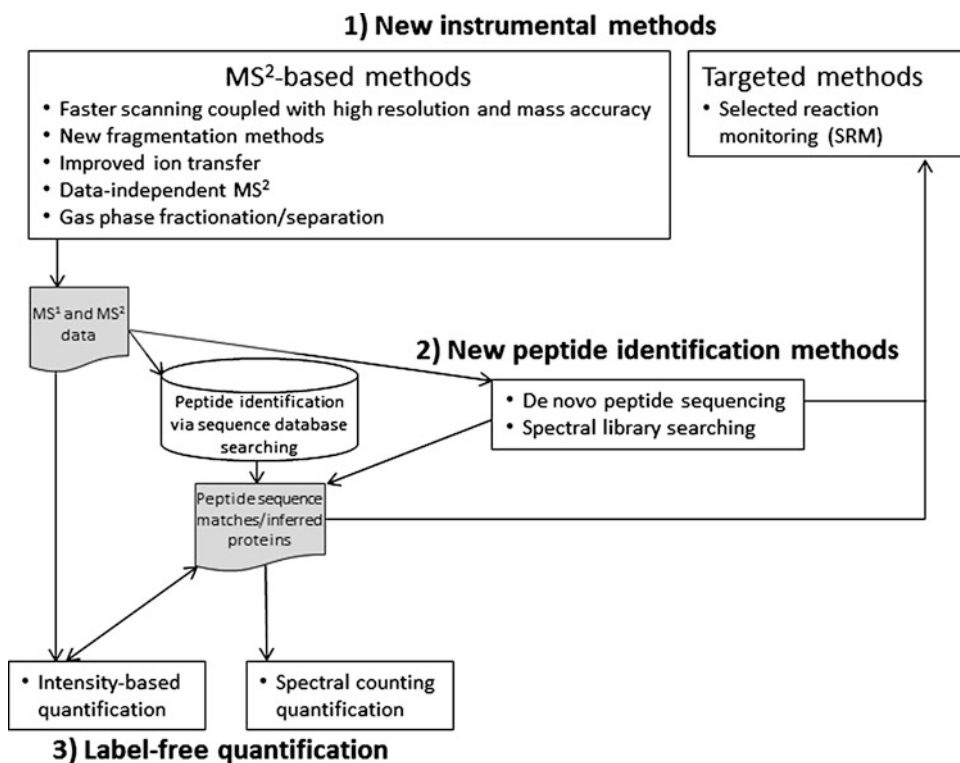


Fig. 1.1 Overview of the three, interconnected technologies reviewed in this chapter

proteome with millions of distinct protein-based molecules. The current proteomics technologies can still only reliably detect a fraction of these molecules. Thus there is continued need for new and improved technologies.

Here, we provide our view on three emerging technologies in MS-based proteomics that are pushing the field in new directions: (1) New instrumental methods; (2) New computational methods for peptide identification; and (3) Label-free quantification. Figure 1.1 provides an overview of the interconnectivity of these technologies. The data produced using new instrumental methods, in particular high resolution and mass accuracy data, enables improved de novo peptide identification, which seeks to overcome the inherent limitations of the currently practiced sequence database searching. Label-free quantification provides a flexible and simple way when comparing samples to determine differentially abundant peptides and inferred proteins. We review recent advances in these three technologies.

1.2 New Instrumental Methods

From the outset, MS instrumentation has been the core technology driving proteomic advances. Fortunately, impressive improvements to the technology have continuously emerged over the last two decades. Most instrument vendors introduce a new model of any given MS instrument every 2–3 years, and those manufactured ~10 years prior to the latest model can scarcely be considered suitable for research. Some of the most fundamental and sought-after metrics for mass spectrometers are resolution, scanning speed, and sensitivity. These are strongly related: in mass spectrometers sensitivity comes at the cost of scanning speed which, in turn, comes at the cost of resolution. Here we review some emerging MS instruments that are redefining what is possible in MS-based proteomic studies. We also discuss emerging methods that are closely linked to improving the performance of the MS instrumentation used for proteomic studies.

1.2.1 Higher Mass Accuracy and Faster Scanning Instruments

Bottom-up proteomics uses nanoLC for peptide separation coupled directly with the MS. Peptides eluting from the nanoLC column are ionized via ESI, and introduced into the mass spectrometer. For complex mixtures (e.g., cell or tissue lysates), the number of peptides vastly exceeds the peak capacity of the separations typically used. Michalski et al. have determined that during a typical 90-min gradient LC run of a complex proteomic mixture, a state-of-the-art mass spectrometer can detect over 100,000 peptide species [19]. Consequently, there are many peptide ions being introduced to the mass spectrometer simultaneously and high resolution is required to differentiate these molecules by their m/z values. Resolution is defined as the ratio of the m/z value to the width of the peak at half its maximum. Therefore a large ratio, for example 50,000/1, is desirable. While not directly related, high mass accuracy usually accompanies high resolution. Mass accuracy is calculated via the following equation: $[(\text{actual } m/z - \text{observed } m/z) / \text{actual } m/z]$. Because this ratio is usually very small, it is multiplied by 10^6 , and reported in units of parts-per-million (ppm). Values of 5 ppm or less are desirable for mass accuracy. Ideally, a mass spectrometer provides sufficient mass accuracy to assign a unique elemental composition, and thus an estimation of amino acid composition, to all peptide peaks in the scanned range. Such a high level of mass accuracy greatly constrains the number of possible amino acid sequences responsible for an observed signal and reduces the incidence of false discoveries when assigning amino acid composition [20]. With 1 ppm measured mass accuracy, the amino acid composition of relatively small peptides with molecular weights in the range of 700–800 Da can be determined [21]. With the help of internal calibration techniques, achieving this level of accuracy now is almost routine [22, 23].

Tandem mass spectrometry is the underlying instrumental analysis method for MS-based proteomics. In its traditional implementation,

detected peptide ions eluting from the nanoLC column are isolated and fragmented, with the m/z values of the fragments being recorded in an MS^2 spectrum. There are numerous ways in which isolated peptides can be fragmented, as will be discussed in Sect. 1.4. The most-used method, collision-induced dissociation (CID), leaks or “bleeds” a small quantity of an inert gas (He, N_2 , Ar) into the chamber where the isolated peptide ions reside. The peptide ions collide with the gas and internalize the energy from the collision. Being in the gas phase, the peptide ions cannot re-distribute the energy to solvent molecules. Instead, the energy is eventually transferred to a vibrational mode which cannot sustain the energy available and results in bond cleavage. This primarily results in cleavage along the peptide bond of the peptide backbone, although one also frequently sees the loss of water, ammonia, carbon monoxide or labile post-translational modifications [24]. The predominant fragments are named accordingly: b-ions are fragments derived from the N-terminus of the peptide, while y-ions are fragments derived from the C-terminus (see Fig. 1.3 in Sect. 1.2). Certain high-energy fragmentation techniques fragment or completely lose the amino acid side chains, and such ions are named d, v, and w- ions.

Ideally, one would acquire a high-quality MS^2 spectrum for each peptide within a complex mixture. Unfortunately this is not the reality, due to two main factors. First, the speed at which an instrument can gather a sufficient population of peptide ions and generate an MS^2 spectrum will determine its effectiveness at sequencing all the detected peptide ions in a sample. Since the peptide signal from the LC column is transient, the more time spent scanning m/z fragments from any peptide ion selected for fragmentation, the more signals from other peptides will be missed. Thus instruments that quickly scan and record MS^2 spectra are desirable. A second factor is the dynamic range of abundance of the peptides present. The electrospray process can generate only a finite amount of ions per unit time, and when an extremely abundant peptide elutes from the column, less abundant peptides will undergo so-called ion suppression.

Instruments with a greater dynamic range or efficiency at selecting low-abundance ions can mitigate these effects; however peptides from the lowest-abundance proteins in a sample remain undetectable unless enrichment or targeted strategies are employed. Thus, instruments with increased sensitivity to low-abundance peptides are desirable. Increased sensitivity is also linked to scan speed, as increased sensitivity means the instrument must spend less time accumulating fragment ions, and can record MS² spectra more rapidly.

Some recently released instruments, combining the desirable qualities of high resolution and mass accuracy and rapid scanning speed, are the Thermo Orbitrap series and the AB Sciex Triple TOF 5600. The Orbitrap mass analyzer allows ions to orbit a central electrode while simultaneously oscillating axially. This axial motion is mass (–to charge) dependent. The Orbitrap analyzer collects an image current of all ions present, each with a characteristic axial frequency. Fourier transform of this image current yields the orbital frequencies present and thus, the *m/z* values present. In this type of mass analyzer, resolution increases with longer scans [25, 26]. The first commercial Orbitrap instruments, coupled with a linear trapping quadrupole, delivered resolving powers of >100,000 with measured mass accuracy of 2–5 ppm and recording of up to three low-resolution MS² scans per second [27]. Recent introduction of the Orbitrap Velos, led to 10 low-resolution MS² spectra recorded per second [28]. The latest installment of the Orbitrap series, Orbitrap Elite, employs a more powerful Orbitrap mass analyzer, [29] providing 2–3 fold higher resolution of up to 240,000, and an improved Fourier transform algorithm, delivering a further 2.3-fold greater resolution. No publications using this instrument exist at the time of writing, however a recent publication describes a related instrument. The Q Exactive, employing the same Orbitrap as the Elite but with a detectorless trapping quadrupole, requires that all MS¹ and MS² scans be performed in the Orbitrap, thus giving high mass accuracy in all mass spectra and allowing stricter filtering criteria when performing database searches for assignment of

peptides to MS² spectra. This instrument also records 10 high mass accuracy MS² spectra in a ~1 s cycle that includes an initial MS¹ scan [30]. When coupled to an ultrahigh pressure LC system delivering a 4-h gradient, the Q Exactive achieves 92% coverage of the yeast proteome.

The AB Sciex Triple TOF 5600 is in fact a quadrupole time-of-flight (Q-TOF) configuration. Relative to other Q-TOF instruments however, the 5600 has improved ion sampling, rapid pulsing of ions towards the TOF and high TOF acceleration voltages, all of which allow up to 100 MS² recorded per second [31]. One of the first publications using this instrument in a proteomics setting determined that 20 MS² scans per 1.3 s cycle gave the most peptide assignments to acquired MS² spectra. This instrument delivers a resolution of 40,000, and, with internal calibration, also produced a measured mass accuracy of 2 ppm. The extremely fast scanning of this instrument is credited for the threefold increase in peptide identifications over an early-model Orbitrap instrument.

1.2.2 Improved Electrospray Ion Transfer Efficiency

Maximized capture in the mass spectrometer of peptide ions generated via ESI increases the instrument's sensitivity. The ESI process generates a divergent ion beam which is collected by a conductance-limited aperture, typically in the form of a skimmer. This configuration captures only a fraction of the ions generated by ESI. The efficiency of ion transfer from an ESI source to the detector has been estimated at <0.1%. To address this bottleneck, Smith and colleagues have produced many refinements to the long-known stacked ring ion guide [32] or ion funnel (Fig. 1.2), yielding successively improved ion transmission while minimizing the *m/z* dependency of ion transmission [33–36]. The ion funnel consists of a series of evenly or progressively further-spaced ring electrodes with successively decreasing inner diameters to help focus the divergent ion beam. Radio frequency [32] or static [37] electric fields are used to

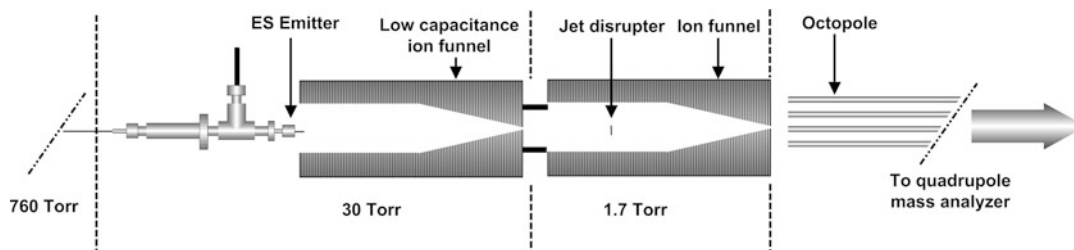


Fig. 1.2 Ion funnel schematic (Adapted with permission [39], Copyright 2008 American Chemical Society)

drive ions through the device sometimes with a DC field superimposed [36]. The ion funnel has achieved collection efficiencies of 50–60% across a typical proteomics m/z range of 200–2,000. However this interface is still not efficiently coupled to a mass spectrometer due to an increased concentration of charged droplets whose repulsion causes losses during transmission [38].

Ion funnels have become widely adopted in commercial mass spectrometers, however the issue of ion transmission remains a barrier to efficient use of all ions generated by ESI. Another factor limiting ion transmission is the pressure gradient between the ESI source and the mass spectrometer. ESI is usually operated at ambient pressure, while the mass spectrometer is operated under high vacuum. This pressure gradient makes efficient ion capture difficult. Operating the electrospray within the low-vacuum region of MS has been used to mitigate this problem and improves ion signal by approximately an order of magnitude [40, 41], giving an estimated 50% ion transmission efficiency [42]. This technology has been termed subambient pressure ionization with nanoelectrospray, or SPIN. Potential, albeit minor hurdles to the widespread adoption of this technique are its compatibility with typical nano-LC flow rates [41] and the robustness of an interface where users are required to introduce a nano-LC-ESI column into the first vacuum stage of a mass spectrometer. However, more efficient use of ions produced by ESI clearly pays dividends in improving instrument sensitivity and will likely continue to see innovations until a majority of ions can be routinely captured in commercial-grade mass spectrometers.

1.2.3 New Fragmentation Methods

Tandem mass spectrometry for amino acid sequence elucidation relies, in part, on the ability to efficiently fragment peptide ions¹. CID is by far the most used fragmentation method due to its simplicity, ease of implementation, and ability to fragment all peptides at least moderately well, in spite of the wide chemical diversity of a typical proteomic sample. As CID is well-suited to other classes of molecules beside peptides, it is present in virtually all commercially available tandem mass spectrometers [43]. CID is sometimes referred to as collisionally-activated dissociation (CAD), particularly when applied to the beam-type version of this fragmentation. This distinction causes one to view CID as a method limited to resonant excitation in an ion trap. Adding to the confusion, one instrument vendor refers to their CAD cell as Higher energy Collisional Dissociation (HCD) [44]. These distinctions are not merely semantic, however. Aside from the differences in hardware required to perform them, resonant excitation CID is slower (~ 30 ms vs. < 1 ms), produces different fragment ion intensities than beam-type CAD, and CID suffers from the so-called “one-third rule”. Under the necessary ion activation conditions for sufficient fragmentation in an ion trap, the resulting fragment ions with $m/z \leq 0.3$ times that of the precursor ion are lost during the activation [45]. This loss of low-mass b/y ions as well as helpful immonium ions hinders the interpretation of the resulting tandem mass spectrum.

¹The terms “fragmentation” and “dissociation” are used interchangeably in the field.

In isotope-tagging experiments for relative quantification between samples, the low-mass region of the MS² spectra contain the reporter fragment ions, crucial to obtaining quantitative information [46, 47]. While a modified version of resonant excitation in an ion trap, called PQD, can be implemented to preserve the low-mass ions, [48] the low fragmentation efficiency of PQD has invited comparison between PQD and HCD [49]. Instrumental improvements in the HCD cell [50] now make HCD a very attractive method for the analysis of low-mass reporter ions in quantitative mass spectrometry.

The use of HCD fragmentation has also been examined for studies of phosphopeptides. While its use resulted in more phosphopeptide and phosphosite identifications than when using CID, [51] it seems that this improvement can be attributed largely to the high mass accuracy scans which are mandated following HCD, as opposed to low-resolution and low mass accuracy scans typically used following CID, rather than any inherent improvement in fragmentation pattern or ion collection.

Electron capture dissociation (ECD) [52, 53] and electron transfer dissociation (ETD) [54–56] are two related fragmentation methods, used in Ion Cyclotron Resonance (ICR) and ion trap mass spectrometers, respectively. Both methods involve an ion/ion interaction between a multiply protonated peptide cation and either a low-energy electron in ECD or an electron-donating anion radical molecule in ETD. The charge-reduced peptide cation dissociates before any energy randomization can occur. This is especially important for peptides carrying PTMs such as phosphorylation or glycosylation [57]. In CID or CAD the labile covalent bonds between these modifications and the peptide are usually preferentially fragmented, limiting fragmentation across the peptide backbone and resulting in less informative MS² spectra for peptide sequence assignment. ECD or ETD meanwhile provide richer MS² spectra from many PTM carrying peptides since much of the fragmentation still occurs along the peptide backbone. This leaves intact the modified amino acid residues and also provides a relatively full complement of sequence-rich fragments, enabling more effective

sequence assignment and increased confidence in the site of modification. This property has made ECD/ETD especially useful in studies of phosphoproteins and glycoproteins [58–61].

Photodissociation methods can also be used to obtain peptide sequence information. Two spectral regimes are commonly being used for this purpose: infrared and (vacuum) ultraviolet. Infrared multiphoton dissociation (IRMPD) typically uses a CO₂ laser emitting tens of watts at a wavelength of 10.6 μm. This wavelength is efficiently absorbed by phosphopeptides and thus IRMPD has been investigated for its utility in analyzing this important post-translational modification [62, 63]. MS² spectra following IRMPD do not suffer from the one-third rule, [64] however the fragmentation typically takes twice as long as resonant excitation CID. IRMPD produces b/y ions, but also yields more internal fragment ions than CID [65].

UV photodissociation typically uses excimer lasers emitting at 157 or 193 nm as the light source [66]. As air absorbs these wavelengths efficiently, the 157 nm light source especially must be placed in the vacuum region of the mass spectrometer, complicating the instrumental requirements. Single-photon UV absorption is sufficient to induce dissociation and in contrast to IRMPD, irradiation times on the order of μs or ns are sufficient. While both 157 and 193 nm light target the peptide backbone, UV photodissociation produces a range of fragments in addition to b/y ions such as a, d, x, v and w ions. The presence of d, v and w fragment ions is evidence of a high-energy fragmentation method; not surprising given that the energy of a single UV photon is approximately double that of a peptide bond [67]. While these fragments can be useful, for instance, in differentiating between leucine and isoleucine, most commercially available peptide identification programs are not optimized for, or capable of, analyzing these ions.

While there currently exists an impressive, if not overwhelming, array of dissociation methods, none can meet all the requirements of every conceivable experiment. Until such a method exists, there remains room for improvement to those currently used, and the development of entirely novel ones.

1.2.4 Data-Independent MS² Analysis

Most tandem MS experiments are performed in a data-dependent manner: the collection of peptide ions entering the mass spectrometer are first recorded in the MS¹ spectrum, and these ions (also called precursor ions) are serially selected for fragmentation and MS² spectra acquisition [68]. It is well-established however, that this method does not provide complete selection of all peptides in complex samples. An alternative method is to perform fragmentation at all peptide ion *m/z* values, regardless of which ions can be detected in an MS¹ spectrum. This method is embodied by two different approaches: with and without isolation of precursor ions within a defined *m/z* window.

In data-dependent MS² spectra acquisition, all ions within a defined, relatively narrow *m/z* window bracketing a precursor of interest are isolated for fragmentation. It however is possible to omit precursor *m/z* isolation and effectively fragment simultaneously all ions present across the entire *m/z* range scanned when acquiring MS¹ spectra. When precursor isolation is omitted, a single LC-MS run could in theory detect the entire proteome. In practice, the usual list of mass spectrometer capabilities is desired: high mass accuracy is tremendously beneficial in assigning fragment ions to precursor ions [69, 70]. High scan speed and MS² spectra acquisition is beneficial in assigning fragment ions to a precursor, based on chromatographic retention time [71, 72]. Dynamic range of the mass spectrometer is also important in achieving deep sequencing, due to the occurrence of co-eluting peaks [73]. One advantage of this approach is that it can be performed on relatively simple instrumentation: only a collision cell (or other means of achieving dissociation [63]) and a single-stage mass analyzer are required. One major challenge in this type of experiment is the data analysis. Knowledge of which precursor ion masses give rise to the observed fragments is necessary for assigning peptide sequence to MS²

spectra using sequence database searching software. When simultaneously fragmenting multiple peptide ions across a large *m/z* range, knowledge of which precursor ion belongs to which fragments is lost. While the precursor mass belonging to sets of fragments can be inferred by relating its retention time in an MS¹ scan to that of the fragments ions in an MS² scan, [71] this is not a trivial process [72]. As such, data-independent MS² in the absence of precursor isolation still struggles with very complex samples, but this approach seems to be re-evaluated each time a breakthrough in hardware performance is made.

Recently, another data-independent acquisition approach was investigated by rapid isolation and fragmentation of peptide ions within narrow (2.5 *m/z*) precursor isolation windows, spanning the entire *m/z* range covered by peptide ions (~400–1,400). These narrow *m/z* “bins” mitigated the need for MS¹ scans, while still providing a tight mass range of potential precursor *m/z* that could be connected to each MS² spectra for sequence assignment. For thorough analysis of a typical, complex protein digest, this approach required over 4 days of mass spectrometry instrument time, but required no sample pre-fractionation [74]. Wider isolation widths have been tested, but the resulting tandem mass spectra are likely to contain more than a single peptide species, resulting in complicated database searches [75]. The use of narrow isolation widths demonstrated the ability for a highly automated method to achieve greater proteome coverage and a wider dynamic range than a data-dependent method. As with experiments that do not use precursor isolation, such studies using narrow isolation widths benefit from instrumental improvements such as high mass accuracy and resolution [76]. It is somewhat surprising how few publications exist on this topic, as it seems well-suited to those experimenters not well versed in multidimensional peptide fractionations who might be attracted to a highly automated method. At this time it is difficult to predict whether the data-independent approach will flourish or flounder, in spite of its demonstrated potential.

1.2.5 Gas-Phase Fractionation and Ion Mobility Separations

Since the complexity of a typical proteomics sample can easily exceed the capacity of a LC-MS system to resolve and detect all peptides present, most fractionation schemes [9–11] occur upstream of the mass spectrometer, and are designed to simplify the mixtures introduced into the mass spectrometer to achieve better sensitivity. However, peptide fractionations usually require considerable manual labor and sample handling.

In contrast to upstream fractionation, a fractionation method has been devised wherein repeated injections of the same unfractionated sample are introduced to the LC-MS, but for each injection a different “fraction” of the standard m/z range is analyzed (e.g. 400–575, 560–740, 730–910 and 900–1,795). This allows the instrument to focus on a smaller m/z range to achieve the most comprehensive detection and fragmentation of peptide ions in this range as possible [77–79]. Since the instrument analyzes or ignores certain portions of the ionized m/z range, this method has been termed “gas-phase fractionation”. For a yeast cell lysate, the analysis of three gas-phase fractions was compared to triplicate analyses of the entire mass range, and found to increase the number of identifications by 30% [80]. A further refinement of this method used *in silico* calculations to determine the optimal m/z bins which would yield equal numbers of theoretical tryptic fragments across the number of bins selected [81]. The authors studied three different organisms of differing complexity, and found that regardless of the biological source, roughly half the tryptic peptides reside below m/z 685 with decreasing ion density as m/z increased. Thus gas-phase fractionation certainly has the power to increase proteomic coverage, but at the cost of performing multiple LC-MS runs. Unlike upstream peptide fractionation methods, gas-phase fractionation does this in an entirely automated fashion, reducing labor and sample handling. However, this method might not be suitable to the analysis of very small samples with low protein amounts where multiple LC-MS analyses are not possible.

Ion mobility spectrometry (IMS) is a gas-phase separation method for electrophoretically separating ions in the presence of a buffer gas. Ions are separated by their mass, charge and mobility; the latter being inversely related to their collisional cross section [82]. IMS devices are frequently coupled to a mass spectrometer (using ion funnels), creating a hyphenated method, IMS-MS. For the purposes of this section it will be assumed that all IMS separations are coupled a mass spectrometer. The time frame of a typical IMS separation is ideally suited to its incorporation in a multidimensional fractionation scheme in proteomics: the peak widths for LC, IMS and TOF-MS are on the order of seconds, ms and μ s, respectively. This allows each subsequent method to acquire tens of measurements of the preceding separation—the minimum required for adequate profiling of a peak [83].

Three versions of IMS are used: linear drift tubes, traveling wave ion guides, and field-asymmetry IMS (FAIMS) [69]. Linear drift tubes and traveling wave ion guides both resemble a stacked ring ion guide (see Sect. 1.3), though differ in the way electric fields are applied in order to propel ions through the device. These differences affect the separation mechanism. The resolution of linear drift tubes and traveling wave ion guides [84] is typically the greatest at 100–150, however similar values have recently been reported with FAIMS [85, 86]. Also, FAIMS typically separates isomers and isobars better than linear drift tubes and hence has been the most widely implemented in proteomics experiments [85–87]. To date, the most successful configuration for FAIMS is the use of parallel plates separated by ~ 2 mm [88]. Under high electric fields, the absolute mobility of an ion deviates from its value at low fields. This difference is exploited in FAIMS by applying an asymmetric radio frequency potential between the two plates. As this potential ejects all ions radially from the device, a DC compensation voltage is required to transmit any ions. This compensation voltage is the discriminating variable in a FAIMS separation [87]. Both Thermo Scientific and AB Sciex have commercially-available FAIMS devices which can be added to their mass spectrometers,

boasting claims of improved selectivity and signal-to-noise ratios. Shvartsburg, Smith and co-workers have made great improvements in the instrumental design of FAIMS devices, improving resolving power [87, 88] and resolving phosphopeptide isomers which differ only in the site of phosphorylation [89]. Waters Corporation has investigated and commercialized a traveling wave ion guide with its mass spectrometers. As the name implies, a DC voltage is passed along the successive ion guide rings, propelling the ions through the device while an rf-field is generated to maintain ions' radial position. By selecting the amplitude and velocity of the DC wave, ions can be separated by mobility or simply transmitted through the device [90, 91].

Ion mobility separations have the ability to add a second dimension of online fractionation to an LC-MS analysis which should greatly simplify the mixture of ions arriving at the mass spectrometer with no increase in analysis time. The resolving power is sufficient to separate different components in a mixture, however a single peptide sequence may have multiple conformations each with different IMS mobility, which de-focuses the ion packed generated by LC-MS. Also, it is not clear whether current computational methods can analyze IMS separations quickly enough to make on-the-fly decisions, as is currently performed in data-dependent LC-MS experiments. Nonetheless, it seems probable that these hurdles can be overcome and that IMS separations will greatly increase the power of proteomics experiments.

1.2.6 Targeted MS

As the collection of known, MS-observable, proteolytically-derived peptides becomes saturated, some researchers are turning away from data-dependent MS analyses. For a known sample type (e.g. identity of the organism, biological state, sample preparation parameters), the observable peptides emanating from its proteome can be predicted and have likely already been observed in other experiments. Thus, generating a comprehensive list of such so-called "proteotypic" peptides should provide a

basis for performing targeted MS experiments in a hypothesis-driven manner [92]. Such methods can then be used, for example, to validate potential biomarkers generated from an initial screening experiment [93] or to follow the proteins in a metabolic pathway following some perturbation [94]. This approach has been greatly advanced by the groups of Aebersold and Carr, who have developed software to predict the most detectable peptides in a mixture, [95, 96] catalogued all experimentally observed peptides [97], demonstrated single copy per cell sensitivity, [94] and are in the process of synthesizing a complete proteotypic peptide library of human serum [98].

A powerful MS method for quantifying several peptides simultaneously is termed selected reaction monitoring (SRM), or sometimes multiple reaction monitoring (MRM). This type of experiment is performed with a triple-quadrupole instrument, and is notoriously selective and sensitive. As ions are electrosprayed into the MS, the first quadrupole transmits a peptide ion at a user-specified m/z value. This ion is then fragmented in the second quadrupole which is not mass-selective, but merely a fragmentation cell. The third quadrupole is then set to transmit the m/z of an expected fragment ion from the precursor peptide ion. This process is repeated, usually for at least three fragments per peptide ion and two proteotypic peptides per protein of interest. Modern mass spectrometers can achieve reliable quantification by dwelling on such a peptide/fragment m/z pair (the "reaction" in SRM, also called a transition) for 10 ms or less. The duty cycle, and thus the sensitivity is inversely related to the number of transitions being monitored, however when the retention time of a peptide is known, the instrument can be scheduled to monitor distinct peptide ions and their transitions at different times. In this way, Kiyonami et al. have quantified 6,000 transitions, relating to 757 peptides in a single LC-MS analysis [99]. They note that this can be extended to 10,000 transitions, targeting 1,000 peptides. Addona et al. [100] have shown that, when using isotopically-labeled standards, this method is very reproducible within and across eight laboratories using

two instrument platforms. Many groups believe that SRM-based targeted proteomics will be the basis for future biomarker validation [101–104].

An important aspect of such large-scale hypothesis-driven efforts is the software. The identification of proteotypic peptides and their SRM transitions can be very time-consuming if performed manually. A variety of software products exist from the instrument manufacturers and from academic groups to assist in the design of SRM experiments [105–107]. The most powerful and popular tool has come from the MacCoss laboratory. Their open-source platform, Skyline, can guide SRM experiments by optimizing collision energy and fragment ion selection, performing quantification, predicting peptide retention time and a host of other functions, for data acquired from the major instrument manufacturers [108–110]. Continued refinements to such software packages will greatly automate and thus expedite the process of developing and optimizing SRM assays capable of quantifying hundreds to thousands of peptides in a single MS analysis. These advancements are transforming MS-based proteomics from just a large-scale discovery technology to a high-throughput assay for monitoring proteins of interest in hypothesis-driven studies.

1.3 New Peptide Identification Methods

1.3.1 Principles of MS² Fragmentation

Tandem mass spectrometry-based proteomics experiments rely on the same principle as Edman degradation, a long standing chemical technique for peptide sequencing [111]. In Edman degradation stepwise degradation from the peptide's n-terminus followed by chromatographic analysis of the released derivatives determines the amino acid sequence. The fragmentation that occurs during the MS² stage mimics Edman degradation because MS² dissociation randomly breaks along the backbones between amino acid residues. This results in two, rarely more, fragment ions,

one each containing the n-terminus and the c-terminus. The m/z values of fragment ions are recorded in the MS² spectra for every selected precursor peptide ion. However, individual fragmentation peaks are not valuable; as in Edman degradation, it is their m/z differences that are informative. As shown in Fig. 1.3, the m/z differences between these peaks determine both the amino acid residue identities and their positions, thus identifying a peptide.

These two fragment ions have predictable structures because as shown in Fig. 1.4, fragmentation can only occur in three places along an ion's backbone. Therefore, the fragment ion will resemble one of six ion structures. The standard nomenclature for these fragment ions identifies both the point of fractionation as well as which terminus retains the charge. Ions a, b and c are n-terminus fragments and x, y and z are c-terminus ions.

Although the exact point of fragmentation depends on many factors, the primary factor is the type of dissociation applied. CID and HCD produce primarily b and y ions, with a few a ions sprinkled in, while ETD produces primarily c and z ions. Their resulting fragmentation patterns differ enough to impact the programs interpreting mass spectra.

1.3.2 Interpretation of MS² Spectra

With just one experiment generating hundreds of thousands of MS¹ and MS² spectra with high resolution, today's mass spectrometers now offer unparalleled mass accuracy and efficiency. Coupled with the increasing use of new dissociation techniques and chromatography methods, mass spectrometers now generate an overwhelming amount of spectral data with different fragmentation patterns and retention time profiles. Unfortunately, widely used software packages for interpretation of mass spectra, that is, for peptide identification, protein inference and validation, were not designed to process this vast amount of data, and they were tuned to process CID-derived data. Because these tools for interpreting mass spectra have failed to keep pace with advances

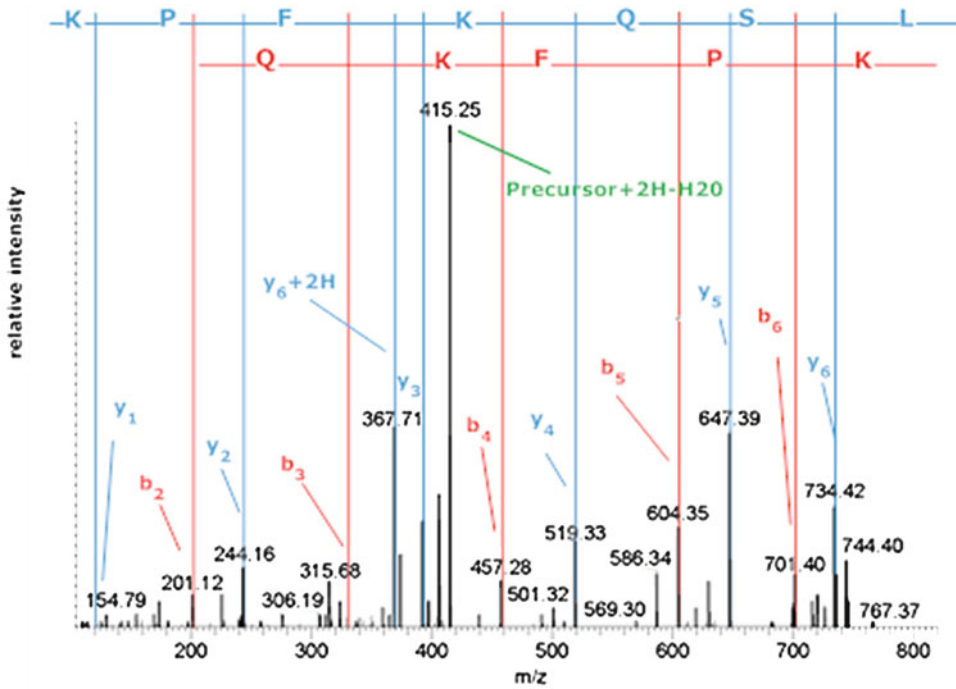


Fig. 1.3 Example peptide MS² spectrum

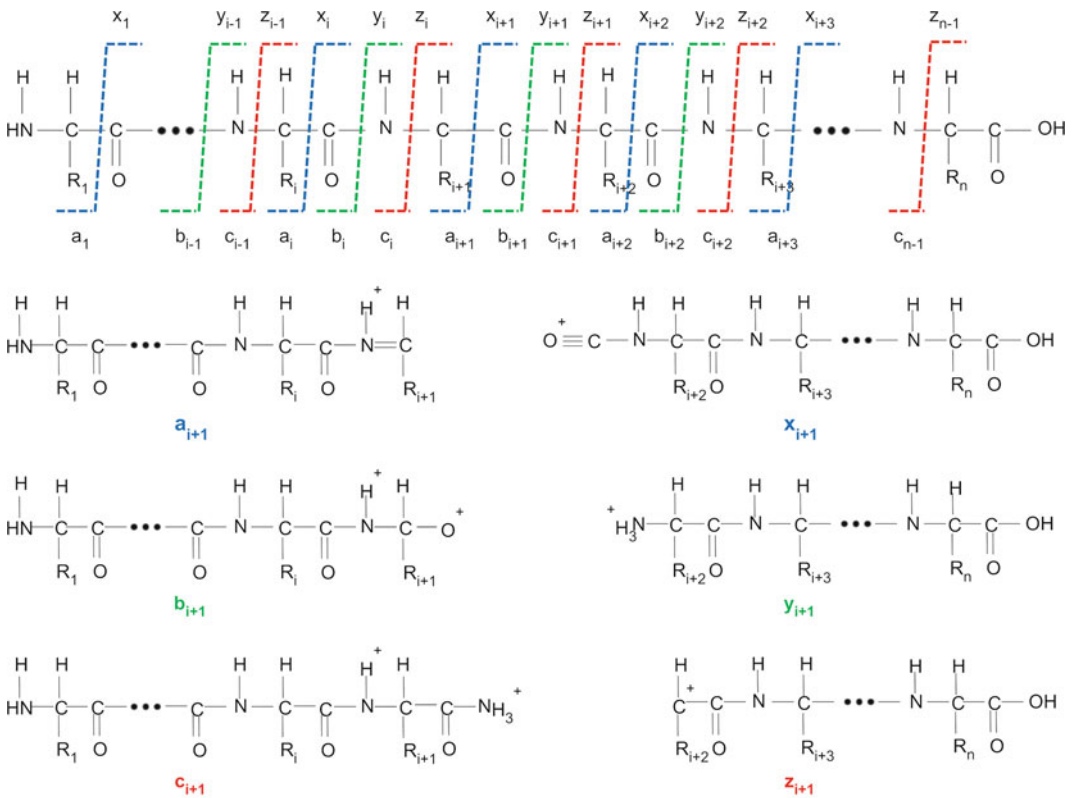


Fig. 1.4 Peptide fragmentation locations and the resulting ions

Table 1.1 Partial list of database search tools

| Program | Reference | Website |
|----------------------|-----------|---|
| SEQUEST | [7] | www.thermoscientific.com |
| Mascot | [8] | www.matrixscience.com/ |
| X!Tandem | [117] | www.thegpm.org/TANDEM/ |
| OMSSA | [118] | www.pubchem.ncbi.nlm.nih.gov/omssa/ |
| Andromeda (MaxQuant) | [119] | www.maxquant.org/ |
| SpectrumMill | | www.chem.agilent.com |
| MyriMatch | [120] | http://fenchurch.mc.vanderbilt.edu/software.php |
| ProteinProspector | [121] | http://prospector.ucsf.edu/prospector/mshome.htm |
| PHENYX | [122] | www.genebio.com/products/phenyx |

in instrument technology, they yield suboptimal proteome characterization.

Interpretation of mass spectra is a multistep process. The data must first be preprocessed to remove noise and identify valid peaks and features, subjects not reviewed here, but several good reviews exist in the literature [112–115]. After preprocessing the sample data, a series of phases culminates in a list of peptides and/or proteins that are confidently deemed present in the sample. These phases are: peptide identification, protein inference, and validation. In the following sections we highlight their main challenges and solutions, and posit an outlook of their future.

1.3.2.1 Peptide Identification

The first phase, peptide identification assigns an amino acid sequence to a spectrum. This is called a peptide spectrum match or PSM. Peptide identification programs have evolved over time, but strategies for assigning PSMs fall into one of four categories: database search, spectral library search, de novo sequencing, and hybrids thereof.

Database Search

During the early days of proteomics experiments, peptide identification was completed via manual de novo sequencing, a tedious process carried out by researchers without the aid of a computer or a database [116]. However, soon proteomics experiments became high throughput and the amount of data generated by them outpaced researcher's ability to manually inspect each spectrum. This drove the invention of alternate means of identifying peptides, mainly database search programs.

Today, researchers avail themselves to the numerous software packages that implement

database search programs, see Table 1.1. Researchers still commonly utilize the first widely used database search programs from the 1990s, SEQUEST [7] and Mascot [8]. Although specific implementations of database search programs differ, they share a common underlying principle introduced by SEQUEST: they compare the observed MS^2 spectra to that of theoretical spectra derived from in-silico enzymatic digestion of a FASTA database. They also share common challenges. One challenge is how to efficiently search the large amount of data available in FASTA databases. Searching all possible peptides from a FASTA database and all of their potential PTMs is prohibitively time-consuming. Even with the use of multiple processors, sequence assignment, including possible PTMs, to hundreds of thousands of MS^2 spectra produced by modern instruments can take days or even weeks. Unfortunately, limiting the peptides only to those with expected enzyme cleavage sites (e.g., lysine and arginine for trypsin cleavage), and limiting the number of PTMs considered, does not adequately narrow the search space. To address this issue, most database search software packages can restrict the search space even further by searching against only those peptides that have a mass within a narrow tolerance window around the observed m/z of its precursor peptide ion. A completely different challenge stems from the fact that different dissociation methods produce very different fragmentation patterns. This was not a problem until recently, because prior to the introduction of ETD, the predominant workhorse of proteomics experiments was CID. But with the introduction of ETD and its increasing adoption comes the

requirement for database search programs to allow multiple fragmentation patterns for the same peptide. Because each type of experiment has its own optimal settings for precursor ion mass tolerance window setting, number of PTMs considered and dissociation methods used, the researcher sets these parameters.

An inconvenient consequence of parameter driven database search is that each different set of parameters produces different results. Therefore, researchers must exercise caution when comparing results between experiments, both within and between laboratories.

Although database search strategies are the predominant choice for peptide identification in shotgun proteomics [123], they do have limitations. First, database search relies on the sequencing data for organism being studied. Thus, if an organism is not yet been sequenced, database searching can only be used to find homologous peptides in different organisms. Second, unexpected, yet important, PTMs, and sequence anomalies will be missed because variants do not exist in the database. Even though some databases take into consideration splice variants, no production quality database search engines make the effort to take advantage the annotation available in databases such as Swiss-Prot or UniProtKB. Therefore, many unexpected, but annotated, PTMs and polymorphisms are missed, which leads to incorrect or missed peptide identifications [124]. Third, false positive identifications occur often because database search programs assign a peptide sequence to each and every spectrum, regardless of quality. Fourth, validation assigns high confidence (>95%) to only 10–30% of the spectra [125]. Finally, each database search engine identifies a partially overlapping but different set of peptides. For instance, SEQUEST may identify a set of 100 peptides and Mascot may identify a set of 100 peptides from the same spectra, but perhaps only 60 of them are common to both SEQUEST and Mascot.

A relatively recent development in shotgun proteomics research combines the results from several different database search engines to identify more peptides with increased confidence.

The idea of combining results from multiple sources is not new. Resing et al. described consensus scoring for multiple peptide identifications from different search engines in 2004 [126] and Alves et al. proposed combining and calibrating confidence scores from multiple search engines into a meta-analytic value for each confidence score [127]. However, software automating integration of separate database search results developed more recently. For instance, a popular tool allowing researchers to combine results from multiple search engines is Scaffold, developed by Searle et al. [125, 128]. By probabilistically combining results from multiple search engines, including SEQUEST, X!Tandem, OMSSA, InSpecT and Mascot, Scaffold increases sensitivity a minimum of 20% with each search engine added [129]. As evidenced by the latest publication of a tool combining results from multiple search engines [130], the idea is garnering more attention and we can expect this trend of new tools for incorporating multiple search engines to continue into the foreseeable future.

Spectral Library Search

Spectral library search strategies are similar to database search strategies, except the observed MS² spectra are compared to collections of experimentally generated spectra rather than hypothetical spectra [131]. These strategies outperform database search strategies in terms of error rates, speed and sensitivity. Using spectral libraries reduces the time spent repeatedly identify the same identifiable peptides by database searching, [132] but can only identify a peptide if it has been previously analyzed by tandem mass spectrometry and its sequence positively identified. A partial list of spectral library search tools is located in Table 1.2.

Libraries of experimental spectra are available from many sources and provide a rich source of spectral data. Spectral libraries for many organisms are stored at the National Institute of Standards and Technology (NIST). Although, the NIST libraries do not target specific PTMs, specialized libraries for specific modifications are available elsewhere, e.g., PhosphoPep [138] for phosphorylation sites in model organisms and

Table 1.2 Partial list of spectral library search tools

| Program | Reference | Website |
|------------------|------------|---|
| SpectraST | [132, 133] | www.peptideatlas.org/spectrast/ |
| NIST MSPepSearch | N/A | http://peptide.nist.gov/ |
| BiblioSpec | [134] | http://proteome.gs.washington.edu/software/bibliospec/v1.0/documentation/index.html |
| X!Hunter | [135] | http://www.thegpm.org/hunter/index.html |
| ProMEX | [136] | http://www.promexdb.org/home.shtml |
| HMMatch | [137] | Not available |

the open source Ub/Ubl spectral library [139] for ubiquitin and ubiquitin-like modifications. In addition, a wealth of spectral data can be downloaded from one of several proteomic data sharing repositories, e.g., PeptideAtlas [140], Pride [141], Peptidome [142], and Tranche (<https://trancheproject.org/>).

As the amount of publicly available spectral data grows, the hope is that one day spectra for all peptides detectable by MS (at least for well-studied organisms) will be contained and annotated in publicly available spectral reference libraries. However, until these reference libraries are sufficiently complete, spectral library search strategies will continue to be underutilized [143]. In the meantime, data in spectral reference libraries are a rich source of data that could be used for purposes other than identifying peptides. For instance, spectral data could be mined to provide important insight into fragmentation patterns, which could in turn lead to improved database search or de novo sequencing [144] as well as the development of SRM methods to target specific peptides [97].

De Novo Sequencing

The limited ability of database search and spectral search strategies to identify the unexpected, for example, PTMs, polymorphisms and sequence anomalies drives the need for peptide identification programs that can efficiently handle the enormous amounts of data without sacrificing confidence in their results. While conceptually unchanged, researchers are again turning to de novo sequencing as an alternative to database search to accurately and confidently identify peptides. De novo sequencing programs can identify PTMs, polymorphisms and sequence anomalies because they compute directly on

spectra to determine the peptide amino acid sequence, process which does not require searching against FASTA databases [145].

De novo sequencing for proteomics has a long and rich history. Spectra were originally sequenced manually, a process which does not scale well. Therefore, as the amount of data from shotgun proteomics grew, researchers turned to computer science for automated de novo sequencing. In the 1980s, several computational algorithms were introduced that helped [146–150] but proved to be terribly slow because they tended to brute force consider all possible amino acid sequences. In 1990, computational algorithms became more efficient when Bartels represented a spectrum as a graph [151]. Although this type of graph is called a spectrum graph by the proteomics community, it should not be confused with spectral graph theory where a graph's spectrum is defined as the set of eigenvalues of a graph's adjacency matrix, nor with a general graph of nodes and edges, where a node does not have a position and an edge can connect any two nodes. In this novel proteomics spectrum graph representation, the vertices represent the spectrum m/z values and two vertices are linked by edges if their mass difference is equivalent to the mass of an amino acid. Figure 1.5 shows a theoretical spectrum graph of the spectrum in Fig. 1.2. Formally, Bartels defined the problem as:

Given amino acid masses $M = \{m_1, \dots, m_{20}\}$, spectrum $S = \{s_1, \dots, s_c\}$, transform it into a spectrum graph $G(V, E)$ such that $V = \{v_1, \dots, v_c\}$ and $G = \{g_1, \dots, g_t\}$ such that v represents a single integer m/z and two vertices $v_n, v_q, q \neq n$ are connected by directed edge e if $|v_q - v_n| \sim m_t$

Although Bartel's approach is now the de facto basis for most de novo peptide sequencing programs, several unresolved issues limited

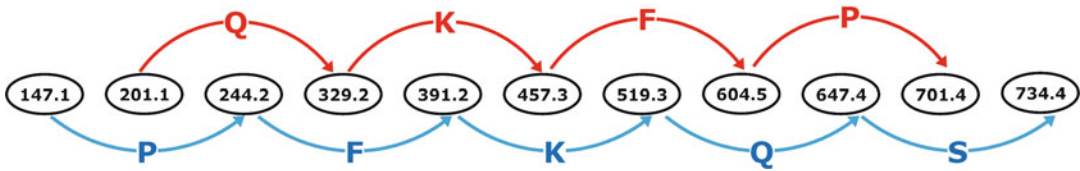


Fig. 1.5 MS² spectrum graph

Table 1.3 Partial list of de novo sequencing tools

| Program | Reference | Website |
|----------|-----------|---|
| PepNovo | [155] | http://proteomics.ucsd.edu/ |
| DiracTag | [157] | http://fenchurch.mc.vanderbilt.edu/software.php |
| PEAKS | [158] | www.bioinform.com/ |
| pNovo | [159] | http://proteomics.ucsd.edu/ |
| Lutefisk | [152] | http://www.hairyfatguy.com/lutefisk/ |
| NovoHMM | [156] | http://people.inf.ethz.ch/befische/proteomics/ |
| DeNovoX | N/A | www.thermo.com |

spectrum graph's, and, therefore, de novo sequencing's, adoption. First, spectrum graph models were instrument specific which required training a new model for each new mass spectrometer. It took until 1997 to implement this strategy when Taylor and Johnson introduced Lutefisk [152]. Second, the predominantly used CID has a propensity for incomplete fragmentation and results in multiple disconnected graphs, graph gaps, limiting the effectiveness of spectrum graph algorithms. Finally, lack of standardized scoring models for spectrum graphs hindered researchers' ability to compare experimental results. SHERENGA, introduced by Dancik et al. in a landmark publication 1999 [153], addressed each of these limitations using a spectral graph-based algorithm. Several research groups have since made additional enhancements to these algorithms, most notably dynamic programming [154] and probabilistic models using networks learned over annotated spectra (e.g., PepNovo [155] and NovoHMM [156]).

Although several de novo sequencing software packages, as shown in Table 1.3, are now available that implement spectrum graph algorithms, de novo sequencing's adoption as a viable option for shotgun proteomics experiments has been slow. A primary contributor to its slow adoption was that affordable mass spectrometry instrumentation lacked the ability to produce high resolution spectra with minimal noise,

completely fragment selected ions and retain potentially important PTMs. Until recently, these issues could only be overcome by using complementary spectra, MS² and MS³ [160], ECD and CID [161], or differentially modified pairs [162] on a FTICR mass spectrometer. The FTICR spectrometer can generate spectra with high resolution (>100,000), thus differentiating valid ions from noise much easier. Furthermore, ECD is complementary to CID, and a more complete fragmentation pattern emerges their spectra are combined into a single artificial spectrum. Finally, ECD inherently uses lower energy than CID allowing for retention and subsequent identification of more PTMs. FTICR mass spectrometers' main drawbacks are that they are extremely expensive and inefficient compared to ion trap mass spectrometers, the main workhorse instruments in MS-based proteomics. Even though FTICR instruments offer 100 times better resolution than an ion trap, each spectrum takes 10 times longer to acquire than on an ion trap instrument.

Recently, via the introduction of the Orbitrap instrument series, the more affordable ion trap mass spectrometers became capable of high resolution (>100,000) and offered ETD, which gives similar spectra to ECD. Taking advantage of these improvements, Datta and Bern expanded on previous pioneering work fusing ECD and CID spectra [163]. In 2009, they introduced Spectrum

Table 1.4 Partial list of hybrid search tools

| Program | Reference | Website |
|------------------------|-----------|---|
| InSpecT | [166] | http://proteomics.ucsd.edu/Software/Inspect.html |
| GutenTag | [165] | http://fields.scripps.edu/ |
| DirecTag | [157] | http://fenchurch.mc.vanderbilt.edu/software.php |
| TagRecon | [167] | http://fenchurch.mc.vanderbilt.edu/software.php |
| PeaksDB | [158] | www.bioinform.com |
| Paragon (ProteinPilot) | [168] | www.absciex.com |

Fusion which uses a global graph partitioning approach to both separate b and y ions and to fuse CID and ETD. The heart of Spectrum Fusion is a supervised machine learning algorithm (tree augmented naïve Bayes network) trained on confidently identified spectra from a prior database search. The result is a synthetic spectrum with only b ions which can then be sequenced by a slightly modified spectrum graph de novo algorithm.

Hybrid Strategies: De Novo & Database/Spectral Library Search

Despite advances in both mass spectrometry instrumentation and software programs, incomplete fragmentation remains an open issue for de novo sequencing strategies. However, when dissociating thousands of ions, they often break along the backbone in enough places so that de novo programs can sequence and identify short peptide sequences which are typically 3–5 short amino acids in length. Again, these ideas are not new. In fact, Mann and Wilm introduced the notion of using short peptides sequences, which they called sequence tags, in 1994 [164], the same year as SEQUEST. However, strategies based on sequence tags did not appear until Tabb et al. published GutenTag program in 2003 [165]. The innovation of GutenTag is that it constructs a model spectrum of the peaks expected from a given sequence tag, compares the observed spectrum and the model spectrum, and generates a correlation score. Tabb et al. went on to provide an enhanced database search tool MyriMatch [120], which is tuned to use these short peptide sequences to infer candidate proteins. Hybrid peptide identification strategies using sequence tags are gaining popularity and several hybrid tools are now available (Table 1.4).

1.3.2.2 Protein Inference

While peptide identification is a necessary phase in proteome profiling, it is not the last one. Proteins must be inferred from the list of peptides identified. However, the task of assembling peptide identifications to infer proteins present in a sample, known as the protein inference problem, is far from trivial [169]. First, the connection between peptides and proteins is lost during enzymatic digestion. This is so because of multiple proteins sharing peptides. The sources of these shared peptides, also known as degenerate peptides, include both natural and artificial phenomena. Degenerate peptides arise often in nature, especially in eukaryotic organisms due to the presence of homologous sequences or splice variants. To make matters worse, errors and redundancies in the database being searched add even more, albeit artificial, degenerate peptides [170]. Regardless of their source, degenerate peptides limit the ability to differentiate between proteins resulting in an unsatisfactory level of ambiguity. This drives the need for validation of results.

1.3.2.3 Validation

MS-based proteomics results are inherently prone to inaccuracies. Without careful filtering, its results are riddled with false positive identifications at both the peptide identification and protein inference levels. To reduce the number of false positives, several scoring models have been proposed and developed to impart a confidence level on identified peptides and inferred proteins. To date, because no single scoring model dominates, different software packages employ their own scoring models. SEQUEST, X!Tandem, Mascot and OMSSA employ variations of a cross correlation (XCorr) score which measures the similarity at different offsets between pre-processed

observed spectra and hypothetical spectra generated by in-silico digestion. Mascot differs slightly from other XCorr based scoring models in that it assesses the probability of a peptide spectrum match being a random event. Other software packages use scoring models based on empirically observed rules, SpectrumMill, or incorporate statistically derived fragmentation frequencies, PHENYX [122].

Each of the thousands of single peptide identifications or protein inferences can be assigned an individual score. However, single case scores do not take into consideration the fact that multiple hypotheses are being tested. Therefore, in addition to using a single statistic, p -value, its close relative for multiple testing, E -value, is often used. p -value, assuming the null hypothesis is true, represents the probability of obtaining a test statistic at least as extreme as the one observed. E -value, assuming the null hypothesis is null, is the expected number of times in multiple testing to obtain a test statistic as extreme as the one that was actually observed. Put more simply, E -values are derived by taking the number of tests multiplied by the p -value. To account for multiple hypotheses testing, many controlling measures have been proposed. Bonferroni correction is used to control Family Wise Error Rate, FWER, which is the probability of finding at least one false positive. However, the Bonferroni correction has been shown to be too conservative given the thousands of hypothesis tests in a single experiment [171].

Less conservative than the Bonferroni correction is the False Discovery Rate (FDR) controlling procedure, introduced by Benjamini and Hochberg [172]. They define FDR as the “expected fraction of mistakes among the rejected hypothesis and suggested to control FDR in multiple testing”. A well-established mechanism to implement FDR for database search results is to search against a decoy FASTA database of invalid peptide sequences, most often concatenated to the end of the target FASTA database with valid peptide sequences [173]. The premise for this approach is that a spectrum will match valid and random (invalid) sequences with equal probability and target and decoy sequences do not overlap. Although decoy databases are intended

to be random, in practice they are most often constructed by reversing, shuffling or randomizing the target FASTA database [174].

With the introduction of FDR as a controlling procedure, publications ensued discussing the proper use of statistical values. Kall et al., argue that using a p -value threshold for FDR is inadequate because the statistical test is performed so many times [175]. It also has the unfortunate property that two different p -value scores can result in the same FDR. To address this problem, Storey and Tibshirani [176] propose a q -score, which when applied to shotgun proteomics, is the defined as the minimum FDR threshold at which a given PSM will be accepted.

Historically, to implement the FDR controlling procedure with a decoy database, researchers accepted all identifications above a certain threshold [177]. This threshold was usually a combination of scores provide by the database search engine. However, problems exist in this strategy, including the need to have separate thresholds for different types of instruments. To overcome problems with the threshold scheme, early validation tools, e.g., QSCORE [178], were developed that employed simple probability, but focused on results from a single search engine.

Because threshold statistical models tend to be instrument specific, researchers turned to machine learning, notably mixture modeling, to build a generic model that could process results from multiple instrument types. Mixture modeling uses models of two normal distributions, one for correct identifications and one for incorrect distributions, to determine a score threshold. Perhaps the most widely used example of mixture modeling for peptide identification validation is Keller et al.’s PeptideProphet [173]. It uses a discriminant score which is derived by converting several scores from the database search programs into a single score. To apply a two-component mixture model, PeptideProphet creates a histogram of discriminant scores and uses curve fitting to draw the correct and incorrect distributions. Using Bayesian statistics, it computes the probability of an identification being correct given its discriminant score. Similar to PeptideProphet is ProteinProphet, which is

used to validate protein inferences. It uses results from PeptideProphet as input to accurately compute the probability that an inferred protein is present in the sample [179] and derives a mixture model of correct and incorrect protein inferences, using an expectation-maximization routine (EM). Since PeptideProphet/ProteinProphet is open source and freely available and integrated into the Trans-Proteomic Pipeline, TPP, it is an attractive option for interpreting mass spectra as evidenced by its use in a number of prominent laboratories.

Although scoring based on mixture modeling can accurately model incorrect and correct score distributions, they are inherently complex and not easily extensible [180]. Therefore, other score models have been proposed. For instance, IDPicker is based on a simple non-parametric Monte Carlo simulation method. IDPicker employs FDR identification aggregation instead of individual identification probabilities, and it is easily extended to accept scoring metrics from multiple search engines, as long as the decoys are provided in the searched database [180].

In a recent departure from the canonical target decoy approach, Kim et al., propose MS-GF which uses generating functions and their derivatives without a decoy database [181]. They argue that by using a decoy database, the proteomics community is de facto acknowledging that it has been unable to solve the following Spectrum Matching problem: “Given a spectrum S and a score threshold T for a spectrum-peptide scoring function, find the probability that a random peptide matches the spectrum S with score equal to or larger than T ” [181]. This problem assumes certain underlying distributions on which probabilistic calculations can be applied. Ideally, the underlying distributions would be purely theoretical in nature to allow the direct calculation of probability and expectation values. However, the sheer number possible parameters makes modeling the theoretical underlying distribution impractical [182]. Instead, p -values and E -values are calculated using heuristic algorithms working on empirically derived distributions. In contrast to the heuristic algorithms, MS-GF demonstrates

that it is possible to compute the precise number of peptides identified in a huge database, solving the Spectrum Matching problem.

1.3.3 Outlook

Although many difficulties exist in thoroughly characterizing a proteome no consensus has been reached by the proteomics community on which the peptide identification, protein inference and validation strategies should be used. This is largely due to the fact that shotgun proteomics is relatively immature and more complex compared to other fields such as genomics. Whereas the genomics community can readily compare results from experiments conducted in different laboratories, the proteomics community has difficulty doing so because reporting of results is not standardized. For instance, some shotgun proteomics researchers will report proteins inferred from a single peptide while others will only report proteins inferred from two or more distinct peptides. If a peptide is shared between multiple proteins, some researchers randomly assign the peptide to a protein, while others apply Occam’s razor or other statistical models. This is compounded by the availability of vastly different FASTA databases for a single organism. The differences mainly stem from their curation processes, or lack thereof, and their sources of deposited sequences.

Reporting standards for shotgun proteomics experiments may be lacking consensus, but serious effort has been made to rectify this problem. In 2002, the Human Proteome Organization, HUPO, launched the Proteomics Standards Initiative, PSI. Its goal was and is to “define community standards for data representation in proteomics to facilitate systematic data capture, comparison, exchange and verification.” [183–187]. Although HUPO sets standards for the broader proteomics community, publishing criteria was still lacking for shotgun proteomics results. To address this, about 30 key people in the proteomics community met in Paris to develop set

of standards focused on publication of shotgun proteomics results. These standards published in 2006 [188] as the Paris Guidelines, and updated in 2009 are slowly being adopted by proteomics journals.

1.4 Label-Free Quantification

The initial application of the MS-based proteomics platform addressed the challenge of cataloging proteins within complex samples. However, biological researchers also need to quantify proteins because proteomes are highly dynamic systems, and their abundances change due to regulation of their synthesis and degradation. Protein activities are dynamically regulated via the addition or removal of PTMs. Therefore, to make MS-proteomics a technology truly useful to researchers who are trying to understand living systems, it must be able to quantify abundance and PTM differences between samples.

The initial technology for quantitative MS-based proteomics involved differential labeling methods with stable isotopes. Isotope labeling methods for quantitative MS-based proteomics have been reviewed in detail [189, 190]. These methods label proteins and/or peptides with stable isotopes (^{15}N , ^{13}C , ^{18}O) through a variety of mechanisms. Stable isotope labeling in cell culture (SILAC) labels proteins via metabolic incorporation of stable-isotope containing amino acids contained in cell culture media [191, 192]. Other methods introduce stable isotopes via reactive chemical tags, such as the isotope affinity tag (ICAT) [193] or isobaric peptide tagging (e.g. iTRAQ, TMT) [194, 195] methods. Labeling with O^{18} is accomplished via enzymatic means at the c-terminus of peptides within complex mixtures [196]. For all of these methods, distinct protein mixtures are first differentially labeled, one with isotopically normal amino acids or chemical tags, and the other with isotopically “heavy” amino acids or chemical tags. Although most labeling methods compare protein abundance between two distinct mixtures, some are

capable of multiplexed analysis, such as iTRAQ labeling which can compare up to eight [197]. After labeling, the mixtures are combined and peptide digests are fractionated and analyzed by MS. Peptide sequences common to both samples, although differentially isotopically labeled, retain the same chemical properties and behave similarly during fractionation. Consequently, differentially isotopically labeled peptides are detected simultaneously and their m/z differences resolved in the MS. Peptides are selected for MS^2 and identified via subsequent sequence database searching. For identified peptides, relative abundance levels between samples are determined via comparison of the mass spectral peak intensities corresponding to the normal or heavy isotope labels.

Although still used prominently, stable isotope labeling has its limitations. One is cost. Stable isotope labeled amino acids or chemical tags are costly to synthesize, and purchase of these can run from hundreds to thousands of dollars, depending on the labeling method used. Another is applicability to only certain biological sample types. SILAC, arguably the most accurate stable isotope labeling method, is only applicable to experiments using cell culture models, although extremely expensive studies of whole organism labeling with stable isotopes in mice and worm have been described [198]. For human and other animal studies, chemical tagging methods, such as iTRAQ or TMT, must be used for stable isotope labeling. Unfortunately, the accuracy of iTRAQ and TMT for measuring relative abundances, which are based on MS^2 fragmentation of labeled peptides, is decreased due to simultaneous fragmentation of multiple peptides in shotgun proteomics [199].

Responding to these limitations, label-free technology has emerged which obviates the need for stable isotope labeling for quantitative proteomics. Two methods underpin the label-free MS-based quantitative proteomics technology: spectral counting and intensity-based measurements. Figure 1.6 details these two methods.

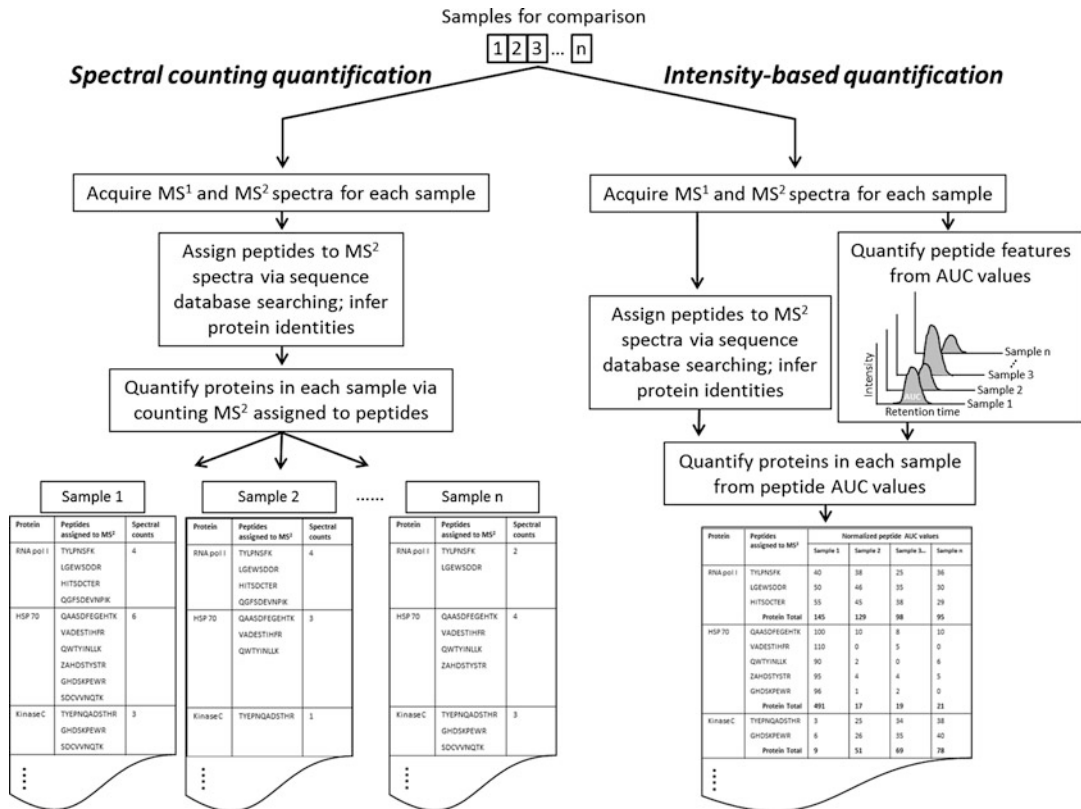


Fig. 1.6 Label-free quantification methods

1.4.1 Spectral Counting Quantification

Spectral counting is based on the core instrumental method used in MS-based shotgun proteomics. Here, peptides separated via LC are detected and selected for CID fragmentation using a data-dependent routine. The fragmentation spectra are recorded as MS² spectra. Peptides are identified by assigning a sequence to each MS² from databases of known protein sequences and a variety of software programs, as described in Sect. 1.2. Protein identities in the starting mixture are inferred from the identification of peptides that are a part of their amino acid sequence. Quantification via spectral counting is based on the observation that the number of peptides identified from MS² spectra is proportional to the abundance of the protein in the starting mixture: more abundant proteins result in more identified peptides while less abundant proteins result in

fewer identified peptides. Protein quantification is achieved by simply counting the number of MS² spectra assigned to peptides within a given protein, without taking into consideration the peptide MS signal intensity. Because quantification is based on peptides assigned to MS² spectra, spectral counting benefits from MS instruments with higher mass accuracy and sensitivity, which increase the number of high confidence peptide identifications [20].

Early on, spectral counting was done in a rather simple manner, simply summing the number of peptide identifications corresponding to each inferred protein. However, as this method increased in popularity, more sophisticated quantification approaches based on spectral counting have emerged. Several extensive reviews have recently appeared on spectral counting [200]. Here we discuss the most commonly used approaches to spectral counting quantification and some representative studies which have used this method.

Table 1.5 Summary of open-source software for label-free quantification

| Program | Reference | Website |
|----------------------------------|-----------|---|
| Spectral counting quantification | | |
| APEX | [205] | http://pfgrc.jcvi.org/index.php/bioinformatics/apex.html/ |
| Census | [210] | http://fields.scripps.edu/census/ |
| emPAI | [202] | http://empai.iab.keio.ac.jp/ |
| PepC | [211] | http://sashimi.svn.sourceforge.net/viewvc/sashimi/trunk/trans_proteomic_pipeline/src/Quantitation/Pepc |
| SI _N ^a | [212] | N/A |
| Intensity-based quantification | | |
| IDEAL-Q | [213] | http://ms.iis.sinica.edu.tw/IDEAL-Q/ |
| MaxQuant | [214] | http://maxquant.org/ |
| MSInspect | [215] | http://proteomics.fhcrc.org/CPL/msinspect/index.html |
| MZMine | [216] | http://mzmine.sourceforge.net/download.shtml |
| PEPPER | [217] | http://www.broadinstitute.org/cancer/software/genepattern/desc/proteomics#pepper |
| SuperHirn | [218] | http://prottools.ethz.ch/muellelu/web/SuperHirn.php |

^aApproach combines both spectral counting and intensity-based quantification

Spectral counting must take into account a protein's length because a longer protein, when enzymatically digested, will produce more peptides than a shorter protein for the MS to detect. Without correction, protein quantification by spectral counting would be biased towards longer proteins. As a consequence, an approach taking into account protein length was developed [201] which provided a normalized spectral abundance factor (NSAF) for each identified protein. The abundance of any given protein within a mixture can be estimated by dividing its NSAF value against the sum of NSAF values for all identified proteins.

An alternative approach to NSAF is the protein abundance index (PAI) [202], which was further improved to the exponential modified PAI, or emPAI [203]. This approach used the number of peptides actually identified from a protein, divided by the estimated total number of peptides expected to be identified for that same protein. The expected peptides were estimated based on the proteins sequence and the sizes of peptides derived from the protein after enzymatic digestion. The relative molar amount of any given protein within a sample can then be calculated by dividing its emPAI value against the sum of all emPAI values within the mixture. The emPAI approach was deployed in a freely available application, emPAI Calc that accepts data from a variety of sequence database searching programs [200].

Another approach, Absolute Protein Expression (APEX), tries to correct for physiochemical variations between peptide sequences that may affect their identification in the MS, and bias spectral counting results. APEX uses a correction factor that attempts to use properties such as amino acid content and length of peptides [204] to assess the probability of any given peptide for MS detection and subsequent identification from MS² spectra. This correction is applied to the spectral counts corresponding to each identified protein, to provide a more accurate measurement of its abundance. APEX has been released as an open source application [205].

Spectral counting has been widely applied. Its application is reviewed in detail elsewhere [200, 206]. Software plays a key role in the automating spectral counting quantification. Table 1.5 shows a summary of the most popular open-source software available. One particularly powerful application uses spectral counting and NSAF values to quantify relative abundance of proteins within functional complexes. Estimation of relative stoichiometry of the different members of protein complexes [207], as well as modeling of protein-protein interaction networks [201] is possible. An interesting application using the emPAI approach identified and quantified relative abundance levels of over 100 proteins in the chicken egg white proteome [208]. APEX was recently used to characterize proteome abundance

differences between mutant strains of the thermophilic anaerobic bacterium *Clostridium thermocellum*, an organism with promise for biofuel production [209].

1.4.2 Intensity-Based Quantification

An alternative to spectral counting is intensity-based measurements of peptide abundance. During a nanoLC-MS analysis, the mass-to-charge (m/z), retention time and signal intensity values are continuously recorded for each detected peptide. This information can be used to reconstruct a chromatographic peak for each peptide. This quantification method estimates the area under the curve (AUC) of the chromatographic peak (Fig. 1.5). The AUC correlates linearly with peptide concentration across a range of low femtomole amounts to tens of picomoles in most contemporary MS instruments [219, 220]. Similar to spectral counting, peptides are identified via MS^2 and sequence database searching, and protein identities are inferred from these peptides. For comparisons of peptide and inferred protein abundance between different samples, each sample is analyzed by nanoLC-MS separately. AUC values calculated for detected peptides in each distinct sample are compared to determine relative abundance. Intensity-based measurements are not used for quantification of different peptides within the same sample, because each peptide sequence ionizes with different efficiency, making comparison based on signal intensity inaccurate.

Although simple in concept, successful implementation of intensity-based quantification relies heavily on sophisticated software. Open-source software choices have been reviewed elsewhere [221]. Some of these choices are summarized in Table 1.5. This software automates critical data processing steps needed to insure accurate results based on AUC values. A recent review by Christin and colleagues [221] thoroughly describes these steps. One key step is proper alignment of peaks corresponding to the same peptide across all separate nanoLC-

MS data sets. Proper alignment, based on peak m/z values and retention time, assures that the AUC values being measured in each sample correspond to the same detected peptide. Use of highly reproducible nanoLC systems with high chromatographic resolving power can help for alignment [222], although ultimately effective alignment via software is critical. High accuracy measurements of peptide m/z values using newer MS instruments has greatly helped with alignment across separate nanoLC-MS datasets. One nice feature of peak alignment, aided by high mass accuracy data, is that a peptide need only be identified by MS^2 in one sample [221]. Peaks in other samples aligning in retention time and accurate m/z can then be confidently assigned to that peptide without the need for their identification from MS^2 spectra.

Another key step is normalization of measured AUC values. Normalization accounts for bias and variability in measured AUC values introduced during sample processing, loading of sample to the nanoLC column, and in-run variability of MS response. A number of normalization procedures have been developed which are effective for minimizing variability and improving accuracy [223, 224].

As with spectral counting, applications of intensity-based quantification are numerous. These are reviewed in detail elsewhere [225, 226]. These different applications have used a variety of publically available software programs for accurate quantification, some of which are summarized in Table 1.5. Here we discuss several representative applications. One interesting, radiation research-relevant example, demonstrated the effectiveness of intensity-based quantification to compare effects of ionizing radiation on colon cancer cells compared to a mock-treated control [227]. Disease biomarker discovery has also been a popular application of intensity-based quantification. Such studies have been done in paraffin embedded archival cancer tissues [228], as well as serum fluid from schizophrenia patients [229].

Overall, label-free quantification addresses many of the limitations of stable-isotope labeling-based technology. Both spectral counting and

intensity-based measurements are cheap and simple, with no need for purchase of costly labeling reagents or extra sample labeling and processing steps. Spectral counting provides the additional benefit of measuring relative abundance of proteins within the same sample, whereas stable isotope labeling only measures relative abundance across separate samples. Intensity-based measurements, when using effective software for aligning peptide peaks across samples, obviates the need for time- and computation-intensive MS² acquisition and subsequent peptide identification via sequence database searching. This method therefore is an attractive choice for biomarker studies, where comparison across many patient samples with high throughput is desirable.

Despite numerous strengths, label-free quantification is not without limitations. Unlike some labeling methods, notably the iTRAQ or TMT methods, multiplexed comparative analysis within a single MS experiment is not possible. Instead, each sample being compared must be analyzed in a separate MS experiment, and preferably with technical replicates to achieve statistical significance [230]. Consequently, large amounts of instrument time are required, which may not be feasible, especially for researchers relying on sample analysis via a fee-for-service facility. Low-abundance proteins also remain a challenge for both methods. Because spectral counting relies on multiple peptides to be identified from each inferred protein to achieve statistical significance, low abundance proteins identified by only a few peptides cannot be accurately quantified. For intensity-based measurements, peptide peaks from low-abundance proteins also suffer from low signal-to-noise ratios, challenging their accurate quantification. Improved instrument sensitivity should only help to increase the ability to identify more peptides derived from low abundance proteins, and improve the effectiveness of both label-free methods. Recently, a promising new method was described [212] which combines spectral counting and intensity-based measurements, thereby capitalizing on the strengths of both methods and providing improved results.

1.5 Conclusions

Consistent with history, technological advances will continue to define and mature the field of MS-based proteomics, catalyzing new milestones of achievement. We anticipate these advances to primarily fall in the areas described in this review: new instrumentation and related methods, and new computational methods and software for identification and quantification of proteins from complex datasets. Continued maturation of MS-based proteomics should one day enable realization of its ultimate goal: comprehensive proteome characterization. Researchers seeking to better understand the effects of radiation on living systems will undoubtedly continue to benefit from the continued advances of this vital technology.

References

1. Griffin TJ, Gygi SP, Ideker T, Rist B, Eng J, Hood L, Aebersold R (2002) Complementary profiling of gene expression at the transcriptome and proteome levels in *Saccharomyces cerevisiae*. *Mol Cell Proteomics* 1(4):323–333
2. Washburn MP, Koller A, Oshiro G, Ulaszek RR, Plouffe D, Deciu C, Winzeler E, Yates JR 3rd (2003) Protein pathway and complex clustering of correlated mRNA and protein expression analyses in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci U S A* 100(6):3107–3112. doi:10.1073/pnas.0634629100 0634629100 [pii]
3. Tanaka K, Waki H, Ido Y, Akita S, Yoshida Y, Yoshida T, Matsuo T (1988) Protein and polymer analyses up to m/z 100,000 by laser ionization time-of-flight mass spectrometry. *Rapid Comm Mass Spectrom* 2(8):151–153
4. Fenn JB, Mann M, Meng CK, Wong SF, Whitehouse CM (1989) Electrospray ionization for mass spectrometry of large biomolecules. *Science* 246(4926):64–71
5. Deterding LJ, Moseley MA, Tomer KB, Jorgenson JW (1991) Nanoscale separations combined with tandem mass spectrometry. *J Chromatogr* 554(1–2):73–82
6. Hunt DF, Yates JR 3rd, Shabanowitz J, Winston S, Hauer CR (1986) Protein sequencing by tandem mass spectrometry. *Proc Natl Acad Sci U S A* 83(17):6233–6237
7. Eng JK, McCormack AL, Yates JRI (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

8. Perkins DN, Pappin DJ, Creasy DM, Cottrell JS (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 20(18):3551–3567. doi:10.1002/(SICI)1522-2683(19991201)20:18<3551::AID-ELPS3551>3.0.CO;2-2 [pii] 10.1002/(SICI)1522-2683(19991201)20:18<3551::AID-ELPS3551>3.0.CO;2-2
9. Gygi SP, Rist B, Griffin TJ, Eng J, Aebersold R (2002) Proteome analysis of low-abundance proteins using multidimensional chromatography and isotope-coded affinity tags. *J Proteome Res* 1(1): 47–54
10. Link AJ, Eng J, Schieltz DM, Carmack E, Mize GJ, Morris DR, Garvik BM, Yates JR 3rd (1999) Direct analysis of protein complexes using mass spectrometry. *Nat Biotechnol* 17(7):676–682
11. Washburn MP, Wolters D, Yates JR 3rd (2001) Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol* 19(3):242–247
12. Gygi SP, Corthals GL, Zhang Y, Rochon Y, Aebersold R (2000) Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology. *Proc Natl Acad Sci U S A* 97(17): 9390–9395
13. Ficarro SB, McClelland ML, Stukenberg PT, Burke DJ, Ross MM, Shabanowitz J, Hunt DF, White FM (2002) Phosphoproteome analysis by mass spectrometry and its application to *Saccharomyces cerevisiae*. *Nat Biotechnol* 20(3):301–305
14. Oda Y, Nagasu T, Chait BT (2001) Enrichment analysis of phosphorylated proteins as a tool for probing the phosphoproteome. *Nat Biotechnol* 19(4): 379–382
15. Zhou H, Watts JD, Aebersold R (2001) A systematic approach to the analysis of protein phosphorylation. *Nat Biotechnol* 19(4):375–378
16. Zhang H, Li XJ, Martin DB, Aebersold R (2003) Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. *Nat Biotechnol* 21(6):660–666
17. Flory MR, Griffin TJ, Martin D, Aebersold R (2002) Advances in quantitative proteomics using stable isotope tags. *Trends Biotechnol* 20(12 Suppl): S23–S29
18. Creasy DM, Cottrell JS (2004) Unimod: protein modifications for mass spectrometry. *Proteomics* 4(6):1534–1536. doi:10.1002/pmic.200300744
19. Michalski A, Cox J, Mann M (2011) More than 100,000 detectable peptide species elute in single shotgun proteomics runs but the majority is inaccessible to data-dependent LC-MS/MS. *J Proteome Res* 10(4):1785–1793. doi:10.1021/pr101060v
20. Mann M, Kelleher NL (2008) Precision proteomics: the case for high resolution and high mass accuracy. *Proc Natl Acad Sci U S A* 105(47):18132–18138. doi:10.1073/pnas.0800788105
21. Zubarev RA, Hakansson P, Sundqvist B (1996) Accuracy requirements for peptide characterization by monoisotopic molecular mass measurements. *Anal Chem* 68(22):4060–4063. doi:10.1021/ac9604651
22. Olsen JV, de Godoy LMF, Li GQ, Macek B, Mortensen P, Pesch R, Makarov A, Lange O, Horning S, Mann M (2005) Parts per million mass accuracy on an orbitrap mass spectrometer via lock mass injection into a C-trap. *Mol Cell Proteomics* 4(12):2010–2021. doi:10.1074/mcp.T500030-MCP200
23. Zhang Y, Wen Z, Washburn MP, Florens LA (2011) Improving proteomics mass accuracy by dynamic offline lock mass. *Anal Chem*. doi:10.1021/ac201867h
24. Papayannopoulos IA (1995) The interpretation of collision-induced dissociation tandem mass-spectra of peptides. *Mass Spectrom Rev* 14(1):49–73. doi:10.1002/mas.1280140104
25. Hardman M, Makarov AA (2003) Interfacing the orbitrap mass analyzer to an electrospray ion source. *Anal Chem* 75(7):1699–1705. doi:10.1021/ac0258047
26. Makarov A (2000) Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis. *Anal Chem* 72(6):1156–1162. doi:10.1021/ac991131p
27. Makarov A, Denisov E, Kholomeev A, Baischun W, Lange O, Strupat K, Horning S (2006) Performance evaluation of a hybrid linear ion trap/orbitrap mass spectrometer. *Anal Chem* 78(7):2113–2120. doi:10.1021/ac0518811
28. Olsen JV, Schwartz JC, Griep-Raming J, Nielsen ML, Damoc E, Denisov E, Lange O, Remes P, Taylor D, Splendore M, Wouters ER, Senko M, Makarov A, Mann M, Horning S (2009) A dual pressure linear ion trap orbitrap instrument with very high sequencing speed. *Mol Cell Proteomics* 8(12):2759–2769. doi:10.1074/mcp.M900375-MCP200
29. Makarov A, Denisov E, Lange O (2009) Performance evaluation of a high-field orbitrap mass analyzer. *J Am Soc Mass Spectrom* 20(8):1391–1396. doi:10.1016/j.jasms.2009.01.005
30. Michalski A, Damoc E, Hauschild JP, Lange O, Wiegand A, Makarov A, Nagaraj N, Cox J, Mann M, Horning S (2011) Mass spectrometry-based proteomics using Q exactive, a high-performance benchtop quadrupole orbitrap mass spectrometer. *Mol Cell Proteomics* 10(9). doi:10.1074/mcp.M111.011015
31. Andrews GL, Simons BL, Young JB, Hawkrige AM, Muddiman DC (2011) Performance characteristics of a New hybrid quadrupole time-of-flight tandem mass spectrometer (TripleTOF 5600). *Anal Chem* 83(13):5442–5446. doi:10.1021/ac200812d
32. Bahr R, Gerlich D, Telye E (1969) *Verhandl DPG* (VI) 4:343
33. Page JS, Tolmachev AV, Tang KQ, Smith RD (2006) Theoretical and experimental evaluation of the low m/z transmission of an electrodynamic ion funnel. *J Am Soc Mass Spectrom* 17(4):586–592. doi:10.1016/j.jasms.2005.12.013

34. Kim T, Tolmachev AV, Harkewicz R, Prior DC, Anderson G, Udseth HR, Smith RD, Bailey TH, Rakov S, Futrell JH (2000) Design and implementation of a new electrodynamic ion funnel. *Anal Chem* 72(10):2247–2255. doi:10.1021/ac991412x
35. Shaffer SA, Tang KQ, Anderson GA, Prior DC, Udseth HR, Smith RD (1997) A novel ion funnel for focusing ions at elevated pressure using electrospray ionization mass spectrometry. *Rapid Commun Mass Spectrom* 11(16):1813–1817
36. Kelly RT, Tolmachev AV, Page JS, Tang KQ, Smith RD (2010) The ion funnel: theory, implementations, and applications. *Mass Spectrom Rev* 29(2):294–312. doi:10.1002/mas.20232
37. Guan SH, Marshall AG (1996) Stacked-ring electrostatic ion guide. *J Am Soc Mass Spectrom* 7(1):101–106. doi:10.1016/1044-0305(95)00605-2
38. Kelly RT, Page JS, Marginean I, Tang KQ, Smith RD (2008) Nanoelectrospray emitter arrays providing interemitter electric field uniformity. *Anal Chem* 80(14):5660–5665. doi:10.1021/ac800508q
39. Page JS, Tang K, Kelly RT, Smith RD (2008) Subambient pressure ionization with nanoelectrospray source and interface for improved sensitivity in mass spectrometry. *Anal Chem* 80(5):1800–1805. doi:10.1021/ac702354b
40. Page JS, Kelly RT, Tang K, Smith RD (2007) Ionization and transmission efficiency in an electrospray ionization-mass spectrometry interface. *J Am Soc Mass Spectrom* 18(9):1582–1590. doi:10.1016/j.jasms.2007.05.018
41. Tang KQ, Page JS, Marginean I, Kelly RT, Smith RD (2011) Improving liquid chromatography-mass spectrometry sensitivity using a subambient pressure ionization with nanoelectrospray (SPIN) interface. *J Am Soc Mass Spectrom* 22(8):1318–1325. doi:10.1007/s13361-011-0135-7
42. Marginean I, Page JS, Tolmachev AV, Tang KQ, Smith RD (2010) Achieving 50% ionization efficiency in subambient pressure ionization with nanoelectrospray. *Anal Chem* 82(22):9344–9349. doi:10.1021/ac1019123
43. McLuckey SA, Mentinova M (2011) Ion/neutral, ion/electron, ion/photon, and ion/ion interactions in tandem mass spectrometry: do we need them all? Are they enough? *J Am Soc Mass Spectrom* 22(1):3–12. doi:10.1007/s13361-010-0004-9
44. McAlister GC, Phanstiel D, Wenger CD, Lee MV, Coon JJ (2010) Analysis of tandem mass spectra by FTMS for improved large-scale proteomics with superior protein quantification. *Anal Chem* 82(1):316–322. doi:10.1021/ac902005s
45. Louris JN, Cooks RG, Syka JEP, Kelley PE, Stafford GC, Todd JFJ (1987) Instrumentation, applications, and energy deposition in quadrupole ion-trap tandem mass-spectrometry. *Anal Chem* 59(13):1677–1685. doi:10.1021/ac00140a021
46. Ross PL, Huang YLN, Marchese JN, Williamson B, Parker K, Hattan S, Khainovski N, Pillai S, Dey S, Daniels S, Purkayastha S, Juhasz P, Martin S, Bartlet-Jones M, He F, Jacobson A, Pappin DJ (2004) Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Mol Cell Proteomics* 3(12):1154–1169. doi:10.1074/mcp.M400129-MCP200
47. Thompson A, Schafer J, Kuhn K, Kienle S, Schwarz J, Schmidt G, Neumann T, Hamon C (2003) Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal Chem* 75(8):1895–1904. doi:10.1021/ac0262560
48. Griffin TJ, Xie HW, Bandhakavi S, Popko J, Mohan A, Carlis JV, Higgins L (2007) iTRAQ reagent-based quantitative proteomic analysis on a linear ion trap mass spectrometer. *J Proteome Res* 6(11):4200–4209. doi:10.1021/pr070291b
49. Bantscheff M, Boesche M, Eberhard D, Matthieson T, Sweetman G, Kuster B (2008) Robust and sensitive iTRAQ quantification on an LTQ orbitrap mass spectrometer. *Mol Cell Proteomics* 7(9):1702–1713. doi:10.1074/mcp.M800029-MCP200
50. Pichler P, Kocher T, Holzmann J, Mohring T, Ammerer G, Mechtler K (2011) Improved precision of iTRAQ and TMT quantification by an axial extraction field in an orbitrap HCD cell. *Anal Chem* 83(4):1469–1474. doi:10.1021/ac102265w
51. Nagaraj N, D'Souza RCJ, Cox J, Olsen JV, Mann M (2010) Feasibility of large-scale phosphoproteomics with higher energy collisional dissociation fragmentation. *J Proteome Res* 9(12):6786–6794. doi:10.1021/pr100637q
52. Zubarev RA, Kelleher NL, McLafferty FW (1998) Electron capture dissociation of multiply charged protein cations. A nonergodic process. *J Am Chem Soc* 120(13):3265–3266. doi:10.1021/ja973478k
53. Bakhtiar R, Guan ZQ (2006) Electron capture dissociation mass spectrometry in characterization of peptides and proteins. *Biotechnol Lett* 28(14):1047–1059. doi:10.1007/s10529-006-9065-z
54. Coon JJ, Ueberheide B, Syka JEP, Dryhurst DD, Ausio J, Shabanowitz J, Hunt DF (2005) Protein identification using sequential ion/ion reactions and tandem mass spectrometry. *Proc Natl Acad Sci U S A* 102(27):9463–9468. doi:10.1073/pnas.0503189102
55. Syka JEP, Coon JJ, Schroeder MJ, Shabanowitz J, Hunt DF (2004) Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc Natl Acad Sci U S A* 101(26):9528–9533. doi:10.1073/pnas.0402700101
56. Mikesh LM, Ueberheide B, Chi A, Coon JJ, Syka JEP, Shabanowitz J, Hunt DF (2006) The utility of ETD mass spectrometry in proteomic analysis. *BBA-Proteins Proteomics* 1764(12):1811–1822. doi:10.1016/j.bbapap.2006.10.003
57. Zubarev RA (2003) Reactions of polypeptide ions with electrons in the gas phase. *Mass Spectrom Rev* 22(1):57–77. doi:10.1002/mas.10042
58. Wiesner J, Premisler T, Sickmann A (2008) Application of electron transfer dissociation (ETD) for the analysis of posttranslational

- modifications. *Proteomics* 8(21):4466–4483. doi:[10.1002/pmic.200800329](https://doi.org/10.1002/pmic.200800329)
59. An HJ, Froehlich JW, Lebrilla CB (2009) Determination of glycosylation sites and site-specific heterogeneity in glycoproteins. *Curr Opin Chem Biol* 13(4):421–426. doi:[10.1016/j.cbpa.2009.07.022](https://doi.org/10.1016/j.cbpa.2009.07.022)
60. Boersema PJ, Mohammed S, Heck AJR (2009) Phosphopeptide fragmentation and analysis by mass spectrometry. *J Mass Spectrom* 44(6):861–878. doi:[10.1002/jms.1599](https://doi.org/10.1002/jms.1599)
61. Schreiber TB, Mausbacher N, Breitkopf SB, Grundner-Culemann K, Daub H (2008) Quantitative phosphoproteomics: an emerging key technology in signal-transduction research. *Proteomics* 8(21):4416–4432. doi:[10.1002/pmic.200800132](https://doi.org/10.1002/pmic.200800132)
62. Crowe MC, Brodbelt JS (2004) Infrared multiphoton dissociation (IRMPD) and collisionally activated dissociation of peptides in a quadrupole ion trap with selective IRMPD of phosphopeptides. *J Am Soc Mass Spectrom* 15(11):1581–1592. doi:[10.1016/j.jasms.2004.07.016](https://doi.org/10.1016/j.jasms.2004.07.016)
63. Crowe MC, Brodbelt JS (2005) Differentiation of phosphorylated and unphosphorylated peptides by high-performance liquid chromatography-electrospray ionization-infrared multiphoton dissociation in a quadrupole ion trap. *Anal Chem* 77(17):5726–5734. doi:[10.1021/ac0509410](https://doi.org/10.1021/ac0509410)
64. Brodbelt JS, Wilson JJ (2009) Infrared multiphoton dissociation in quadrupole ion traps. *Mass Spectrom Rev* 28(3):390–424. doi:[10.1002/mas.20216](https://doi.org/10.1002/mas.20216)
65. Little DP, Speir JP, Senko MW, Oconnor PB, McLafferty FW (1994) Infrared multiphoton dissociation of large multiply-charged ions for biomolecule sequencing. *Anal Chem* 66(18):2809–2815. doi:[10.1021/ac00090a004](https://doi.org/10.1021/ac00090a004)
66. Ly T, Julian RR (2009) Ultraviolet photodissociation: developments towards applications for mass-spectrometry-based proteomics. *Angew Chem Int Ed* 48(39):7130–7137. doi:[10.1002/anie.200900613](https://doi.org/10.1002/anie.200900613)
67. Reilly JP (2009) Ultraviolet photofragmentation of biomolecular ions. *Mass Spectrom Rev* 28(3):425–447. doi:[10.1002/mas.20214](https://doi.org/10.1002/mas.20214)
68. Gatlin CL, Eng JK, Cross ST, Detter JC, Yates JR (2000) Automated identification of amino acid sequence variations in proteins by HPLC/microspray tandem mass spectrometry. *Anal Chem* 72(4):757–763. doi:[10.1021/ac991025n](https://doi.org/10.1021/ac991025n)
69. Masselon C, Anderson GA, Harkewicz R, Bruce JE, Pasa-Tolic L, Smith RD (2000) Accurate mass multiplexed tandem mass spectrometry for high-throughput polypeptide identification from mixtures. *Anal Chem* 72(8):1918–1924. doi:[10.1021/ac991133+](https://doi.org/10.1021/ac991133+)
70. Purvine S, Eppel JT, Yi EC, Goodlett DR (2003) Shotgun collision-induced dissociation of peptides using a time of flight mass analyzer. *Proteomics* 3(6):847–850. doi:[10.1002/pmic.200300362](https://doi.org/10.1002/pmic.200300362)
71. Silva JC, Denny R, Dorschel CA, Gorenstein M, Kass IJ, Li GZ, McKenna T, Nold MJ, Richardson K, Young P, Geromanos S (2005) Quantitative proteomic analysis by accurate mass retention time pairs. *Anal Chem* 77(7):2187–2200. doi:[10.1021/ac048455k](https://doi.org/10.1021/ac048455k)
72. Geiger T, Cox J, Mann M (2010) Proteomics on an orbitrap benchtop mass spectrometer using All-ion fragmentation. *Mol Cell Proteomics* 9(10):2252–2261. doi:[10.1074/mcp.M110.001537](https://doi.org/10.1074/mcp.M110.001537)
73. Li LJ, Masselon CD, Anderson GA, Pasa-Tolic L, Lee SW, Shen YF, Zhao R, Lipton MS, Conrads TP, Tolic N, Smith RD (2001) High-throughput peptide identification from protein digests using data-dependent multiplexed tandem FTICR mass spectrometry coupled with capillary liquid chromatography. *Anal Chem* 73(14):3312–3322. doi:[10.1021/ac010192w](https://doi.org/10.1021/ac010192w)
74. Panchoaud A, Scherl A, Shaffer SA, von Haller PD, Kulasekara HD, Miller SI, Goodlett DR (2009) Precursor acquisition independent from ion count: how to dive deeper into the proteomics ocean. *Anal Chem* 81(15):6481–6488. doi:[10.1021/ac900888s](https://doi.org/10.1021/ac900888s)
75. Venable JD, Dong MQ, Wohlschlegel J, Dillin A, Yates JR (2004) Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra. *Nat Methods* 1(1):39–45. doi:[10.1038/nmeth705](https://doi.org/10.1038/nmeth705)
76. Panchoaud A, Jung S, Shaffer SA, Aitchison JD, Goodlett DR (2011) Faster, quantitative, and accurate precursor acquisition independent from ion count. *Anal Chem* 83(6):2250–2257. doi:[10.1021/ac103079q](https://doi.org/10.1021/ac103079q)
77. Davis MT, Spahr CS, McGinley MD, Robinson JH, Bures EJ, Beierle J, Mort J, Yu W, Luethy R, Patterson SD (2001) Towards defining the urinary proteome using liquid chromatography-tandem mass spectrometry - II. Limitations of complex mixture analyses. *Proteomics* 1(1):108–117. doi:[10.1002/1615-9861\(200101\)1:1<108:aid-prot108>3.0.co;2-5](https://doi.org/10.1002/1615-9861(200101)1:1<108:aid-prot108>3.0.co;2-5)
78. Patterson SD, Spahr CS, Daugas E, Susin SA, Irinopoulou T, Koehler C, Kroemer G (2000) Mass spectrometric identification of proteins released from mitochondria undergoing permeability transition. *Cell Death Differ* 7(2):137–144. doi:[10.1038/sj.cdd.4400640](https://doi.org/10.1038/sj.cdd.4400640)
79. Spahr CS, Davis MT, McGinley MD, Robinson JH, Bures EJ, Beierle J, Mort J, Courchesne PL, Chen K, Wahl RC, Yu W, Luethy R, Patterson SD (2001) Towards defining the urinary proteome using liquid chromatography-tandem mass spectrometry I. Profiling an unfractionated tryptic digest. *Proteomics* 1(1):93–107. doi:[10.1002/1615-9861\(200101\)1:1<93:aid-prot93>3.0.co;2-3](https://doi.org/10.1002/1615-9861(200101)1:1<93:aid-prot93>3.0.co;2-3)
80. Yi EC, Marelli M, Lee H, Purvine SO, Aebersold R, Aitchison JD, Goodlett DR (2002) Approaching complete peroxisome characterization by gas-phase fractionation. *Electrophoresis* 23(18):3205–3216. doi:[10.1002/1522-2683\(200209\)23:18<3205:aid-elps3205>3.0.co;2-y](https://doi.org/10.1002/1522-2683(200209)23:18<3205:aid-elps3205>3.0.co;2-y)
81. Scherl A, Shaffer SA, Taylor GK, Kulasekara HD, Miller SI, Goodlett DR (2008) Genome-specific gas-phase fractionation strategy for improved

- shotgun proteomic profiling of proteotypic peptides. *Anal Chem* 80(4):1182–1191. doi:[10.1021/ac701680f](https://doi.org/10.1021/ac701680f)
82. Harvey SR, MacPhee CE, Barran PE (2011) Ion mobility mass spectrometry for peptide analysis. *Methods* 54(4):454–461. doi:[10.1016/j.ymeth.2011.05.004](https://doi.org/10.1016/j.ymeth.2011.05.004)
83. Valentine SJ, Kulchania M, Barnes CAS, Clemmer DE (2001) Multidimensional separations of complex peptide mixtures: a combined high-performance liquid chromatography/ion mobility/time-of-flight mass spectrometry approach. *Int J Mass Spectrom* 212(1–3):97–109. doi:[10.1016/s1387-3806\(01\)00511-5](https://doi.org/10.1016/s1387-3806(01)00511-5)
84. Srebalus CA, Li JW, Marshall WS, Clemmer DE (1999) Gas phase separations of electrosprayed peptide libraries. *Anal Chem* 71(18):3918–3927. doi:[10.1021/ac9903757](https://doi.org/10.1021/ac9903757)
85. Shvartsburg AA, Danielson WF, Smith RD (2010) High-resolution differential ion mobility separations using helium-rich gases. *Anal Chem* 82(6):2456–2462. doi:[10.1021/ac902852a](https://doi.org/10.1021/ac902852a)
86. Shvartsburg AA, Tang KQ, Smith RD (2010) Differential ion mobility separations of peptides with resolving power exceeding 50. *Anal Chem* 82(1):32–35. doi:[10.1021/ac902133n](https://doi.org/10.1021/ac902133n)
87. Shvartsburg AA, Prior DC, Tang KQ, Smith RD (2010) High-resolution differential ion mobility separations using planar analyzers at elevated dispersion fields. *Anal Chem* 82(18):7649–7655. doi:[10.1021/ac101413k](https://doi.org/10.1021/ac101413k)
88. Shvartsburg AA, Li FM, Tang KQ, Smith RD (2006) High-resolution field asymmetric waveform ion mobility spectrometry using new planar geometry analyzers. *Anal Chem* 78(11):3706–3714. doi:[10.1021/ac052020v](https://doi.org/10.1021/ac052020v)
89. Shvartsburg AA, Singer D, Smith RD, Hoffmann R (2011) Ion mobility separation of isomeric phosphopeptides from a protein with variant modification of adjacent residues. *Anal Chem* 83(13):5078–5085. doi:[10.1021/ac200985s](https://doi.org/10.1021/ac200985s)
90. Giles K, Pringle SD, Worthington KR, Little D, Wildgoose JL, Bateman RH (2004) Applications of a travelling wave-based radio-frequency-only stacked ring ion guide. *Rapid Commun Mass Spectrom* 18(20):2401–2414. doi:[10.1002/rcm.1641](https://doi.org/10.1002/rcm.1641)
91. Pringle SD, Giles K, Wildgoose JL, Williams JP, Slade SE, Thalassinou K, Bateman RH, Bowers MT, Scrivens JH (2007) An investigation of the mobility separation of some peptide and protein ions using a new hybrid quadrupole/travelling wave IMS/oat-ToF instrument. *Int J Mass Spectrom* 261(1):1–12. doi:[10.1016/j.ijms.2006.07.021](https://doi.org/10.1016/j.ijms.2006.07.021)
92. Schmidt A, Claassen M, Aebersold R (2009) Directed mass spectrometry: towards hypothesis-driven proteomics. *Curr Opin Chem Biol* 13(5–6):510–517. doi:[10.1016/j.cbpa.2009.08.016](https://doi.org/10.1016/j.cbpa.2009.08.016)
93. Paulovich AG, Whiteaker JR, Hoofnagle AN, Wang P (2008) The interface between biomarker discovery and clinical validation: the tar pit of the protein biomarker pipeline. *Proteomics Clin Appl* 2(10–11):1386–1402. doi:[10.1002/prca.200780174](https://doi.org/10.1002/prca.200780174)
94. Picotti P, Bodenmiller B, Mueller LN, Domon B, Aebersold R (2009) Full dynamic range proteome analysis of *S. Cerevisiae* by targeted proteomics. *Cell* 138(4):795–806. doi:[10.1016/j.cell.2009.05.051](https://doi.org/10.1016/j.cell.2009.05.051)
95. Fusaro VA, Mani DR, Mesirov JP, Carr SA (2009) Prediction of high-responding peptides for targeted protein assays by mass spectrometry. *Nat Biotechnol* 27(2):190–198. doi:[10.1038/nbt.1524](https://doi.org/10.1038/nbt.1524)
96. Mallick P, Schirle M, Chen SS, Flory MR, Lee H, Martin D, Raught B, Schmitt R, Werner T, Kuster B, Aebersold R (2007) Computational prediction of proteotypic peptides for quantitative proteomics. *Nat Biotechnol* 25(1):125–131. doi:[10.1038/nbt1275](https://doi.org/10.1038/nbt1275)
97. Deutsch EW, Lam H, Aebersold R (2008) PeptideAtlas: a resource for target selection for emerging targeted proteomics workflows. *EMBO Rep* 9(5):429–434. doi:[10.1038/embor.2008.56](https://doi.org/10.1038/embor.2008.56)
98. Farrah T, Deutsch EW, Omenn GS, Campbell DS, Sun Z, Bletz JA, Mallick P, Katz JE, Malmstrom J, Ossola R, Watts JD, Lin BAY, Zhang H, Moritz RL, Aebersold R (2011) A high-confidence human plasma proteome reference set with estimated concentrations in PeptideAtlas. *Mol Cell Proteomics* 10(9). doi:[10.1074/mcp.M110.006353](https://doi.org/10.1074/mcp.M110.006353)
99. Kiyonami R, Schoen A, Prakash A, Peterman S, Zabrouskov V, Picotti P, Aebersold R, Huhmer A, Domon B (2011) Increased selectivity, analytical precision, and throughput in targeted proteomics. *Mol Cell Proteomics* 10(2). doi:[10.1074/mcp.M110.002931](https://doi.org/10.1074/mcp.M110.002931)
100. Addona TA, Abbatello SE, Schilling B, Skates SJ, Mani DR, Bunk DM, Spiegelman CH, Zimmerman LJ, Ham AJ, Keshishian H, Hall SC, Allen S, Blackman RK, Borchers CH, Buck C, Cardasis HL, Cusack MP, Dodder NG, Gibson BW, Held JM, Hiltke T, Jackson A, Johansen EB, Kinsinger CR, Li J, Mesri M, Neubert TA, Niles RK, Pulsipher TC, Ransohoff D, Rodriguez H, Rudnick PA, Smith D, Tabb DL, Tegeler TJ, Variyath AM, Vega-Montoto LJ, Wahlander A, Waldemarson S, Wang M, Whiteaker JR, Zhao L, Anderson NL, Fisher SJ, Liebler DC, Paulovich AG, Regnier FE, Tempst P, Carr SA (2009) Multi-site assessment of the precision and reproducibility of multiple reaction monitoring-based measurements of proteins in plasma. *Nat Biotechnol* 27(7):633–641. doi:[nbt.1546](https://doi.org/10.1038/nbt.1546) [pii] [10.1038/nbt.1546](https://doi.org/10.1038/nbt.1546)
101. Calvo E, Camafeita E, Fernandez-Gutierrez B, Lopez JA (2011) Applying selected reaction monitoring to targeted proteomics. *Expert Rev Proteomics* 8(2):165–173. doi:[10.1586/epr.11.11](https://doi.org/10.1586/epr.11.11)
102. Chiu CL, Randall S, Molloy MP (2009) Recent progress in selected reaction monitoring MS-driven plasma protein biomarker analysis. *Bioanalysis* 1(4):847–855. doi:[10.4155/bio.09.56](https://doi.org/10.4155/bio.09.56)

103. Elschenbroich S, Kislinger T (2011) Targeted proteomics by selected reaction monitoring mass spectrometry: applications to systems biology and biomarker discovery. *Mol Biosyst* 7(2):292–303. doi:[10.1039/c0mb00159g](https://doi.org/10.1039/c0mb00159g)
104. Surinova S, Schiess R, Huttenhain R, Cerciello F, Wollscheid B, Aebersold R (2011) On the development of plasma protein biomarkers. *J Proteome Res* 10(1):5–16. doi:[10.1021/pr1008515](https://doi.org/10.1021/pr1008515)
105. Martin DB, Holzman T, May D, Peterson A, Eastham A, Eng J, McIntosh M (2008) MRMer, an interactive open source and cross-platform system for data extraction and visualization of multiple reaction monitoring experiments. *Mol Cell Proteomics* 7(11):2270–2278. doi:[10.1074/mcp.M700504-MCP200](https://doi.org/10.1074/mcp.M700504-MCP200)
106. Mead JA, Bianco L, Ottone V, Barton C, Kay RG, Lilley KS, Bond NJ, Bessant C (2009) MRMAid, the web-based tool for designing multiple reaction monitoring (MRM) transitions. *Mol Cell Proteomics* 8(4):696–705. doi:[10.1074/mcp.M800192-MCP200](https://doi.org/10.1074/mcp.M800192-MCP200)
107. Sherwood CA, Eastham A, Lee LW, Peterson A, Eng JK, Shteynberg D, Mendoza L, Deutsch EW, Risler J, Tasman N, Aebersold R, Lam H, Martin DB (2009) MaRiMba: a software application for spectral library-based MRM transition list assembly. *J Proteome Res* 8(10):4396–4405. doi:[10.1021/pr900010h](https://doi.org/10.1021/pr900010h)
108. MacLean B, Tomazela DM, Shulman N, Chambers M, Finney GL, Frewen B, Kern R, Tabb DL, Liebler DC, MacCoss MJ (2010) Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* 26(7):966–968. doi:[10.1093/bioinformatics/btq054](https://doi.org/10.1093/bioinformatics/btq054)
109. MacLean B, Tomazela DM, Abbatiello SE, Zhang SC, Whiteaker JR, Paulovich AG, Carr SA, MacCoss MJ (2010) Effect of collision energy optimization on the measurement of peptides by selected reaction monitoring (SRM) mass spectrometry. *Anal Chem* 82(24):10116–10124. doi:[10.1021/ac102179j](https://doi.org/10.1021/ac102179j)
110. Prakash A, Tomazela DM, Frewen B, MacLean B, Merrihew G, Peterman S, MacCoss MJ (2009) Expediting the development of targeted SRM assays: using data from shotgun proteomics to automate method development. *J Proteome Res* 8(6):2733–2739. doi:[10.1021/pr801028b](https://doi.org/10.1021/pr801028b)
111. Chait BT, Wang R, Beavis RC, Kent SBH (1993) Protein ladder sequencing. *Science* 262(5130):89–92
112. Cruz-Marcelo A, Guerra R, Vannucci M, Li Y, Lau CC, Man TK (2008) Comparison of algorithms for pre-processing of SELDI-TOF mass spectrometry data. *Bioinformatics* 24(19):2129–2136. doi:[10.1093/bioinformatics/btn398](https://doi.org/10.1093/bioinformatics/btn398) [pii]
113. Roy P, Truntzer C, Maucourt-Boulch D, Jouve T, Molinari N (2011) Protein mass spectra data analysis for clinical biomarker discovery: a global review. *Brief Bioinform* 12(2):176–186. doi:[10.1093/bib/bbq019](https://doi.org/10.1093/bib/bbq019) [pii]
114. Sellers KF, Miecznikowski JC (2010) Feature detection techniques for preprocessing proteomic data. *Int J Biomed Imaging* 2010:896718. doi:[10.1155/2010/896718](https://doi.org/10.1155/2010/896718)
115. Wegdam W, Moerland PD, Buist MR, Loren V, van Themaat E, Bleijlevens B, Hoefsloot HC, de Koster CG, Aerts JM (2009) Classification-based comparison of pre-processing methods for interpretation of mass spectrometry generated clinical datasets. *Proteome Sci* 7:19. doi:[1477-5956-7-19](https://doi.org/10.1186/1477-5956-7-19) [pii] [10.1186/1477-5956-7-19](https://doi.org/10.1186/1477-5956-7-19)
116. Addona T, Clauser K (2002) De Novo Peptide De Novo Peptide Sequencing via Manual Interpretation of MS/MS Spectra. *Curr Protoc Protein Sci* 16.11.1–16.11.19
117. Craig R, Beavis RC (2004) TANDEM: matching proteins with tandem mass spectra. *Bioinformatics* 20(9):1466–1467. doi:[10.1093/bioinformatics/bth092](https://doi.org/10.1093/bioinformatics/bth092)
118. Geer LY, Markey SP, Kowalak JA, Wagner L, Xu M, Maynard DM, Yang X, Shi W, Bryant SH (2004) Open mass spectrometry search algorithm. *J Proteome Res* 3(5):958–964. doi:[10.1021/pr0499491](https://doi.org/10.1021/pr0499491)
119. Cox J, Neuhauser N, Michalski A, Scheltema RA, Olsen JV, Mann M (2011) Andromeda: a peptide search engine integrated into the MaxQuant environment. *J Proteome Res* 10(4):1794–1805. doi:[10.1021/Pr101065j](https://doi.org/10.1021/Pr101065j)
120. Tabb DL, Fernando CG, Chambers MC (2007) MyriMatch: highly accurate tandem mass spectral peptide identification by multivariate hypergeometric analysis. *J Proteome Res* 6(2):654–661. doi:[10.1021/Pr0604054](https://doi.org/10.1021/Pr0604054)
121. Clauser KR, Baker P, Burlingame AL (1999) Role of accurate mass measurement (+/– 10 ppm) in protein identification strategies employing MS or MS MS and database searching. *Anal Chem* 71(14):2871–2882
122. Colinge J, Masselot A, Giron M, Dessingy T, Magnin J (2003) OLAV: towards high-throughput tandem mass spectrometry data identification. *Proteomics* 3(8):1454–1463. doi:[10.1002/pmic.200300485](https://doi.org/10.1002/pmic.200300485)
123. Casado-Vela J (2011) Lights and shadows of proteomic technologies for the study of protein species including isoforms, splicing variants and protein post-translational modifications (vol 11, pg 590, 2011). *Proteomics* 11(7):1370–1370
124. Nilsson T, Mann M, Aebersold R, Yates JR, Bairoch A, Bergeron JJM (2010) Mass spectrometry in high-throughput proteomics: ready for the big time. *Nat Methods* 7(9):681–685. doi:[10.1038/nmeth0910-681](https://doi.org/10.1038/nmeth0910-681)
125. Searle BC, Turner M, Nesvizhskii AI (2008) Improving sensitivity by probabilistically combining results from multiple MS/MS search methodologies. *J Proteome Res* 7(1):245–253. doi:[10.1021/Pr070540w](https://doi.org/10.1021/Pr070540w)
126. Resing KA, Meyer-Arendt K, Mendoza AM, Aveline-Wolf LD, Jonscher KR, Pierce KG, Old

- WM, Cheung HT, Russell S, Wattawa JL, Goehle GR, Knight RD, Ahn NG (2004) Improving reproducibility and sensitivity in identifying human proteins by shotgun proteomics. *Anal Chem* 76(13):3556–3568
127. Alves G, Wu WW, Wang GH, Shen RF, Yu YK (2008) Enhancing peptide identification confidence by combining search methods. *J Proteome Res* 7(8):3102–3113. doi:[10.1021/Pr700798h](https://doi.org/10.1021/Pr700798h)
128. Searle BC, Turner M (2006) Improving computer interpretation of linear ion trap proteomics data using Scaffold. *Mol Cell Proteomics* 5(10):S297–S297
129. Searle BC (2010) Scaffold: a bioinformatic tool for validating MS/MS-based proteomic studies. *Proteomics* 10(6):1265–1269. doi:[10.1002/pmic.200900437](https://doi.org/10.1002/pmic.200900437)
130. Kwon T, Choi H, Vogel C, Nesvizhskii AI, Marcotte EM (2011) MSblender: a probabilistic approach for integrating peptide identifications from multiple database search engines. *J Proteome Res* 10(7):2949–2958. doi:[10.1021/Pr2002116](https://doi.org/10.1021/Pr2002116)
131. Yates JR, Morgan SF, Gatlin CL, Griffin PR, Eng JK (1998) Method to compare collision-induced dissociation spectra of peptides: potential for library searching and subtractive analysis. *Anal Chem* 70(17):3557–3565
132. Lam H, Deutsch EW, Eddes JS, Eng JK, King N, Stein SE, Aebersold R (2007) Development and validation of a spectral library searching method for peptide identification from MS/MS. *Proteomics* 7(5):655–667
133. Lam H, Deutsch EW, Eddes JS, Eng JK, Stein SE, Aebersold R (2008) Building consensus spectral libraries for peptide identification in proteomics. *Nat Methods* 5(10):873–875. doi:[10.1038/Nmeth.1254](https://doi.org/10.1038/Nmeth.1254)
134. Frewen BE, Merrihew GE, Wu CC, Noble WS, MacCoss MJ (2006) Analysis of peptide MS/MS spectra from large-scale proteomics experiments using spectrum libraries. *Anal Chem* 78(16):5678–5684. doi:[10.1021/Ac060279n](https://doi.org/10.1021/Ac060279n)
135. Craig R, Cortens JC, Fenyo D, Beavis RC (2006) Using annotated peptide mass spectrum libraries for protein identification. *J Proteome Res* 5(8):1843–1849. doi:[10.1021/Pr0602085](https://doi.org/10.1021/Pr0602085)
136. Hummel J, Niemann M, Wienkoop S, Schulze W, Steinhäuser D, Selbig J, Walther D, Weckwerth W (2007) ProMEX: a mass spectral reference database for proteins and protein phosphorylation sites. *BMC Bioinformatics* 8. doi:[Artn 216 Doi 10.1186/1471-2105-8-216](https://doi.org/10.1186/1471-2105-8-216)
137. Wu X, Tseng CW, Edwards N (2007) HMMatch: peptide identification by spectral matching of tandem mass spectra using hidden Markov models. *J Comput Biol* 14(8):1025–1043. doi:[10.1089/cmb.2007.0071](https://doi.org/10.1089/cmb.2007.0071)
138. Bodenmiller B, Campbell D, Gerrits B, Lam H, Jovanovic M, Picotti P, Schlapbach R, Aebersold R (2008) PhosphoPep—a database of protein phosphorylation sites in model organisms. *Nat Biotechnol* 26(12):1339–1340. doi:[10.1038/Nbt1208-1339](https://doi.org/10.1038/Nbt1208-1339)
139. Srikumar T, Jeram SM, Lam H, Raught B (2010) A ubiquitin and ubiquitin-like protein spectral library. *Proteomics* 10(2):337–342. doi:[10.1002/pmic.200900627](https://doi.org/10.1002/pmic.200900627)
140. Desiere F, Deutsch EW, King NL, Nesvizhskii AI, Mallick P, Eng J, Chen S, Eddes J, Loevenich SN, Aebersold R (2006) The PeptideAtlas project. *Nucleic Acids Res* 34(Database issue):D655–D658
141. Vizcaino JA, Cote R, Reisinger F, Foster JM, Mueller M, Rameseder J, Hermjakob H, Martens L (2009) A guide to the proteomics identifications database proteomics data repository. *Proteomics* 9(18):4276–4283. doi:[10.1002/pmic.200900402](https://doi.org/10.1002/pmic.200900402)
142. Jones P, Cote RG, Martens L, Quinn AF, Taylor CF, Derache W, Hermjakob H, Apweiler R (2006) PRIDE: a public repository of protein and peptide identifications for the proteomics community. *Nucleic Acids Res* 34:D659–D663. doi:[10.1093/Nar/Gkj138](https://doi.org/10.1093/Nar/Gkj138)
143. Nesvizhskii AI (2010) A survey of computational methods and error rate estimation procedures for peptide and protein identification in shotgun proteomics. *J Proteomics* 73(11):2092–2123. doi:[S1874-3919\(10\)00249-6 \[pii\] 10.1016/j.jprot.2010.08.009](https://doi.org/10.1016/j.jprot.2010.08.009)
144. Frank AM (2009) Predicting intensity ranks of peptide fragment ions. *J Proteome Res* 8(5):2226–2240. doi:[10.1021/Pr800677f](https://doi.org/10.1021/Pr800677f)
145. Cox J, Mann M (2009) Computational principles of determining and improving mass precision and accuracy for proteome measurements in an orbitrap. *J Am Soc Mass Spectrom* 20(8):1477–1485. doi:[10.1016/j.jasms.2009.05.007](https://doi.org/10.1016/j.jasms.2009.05.007)
146. Hamm CW, Wilson WE, Harvan DJ (1986) Peptide sequencing program. *Comput Appl Biosci* 2(2):115–118
147. Ishikawa K, Niwa Y (1986) Computer-aided peptide sequencing by fast-atom-bombardment mass-spectrometry. *Biomed Environ Mass* 13(7):373–380
148. Sakurai T, Matsuo T, Matsuda H, Katakuse I (1984) Paas-3: a computer-program to determine probable sequence of peptides from mass-spectrometric data. *Biomed Mass Spectrom* 11(8):396–399
149. Scoble HA, Biller JE, Biemann K (1987) A graphics display-oriented strategy for the amino-acid sequencing of peptides by tandem mass-spectrometry. *Fresen Z Anal Chem* 327(2):239–245
150. Siegel MM, Bauman N (1988) An efficient algorithm for sequencing peptides using fast atom bombardment mass-spectral data. *Biomed Environ Mass* 15(6):333–343
151. Bartels C (1990) Fast algorithm for peptide sequencing by mass-spectroscopy. *Biomed Environ Mass* 19(6):363–368
152. Taylor JA, Johnson RS (2001) Implementation and uses of automated de novo peptide sequencing by tandem mass spectrometry. *Anal Chem* 73(11):2594–2604

153. Dancik V, Addona TA, Clauser KR, Vath JE, Pevzner PA (1999) De novo peptide sequencing via tandem mass spectrometry. *J Comput Biol* 6(3–4):327–342
154. Chen T, Kao MY, Tepel M, Rush J, Church GM (2001) A dynamic programming approach to de novo peptide sequencing via tandem mass spectrometry. *J Comput Biol* 8(3):325–337
155. Frank A, Pevzner P (2005) PepNovo: De novo peptide sequencing via probabilistic network modeling. *Anal Chem* 77(4):964–973. doi:[10.1021/Ac048788h](https://doi.org/10.1021/Ac048788h)
156. Fischer B, Roth V, Roos F, Grossmann J, Baginsky S, Widmayer P, Gruissem W, Buhmann JM (2005) NovoHMM: a hidden Markov model for de novo peptide sequencing. *Anal Chem* 77(22):7265–7273. doi:[10.1021/Ac0508853](https://doi.org/10.1021/Ac0508853)
157. Tabb DL, Ma ZQ, Martin DB, Ham AJL, Chambers MC (2008) DirecTag: accurate sequence tags from peptide MS/MS through statistical scoring. *J Proteome Res* 7(9):3838–3846. doi:[10.1021/Pr800154p](https://doi.org/10.1021/Pr800154p)
158. Ma B, Zhang K, Hendrie C, Liang C, Li M, Kirby AD, Lajoie G (2003) Peaks: powerful software for peptide de novo sequencing by tandem mass spectrometry. *Rapid Commun Mass Spectrom* 17:2337–2342
159. Chi H, Sun RX, Yang B, Song CQ, Wang LH, Liu C, Fu Y, Yuan ZF, Wang HP, He SM, Dong MQ (2010) PNovo: De novo peptide sequencing and identification using HCD spectra. *J Proteome Res* 9(5):2713–2724. doi:[10.1021/Pr100182k](https://doi.org/10.1021/Pr100182k)
160. Zhang ZQ, McElvain JS (2000) De novo peptide sequencing by two dimensional fragment correlation mass spectrometry. *Anal Chem* 72(11):2337–2350
161. Savitski MM, Nielsen ML, Kjeldsen F, Zubarev RA (2005) Proteomics-grade de novo sequencing approach. *J Proteome Res* 4(6):2348–2354. doi:[10.1021/pr050288x](https://doi.org/10.1021/pr050288x)
162. Bandeira N, Tsur D, Frank A, Pevzner PA (2007) Protein identification by spectral networks analysis. *Proc Natl Acad Sci U S A* 104(15):6140–6145. doi:[10.1073/pnas.0701130104](https://doi.org/10.1073/pnas.0701130104)
163. Datta R, Bern M (2009) Spectrum fusion: using multiple mass spectra for de novo peptide sequencing. *J Comput Biol* 16(8):1169–1182. doi:[10.1089/cmb.2009.0122](https://doi.org/10.1089/cmb.2009.0122)
164. Mann M, Wilm M (1994) Error tolerant identification of peptides in sequence databases by peptide sequence tags. *Anal Chem* 66(24):4390–4399
165. Tabb DL, Saraf A, Yates JR (2003) GutenTag: high-throughput sequence tagging via an empirically derived fragmentation model. *Anal Chem* 75(23):6415–6421. doi:[10.1021/Ac0347462](https://doi.org/10.1021/Ac0347462)
166. Tanner S, Shu HJ, Frank A, Wang LC, Zandi E, Mumby M, Pevzner PA, Bafna V (2005) InsPecT: identification of posttranslational modified peptides from tandem mass spectra. *Anal Chem* 77(14):4626–4639. doi:[10.1021/Ac050102d](https://doi.org/10.1021/Ac050102d)
167. Dasari S, Chambers MC, Slebos RJ, Zimmerman LJ, Ham AJL, Tabb DL (2010) TagRecon: high-throughput mutation identification through sequence tagging. *J Proteome Res* 9(4):1716–1726. doi:[10.1021/pr900850m](https://doi.org/10.1021/pr900850m)
168. Shilov IV, Seymour SL, Patel AA, Loboda A, Tang WH, Keating SP, Hunter CL, Nuwaysir LM, Schaeffer DA (2007) The paragon algorithm, a next generation search engine that uses sequence temperature values and feature probabilities to identify peptides from tandem mass spectra. *Mol Cell Proteomics* 6(9):1638–1655. doi:[10.1074/mcp.T600050-MCP200](https://doi.org/10.1074/mcp.T600050-MCP200)
169. Nesvizhskii AI, Aebersold R (2005) Interpretation of shotgun proteomic data: the protein inference problem. *Mol Cell Proteomics* 4(10):1419–1440
170. Nesvizhskii AI, Aebersold R (2004) Analysis, statistical validation and dissemination of large-scale proteomics datasets generated by tandem MS. *Drug Discov Today* 9(4):173–181. doi:[10.1016/S1359-6446\(03\)02978-7](https://doi.org/10.1016/S1359-6446(03)02978-7) [S1359644603029787 \[pii\]](https://doi.org/10.1016/S1359-6446(03)02978-7)
171. States DJ, Omenn GS, Blackwell TW, Fermin D, Eng J, Speicher DW, Hanash SM (2006) Challenges in deriving high-confidence protein identifications from data gathered by a HUPO plasma proteome collaborative study. *Nat Biotechnol* 24(3):333–338. doi:[10.1038/Nbt1183](https://doi.org/10.1038/Nbt1183)
172. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate—a practical and powerful approach to multiple testing. *J Roy Stat Soc B Met* 57(1):289–300
173. Keller A, Nesvizhskii AI, Kolker E, Aebersold R (2002) Empirical statistical model to estimate the accuracy of peptide identifications made by ms/ms and database search. *Anal Chem* 74(20):5383–5392
174. Bianco L, Mead JA, Bessant C (2009) Comparison of novel decoy database designs for optimizing protein identification searches using ABRAF sPRG2006 standard MS/MS data sets. *J Proteome Res* 8(4):1782–1791. doi:[10.1021/Pr800792z](https://doi.org/10.1021/Pr800792z)
175. Kall L, Storey JD, MacCoss MJ, Noble WS (2008) Assigning significance to peptides identified by tandem mass spectrometry using decoy databases. *J Proteome Res* 7(1):29–34. doi:[10.1021/pr700600n](https://doi.org/10.1021/pr700600n)
176. Storey JD, Tibshirani R (2003) Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A* 100(16):9440–9445. doi:[10.1073/pnas.1530509100](https://doi.org/10.1073/pnas.1530509100) [1530509100 \[pii\]](https://doi.org/10.1073/pnas.1530509100)
177. Qian W-J, Liu T, Monroe ME, Strittmatter EF, Jacobs JM, Kangas LJ, Petritis K, Camp DG, Smith RD (2005) Probability-based evaluation of peptide and protein identifications from tandem mass spectrometry and SEQUEST analysis: the human proteome. *J Proteome Res* 4(1):53–62
178. Moore RE, Young MK, Lee TD (2002) Qscore: an algorithm for evaluating SEQUEST database search results. *J Am Soc Mass Spectrom* 13(4):378–386
179. Nesvizhskii AI, Keller A, Kolker E, Aebersold R (2003) A statistical model for identifying proteins by tandem mass spectrometry. *Anal Chem* 75(17):4646–4658

180. Ma ZQ, Dasari S, Chambers MC, Litton MD, Sobecki SM, Zimmerman LJ, Halvey PJ, Schilling B, Drake PM, Gibson BW, Tabb DL (2009) ID-Picker 2.0: improved protein assembly with high discrimination peptide identification filtering. *J Proteome Res* 8(8):3872–3881. doi:[10.1021/pr900360j](https://doi.org/10.1021/pr900360j)
181. Kim S, Mischerikow N, Bandeira N, Navarro JD, Wich L, Mohammed S, Heck AJ, Pevzner PA (2010) The generating function of CID, ETD and CID/ETD pairs of tandem mass spectra: applications to database search. *Mol Cell Proteomics*. doi:[M110.003731 \[pii\] 10.1074/mcp.M110.003731](https://doi.org/10.1074/mcp.M110.003731)
182. Fenyo D, Beavis RC (2003) A method for assessing the statistical significance of mass spectrometry-based protein identifications using general scoring schemes. *Anal Chem* 75(4):768–774. doi:[10.1021/Ac0258709](https://doi.org/10.1021/Ac0258709)
183. Taylor CF (2006) A capital workshop for the HUPO proteomics standards initiative. *J Proteome Res* 5(12):3229–3230
184. Hermjakob H (2006) The HUPO proteomics standards initiative—overcoming the fragmentation of proteomics data. *Proteomics* 6(1):34–38. doi:[10.1002/pmic.200600537](https://doi.org/10.1002/pmic.200600537)
185. Taylor CF, Hermjakob H, Julian RK, Garavelli JS, Aebersold R, Apweiler R (2006) The work of the human proteome organisation’s proteomics standards initiative (HUPO PSI). *Omics* 10(2):145–151
186. Orchard S, Kersey P, Hermjakob H, Apweiler R (2003) Meeting review: The HUPO proteomics standards initiative meeting: towards common standards for exchanging proteomics data—Hinxton, Cambridge, UK, 19–20 October 2002. *Comp Funct Genom* 4(1):16–19. doi:[10.1002/Cfg.232](https://doi.org/10.1002/Cfg.232)
187. Orchard S, Kersey P, Zhu WM, Montecchi-Palazzi L, Hermjakob H, Apweiler R (2003) Meeting review: progress in establishing common standards for exchanging proteomics data: the second meeting of the HUPO proteomics standards initiative. *Comp Funct Genom* 4(2):203–206. doi:[10.1002/Cfg.279](https://doi.org/10.1002/Cfg.279)
188. Bradshaw RA, Burlingame AL, Carr S, Aebersold R (2006) Reporting protein identification data—the next generation of guidelines. *Mol Cell Proteomics* 5(5):787–788
189. Becker GW (2008) Stable isotopic labeling of proteins for quantitative proteomic applications. *Brief Funct Genomic Proteomic* 7(5):371–382. doi:[eln047 \[pii\] 10.1093/bfgp/eln047](https://doi.org/10.1093/bfgp/eln047)
190. Gevaert K, Impens F, Ghesquiere B, Van Damme P, Lambrechts A, Vandekerckhove J (2008) Stable isotopic labeling in proteomics. *Proteomics* 8(23–24):4873–4885. doi:[10.1002/pmic.200800421](https://doi.org/10.1002/pmic.200800421)
191. Oda Y, Huang K, Cross FR, Cowburn D, Chait BT (1999) Accurate quantitation of protein expression and site-specific phosphorylation. *Proc Natl Acad Sci U S A* 96(12):6591–6596
192. Ong SE, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, Mann M (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
193. Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol* 17(10):994–999
194. Dayon L, Hainard A, Licker V, Turck N, Kuhn K, Hochstrasser DF, Burkhard PR, Sanchez JC (2008) Relative quantification of proteins in human cerebrospinal fluids by MS/MS using 6-plex isobaric tags. *Anal Chem* 80(8):2921–2931. doi:[10.1021/ac702422x](https://doi.org/10.1021/ac702422x)
195. Ross PL, Huang YN, Marchese JN, Williamson B, Parker K, Hattan S, Khainovski N, Pillai S, Dey S, Daniels S, Purkayastha S, Juhasz P, Martin S, Bartlett-Jones M, He F, Jacobson A, Pappin DJ (2004) Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Mol Cell Proteomics* 3(12):1154–1169
196. Reynolds KJ, Fenselau C (2004) Quantitative protein analysis using proteolytic [¹⁸O] water labeling. *Curr Protoc Protein Sci* 23:23–24. doi:[10.1002/0471140864.ps2304s34](https://doi.org/10.1002/0471140864.ps2304s34)
197. Ow SY, Cardona T, Taton A, Magnuson A, Lindblad P, Stensjo K, Wright PC (2008) Quantitative shotgun proteomics of enriched heterocysts from *Nostoc* sp. PCC 7120 using 8-plex isobaric peptide tags. *J Proteome Res* 7(4):1615–1628. doi:[10.1021/pr700604v](https://doi.org/10.1021/pr700604v)
198. Wu CC, MacCoss MJ, Howell KE, Matthews DE, Yates JR 3rd (2004) Metabolic labeling of mammalian organisms with stable isotopes for quantitative proteomic analysis. *Anal Chem* 76(17):4951–4959
199. Savitski MM, Sweetman G, Askenazi M, Marto JA, Lang M, Zinn N, Bantscheff M (2011) Delayed fragmentation and optimized isolation width settings for improvement of protein identification and accuracy of isobaric mass tag quantification on orbitrap-type mass spectrometers. *Anal Chem* 83(23):8959–8967. doi:[10.1021/ac201760x](https://doi.org/10.1021/ac201760x)
200. Neilson KA, Ali NA, Muralidharan S, Mirzaei M, Mariani M, Assadourian G, Lee A, van Sluyter SC, Haynes PA (2011) Less label, more free: approaches in label-free quantitative mass spectrometry. *Proteomics* 11(4):535–553. doi:[10.1002/pmic.201000553](https://doi.org/10.1002/pmic.201000553)
201. Sardiou ME, Cai Y, Jin J, Swanson SK, Conaway RC, Conaway JW, Florens L, Washburn MP (2008) Probabilistic assembly of human protein interaction networks from label-free quantitative proteomics. *Proc Natl Acad Sci U S A* 105(5):1454–1459. doi:[0706983105 \[pii\] 10.1073/pnas.0706983105](https://doi.org/10.1073/pnas.0706983105)
202. Ishihama Y, Oda Y, Tabata T, Sato T, Nagasu T, Rappsilber J, Mann M (2005) Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Mol*

- Cell Proteomics 4(9):1265–1272. doi:[M500061-MCP200](https://doi.org/10.1074/mcp.M500061-MCP200) [pii] [10.1074/mcp.M500061-MCP200](https://doi.org/10.1074/mcp.M500061-MCP200)
203. Rappsilber J, Ryder U, Lamond AI, Mann M (2002) Large-scale proteomic analysis of the human spliceosome. *Genome Res* 12(8):1231–1245. doi:[10.1101/gr.473902](https://doi.org/10.1101/gr.473902)
 204. Mallick P, Schirle M, Chen SS, Flory MR, Lee H, Martin D, Ranish J, Raught B, Schmitt R, Werner T, Kuster B, Aebersold R (2007) Computational prediction of proteotypic peptides for quantitative proteomics. *Nat Biotechnol* 25(1):125–131. doi:[nbt1275](https://doi.org/10.1038/nbt1275) [pii] [10.1038/nbt1275](https://doi.org/10.1038/nbt1275)
 205. Braisted JC, Kuntumalla S, Vogel C, Marcotte EM, Rodrigues AR, Wang R, Huang ST, Ferlanti ES, Saeed AI, Fleischmann RD, Peterson SN, Pieper R (2008) The APEX quantitative proteomics tool: generating protein quantitation estimates from LC-MS/MS proteomics results. *BMC Bioinformatics* 9:529. doi:[1471-2105-9-529](https://doi.org/10.1186/1471-2105-9-529) [pii] [10.1186/1471-2105-9-529](https://doi.org/10.1186/1471-2105-9-529)
 206. Lundgren DH, Hwang SI, Wu L, Han DK (2010) Role of spectral counting in quantitative proteomics. *Expert Rev Proteomics* 7(1):39–53. doi:[10.1586/epr.09.69](https://doi.org/10.1586/epr.09.69)
 207. Paoletti AC, Parmely TJ, Tomomori-Sato C, Sato S, Zhu D, Conaway RC, Conaway JW, Florens L, Washburn MP (2006) Quantitative proteomic analysis of distinct mammalian Mediator complexes using normalized spectral abundance factors. *Proc Natl Acad Sci U S A* 103(50):18928–18933. doi:[0606379103](https://doi.org/10.1073/pnas.0606379103) [pii] [10.1073/pnas.0606379103](https://doi.org/10.1073/pnas.0606379103)
 208. Mann K, Mann M (2008) The chicken egg yolk plasma and granule proteomes. *Proteomics* 8(1):178–191. doi:[10.1002/pmic.200700790](https://doi.org/10.1002/pmic.200700790)
 209. Olson DG, Tripathi SA, Giannone RJ, Lo J, Caiazza NC, Hogsett DA, Hettich RL, Guss AM, Dubrovsky G, Lynd LR (2010) Deletion of the Cel48S cellulase from *Clostridium thermocellum*. *Proc Natl Acad Sci U S A* 107(41):17727–17732. doi:[1003584107](https://doi.org/10.1073/pnas.1003584107) [pii] [10.1073/pnas.1003584107](https://doi.org/10.1073/pnas.1003584107)
 210. Park SK, Venable JD, Xu T, Yates JR 3rd (2008) A quantitative analysis software tool for mass spectrometry-based proteomics. *Nat Methods* 5(4):319–322. doi:[nmeth.1195](https://doi.org/10.1038/nmeth.1195) [pii] [10.1038/nmeth.1195](https://doi.org/10.1038/nmeth.1195)
 211. Heinecke NL, Pratt BS, Vaisar T, Becker L (2010) PepC: proteomics software for identifying differentially expressed proteins based on spectral counting. *Bioinformatics* 26(12):1574–1575. doi:[btq171](https://doi.org/10.1093/bioinformatics/btq171) [pii] [10.1093/bioinformatics/btq171](https://doi.org/10.1093/bioinformatics/btq171)
 212. Griffin NM, Yu J, Long F, Oh P, Shore S, Li Y, Koziol JA, Schnitzer JE (2010) Label-free, normalized quantification of complex mass spectrometry data for proteomic analysis. *Nat Biotechnol* 28(1):83–89. doi:[nbt.1592](https://doi.org/10.1038/nbt.1592) [pii] [10.1038/nbt.1592](https://doi.org/10.1038/nbt.1592)
 213. Tsou CC, Tsai CF, Tsui YH, Sudhir PR, Wang YT, Chen YJ, Chen JY, Sung TY, Hsu WL (2010) IDEAL-Q, an automated tool for label-free quantitation analysis using an efficient peptide alignment approach and spectral data validation. *Mol Cell Proteomics* 9(1):131–144. doi:[M900177-MCP200](https://doi.org/10.1074/mcp.M900177-MCP200) [pii] [10.1074/mcp.M900177-MCP200](https://doi.org/10.1074/mcp.M900177-MCP200)
 214. Cox J, Mann M (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* 26(12):1367–1372. doi:[nbt.1511](https://doi.org/10.1038/nbt.1511) [pii] [10.1038/nbt.1511](https://doi.org/10.1038/nbt.1511)
 215. Bellew M, Coram M, Fitzgibbon M, Igra M, Randolph T, Wang P, May D, Eng J, Fang R, Lin C, Chen J, Goodlett D, Whiteaker J, Paulovich A, McIntosh M (2006) A suite of algorithms for the comprehensive analysis of complex protein mixtures using high-resolution LC-MS. *Bioinformatics* 22(15):1902–1909. doi:[btl276](https://doi.org/10.1093/bioinformatics/btl276) [pii] [10.1093/bioinformatics/btl276](https://doi.org/10.1093/bioinformatics/btl276)
 216. Pluskal T, Castillo S, Villar-Briones A, Oresic M (2010) MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics* 11:395. doi:[1471-2105-11-395](https://doi.org/10.1186/1471-2105-11-395) [pii] [10.1186/1471-2105-11-395](https://doi.org/10.1186/1471-2105-11-395)
 217. Jaffe JD, Mani DR, Leptos KC, Church GM, Gillette MA, Carr SA (2006) PEPPer, a platform for experimental proteomic pattern recognition. *Mol Cell Proteomics* 5(10):1927–1941. doi:[M600222-MCP200](https://doi.org/10.1074/mcp.M600222-MCP200) [pii] [10.1074/mcp.M600222-MCP200](https://doi.org/10.1074/mcp.M600222-MCP200)
 218. Mueller LN, Rinner O, Schmidt A, Letarte S, Bodenmiller B, Brusniak MY, Vitek O, Aebersold R, Muller M (2007) SuperHirn—a novel tool for high resolution LC-MS-based peptide/protein profiling. *Proteomics* 7(19):3470–3480. doi:[10.1002/pmic.200700057](https://doi.org/10.1002/pmic.200700057)
 219. Bondarenko PV, Chelius D, Shaler TA (2002) Identification and relative quantitation of protein mixtures by enzymatic digestion followed by capillary reversed-phase liquid chromatography-tandem mass spectrometry. *Anal Chem* 74(18):4741–4749
 220. Chelius D, Bondarenko PV (2002) Quantitative profiling of proteins in complex mixtures using liquid chromatography and mass spectrometry. *J Proteome Res* 1(4):317–323
 221. Christin C, Bischoff R, Horvatovich P (2011) Data processing pipelines for comprehensive profiling of proteomics samples by label-free LC-MS for biomarker discovery. *Talanta* 83(4):1209–1224. doi:[S0039-9140\(10\)00825-8](https://doi.org/10.1016/j.talanta.2010.10.029) [pii] [10.1016/j.talanta.2010.10.029](https://doi.org/10.1016/j.talanta.2010.10.029)
 222. Contrepois K, Ezan E, Mann C, Fenaille F (2010) Ultra-high performance liquid chromatography-mass spectrometry for the fast profiling of histone post-translational modifications. *J Proteome Res* 9(10):5501–5509. doi:[10.1021/pr100497a](https://doi.org/10.1021/pr100497a)
 223. Callister SJ, Barry RC, Adkins JN, Johnson ET, Qian WJ, Webb-Robertson BJ, Smith RD, Lipton MS (2006) Normalization approaches for removing systematic biases associated with mass spectrometry and label-free proteomics. *J Proteome Res* 5(2):277–286. doi:[10.1021/pr050300l](https://doi.org/10.1021/pr050300l)
 224. Kultima K, Nilsson A, Scholz B, Rossbach UL, Falth M, Andren PE (2009) Development and eval-

- uation of normalization methods for label-free relative quantification of endogenous peptides. *Mol Cell Proteomics* 8(10):2285–2295. doi:[M800514-MCP200](https://doi.org/10.1074/mcp.M800514-MCP200) [pii] [10.1074/mcp.M800514-MCP200](https://doi.org/10.1074/mcp.M800514-MCP200)
225. Rajcevic U, Niclou SP, Jimenez CR (2009) Proteomics strategies for target identification and biomarker discovery in cancer. *Front Biosci* 14:3292–3303. doi:[3452](https://doi.org/10.1093/fbs/kfn147) [pii]
226. Zhu W, Smith JW, Huang CM (2010) Mass spectrometry-based label-free quantitative proteomics. *J Biomed Biotechnol* 2010:840518. doi:[10.1155/2010/840518](https://doi.org/10.1155/2010/840518)
227. Lengqvist J, Andrade J, Yang Y, Alvelius G, Lewensohn R, Lehtio J (2009) Robustness and accuracy of high speed LC-MS separations for global peptide quantitation and biomarker discovery. *J Chromatogr B Analyt Technol Biomed Life Sci* 877(13):1306–1316. doi:[S1570-0232\(09\)00138-X](https://doi.org/10.1016/j.jchromb.2009.02.052) [pii] [10.1016/j.jchromb.2009.02.052](https://doi.org/10.1016/j.jchromb.2009.02.052)
228. Huang SK, Darfler MM, Nicholl MB, You J, Bemis KG, Tegeler TJ, Wang M, Wery JP, Chong KK, Nguyen L, Scolyer RA, Hoon DS (2009) LC/MS-based quantitative proteomic analysis of paraffin-embedded archival melanomas reveals potential proteomic biomarkers associated with metastasis. *PLoS One* 4(2):e4430. doi:[10.1371/journal.pone.0004430](https://doi.org/10.1371/journal.pone.0004430)
229. Huang JT, McKenna T, Hughes C, Leweke FM, Schwarz E, Bahn S (2007) CSF biomarker discovery using label-free nano-LC-MS based proteomic profiling: technical aspects. *J Sep Sci* 30(2): 214–225
230. Pavelka N, Fournier ML, Swanson SK, Pelizzola M, Ricciardi-Castagnoli P, Florens L, Washburn MP (2008) Statistical similarities between transcriptomics and quantitative shotgun proteomics data. *Mol Cell Proteomics* 7(4):631–644. doi:[M700240-MCP200](https://doi.org/10.1074/mcp.M700240-MCP200) [pii] [10.1074/mcp.M700240-MCP200](https://doi.org/10.1074/mcp.M700240-MCP200)