Journal of proteome research

pubs.acs.org/jpr

Tutorial

An Introduction to Mass Spectrometry-Based Proteomics

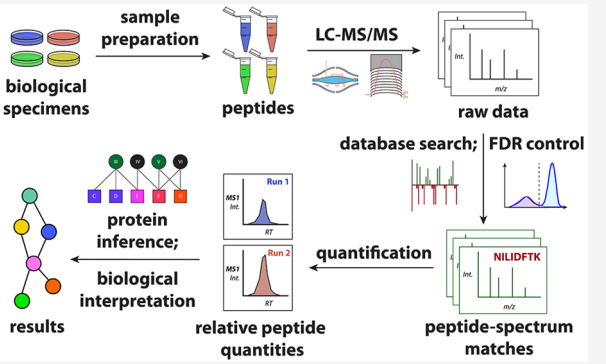
Steven R. Shuken*

Cite This: J. Proteome Res. 2023, 22, 2151−2171

Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: Mass spectrometry is unmatched in its versatility for studying practically any aspect of the proteome. Because the foundations of mass spectrometry-based proteomics are complex and span multiple scientific fields, proteomics can be perceived as having a high barrier to entry. This tutorial is intended to be an accessible illustrated guide to the technical details of a relatively simple quantitative proteomic experiment. An attempt is made to explain the relevant concepts to those with limited knowledge of mass spectrometry and a basic understanding of proteins. An experimental overview is provided, from the beginning of sample preparation to the analysis of protein group quantities, with explanations of how the data are acquired, processed, and analyzed. A selection of advanced topics is briefly surveyed and works for further reading are cited. To conclude, a brief discussion of the future of proteomics is given, considering next-generation protein sequencing technologies that may complement mass spectrometry to create a fruitful future for proteomics.



KEYWORDS: *mass spectrometry, proteomics, bottom-up, data-dependent acquisition, label-free quantification, untargeted proteomics*

1. INTRODUCTION

The field of mass spectrometry (MS)-based proteomics consists of a vast range of experiments that ask a diverse array of questions about proteins, including questions about protein sequences and abundance levels, the contents of subcellular compartments, protein functions, three-dimensional structures, chemical reactivities, protein−protein interactions, and more.[1] In MS-based proteomics, the primary instruments used to answer these questions are mass spectrometers, which measure ion mass-to-charge (m/z) values and signal intensities. MS can be performed on a range of scales, from the identification of a single protein to several thousands of proteins, and on samples with varying levels of complexity.

As an introduction to MS-based proteomics, this tutorial focuses on one of the most common and technologically simple MS-based proteomic experiments: a quantitative untargeted bottom-up proteomic experiment using data-dependent acquisition (DDA) on a complex sample such as a cell culture, biological tissue or fluid, or plant/fungal material (Figure 1). In this type of study, proteins are extracted from the sample and then "digested" into pieces of protein, called peptides, using one or more enzymes. Peptides are then analyzed by MS, with peptides ∼7−30 amino acids in length being the most suitable for analysis. When MS data are acquired in DDA mode, peptides are detected and then immediately selected one-by-one for fragmentation, which enables their sequences to be assigned in downstream data analysis. The phrase "bottom-up" refers to the subsequent inference of information about the proteins from the peptide analysis (Figure 1). In untargeted proteomics, which is focused on here, the data are analyzed to indiscriminately identify proteins and estimate their relative abundances. Peptides have several advantages over intact proteins, including having a less disperse distribution of sizes, being generally amenable to separation by reversed-phase high-performance liquid chromatography (HPLC), and usually yielding more easily interpretable fragment spectra. (Intact protein analysis has advantages which are discussed in the Advanced Topics and Further Reading section, but untargeted proteomics is usually performed on peptides.) Each step of the experiment, starting with sample preparation and ending with protein group quantification, is described in detail below. The technologies discussed here form the basis of most subfields of MS-based proteomics, a selection of which (including alternatives to untargeted proteomics, bottom-up mass spectrometry, and DDA) are briefly surveyed in the Advanced Topics and Further Reading section. Deep and broad reviews of mass spectrometry and its uses in proteomics can be found elsewhere.[2−4]

Received: December 22, 2022
Published: June 1, 2023

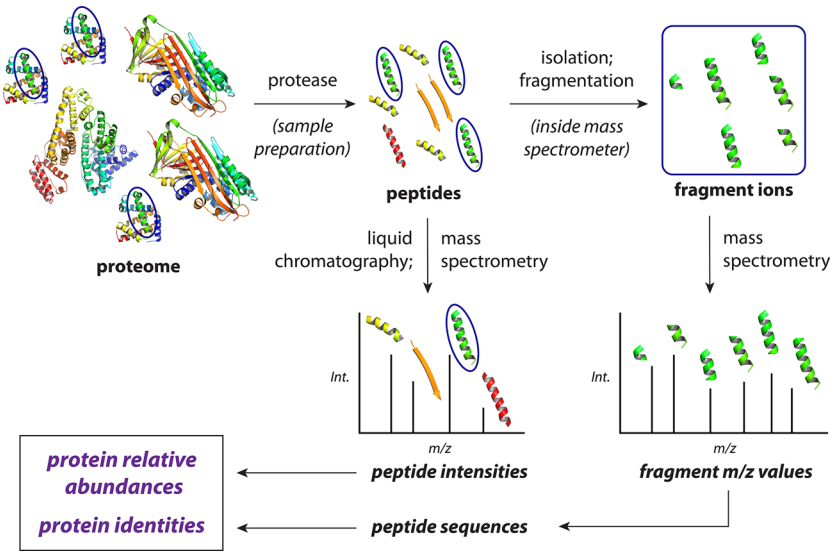


Figure 1. Simplified conceptual overview of a typical untargeted bottom-up mass spectrometry-based proteomic experiment. Using a protease, proteins are digested into peptides, which are separated by liquid chromatography, sprayed into a mass spectrometer, and enter the gas phase as ions. The mass spectrometer measures the ionized peptides' mass-to-charge values (m/z) and electrical signals produced by peptides' interactions with detectors. The intensities (Int.) of the electrical signals are plotted along the y-axis of each spectrum. Each of a selected set of peptides is then separately isolated and fragmented inside the mass spectrometer, and the fragments are measured. The resulting data are analyzed to identify the amino acid sequences of the peptides and measure their relative abundances, which are used to infer the identities and relative abundances of the proteins. One peptide (green) is highlighted with blue ovals as an example.

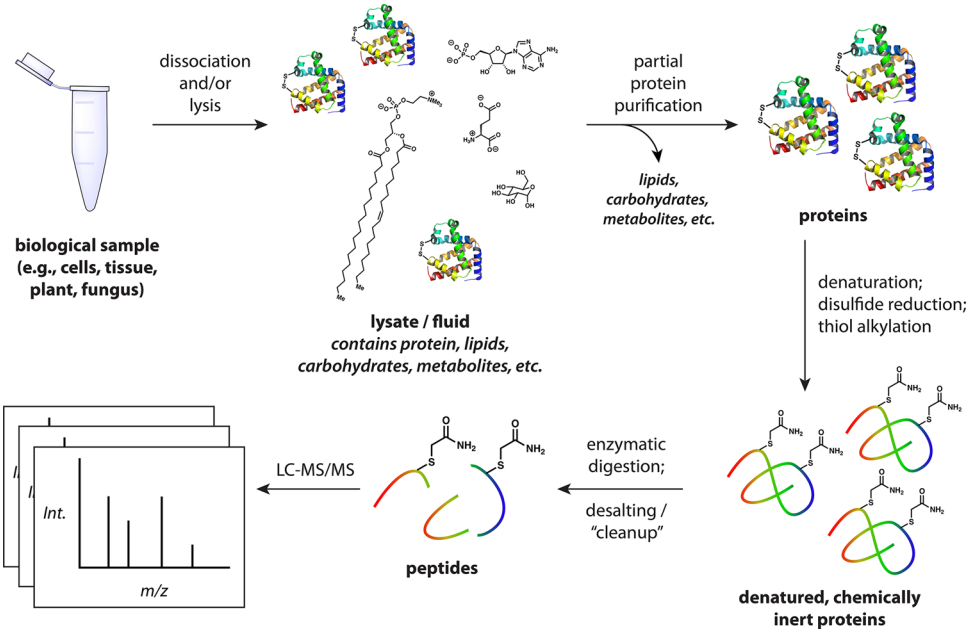


Figure 2. Generic sample preparation workflow for bottom-up mass spectrometry-based proteomics. Some steps may be rearranged, removed, or added to the workflow depending on the experiment. Reagents used for each step vary.

2. EXPERIMENTAL WORKFLOW

2.1. Sample Preparation

Every proteomic experiment begins with sample preparation. A generic sample preparation workflow for bottom-up proteomics is diagrammed in Figure 2 and described below. More specific details of particular workflows have been published; some effective protocols with which the author is familiar are cited here,[5−7] but the reader is encouraged to follow sample preparation protocols used in studies that best match their own experimental plan.

First, if tissues or plant samples are being analyzed, they are homogenized/dissociated. Cells are lysed. Often, the protein amount is measured using a commercial assay such as the bicinchoninic acid (BCA) assay;[8] the amount of material necessary varies widely depending on the technology and application, from nanograms to milligrams, but for a simple bottom-up label-free untargeted analysis performed with an orbitrap-based spectrometer, 5−50 μg per sample is sufficient (though it is advisible to prepare extra material). The proteins are partially purified (separated from other types of molecules in the lysate, e.g., lipids and carbohydrates) by precipitation, e.g., in

https://doi.org/10.1021/acs.jproteome.2c00838
J. Proteome Res. 2023, 22, 2151−2171

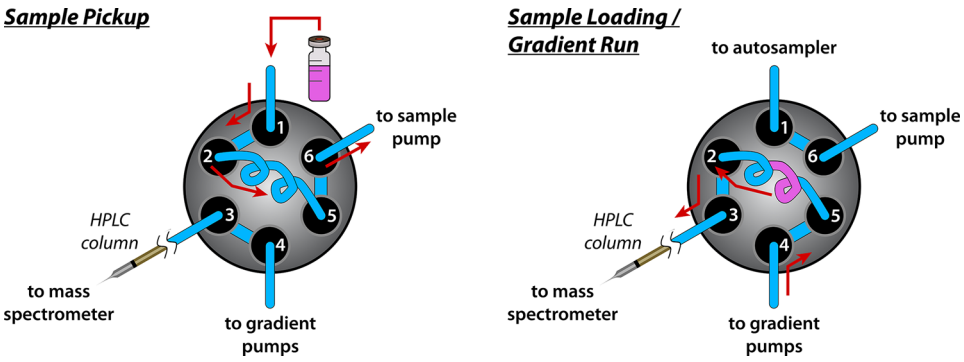


Figure 3. Simplified illustration of an example of an HPLC method for LC-MS/MS. In this example, a six-port valve has port 1 connected to the autosampler, ports 2 and 5 linked by the sample loop (drawn as a coil), port 3 connected to the column which is positioned in front of the mass spectrometer, port 4 connected to the gradient pumps, and port 6 connected to the sample pump. Liquid flow is indicated by red arrows. In the first step, valves 1 and 2 are connected and the sample pump applies negative pressure (pulls), transferring some amount of the sample into the sample loop. In the second step, port 4 is connected to port 5 and port 2 is connected to port 3 so that the gradient pumps, sample loop, and column are all connected. The gradient pumps apply positive pressure, pushing the sample out of the sample loop onto the column and spraying the eluate into the mass spectrometer.

cold acetone,[9] on a solid surface such as beads or a filter,[5,7] or in a biphasic water−methanol−chloroform mixture.[10] The proteins are then washed and resuspended in an aqueous buffer. After this step (or before this step, during lysis), the partially purified proteins are denatured (unfolded), typically with urea and/or a detergent such as sodium deoxycholate (DOC). The proteins are then rendered chemically inert by disulfide bond reduction followed by thiol alkylation, usually with dithiothreitol (DTT) or tris(2-carboxyethyl)phosphine (TCEP) followed by iodoacetamide (IAA). By converting disulfides and thiols to thioethers (Figure 2), this prevents the formation of thiol oxidation products or disulfide bonds throughout the rest of the workflow, which would complicate the analysis.[11]

The reduced and alkylated proteins are then "digested," i.e., some of their peptide bonds are hydrolyzed by proteolysis, most commonly with LysC and trypsin (either simultaneously or sequentially) at 37 °C for a combined duration of at least 1 h[7] and often over 14 h.[6] Trypsin cleaves the peptide bond on the C-terminal side of lysine (K) and arginine (R), though its efficiency is reduced in certain cases, especially when the K/R is followed by a proline residue.[12] LysC cleaves C-terminal to K, cleaving efficiently at some sites at which trypsin's efficiency is reduced, including lysines flanked by acidic residues.[13] LysC is also more stable under denaturing conditions than trypsin.[14] This combination is preferred over other proteases because (1) it is generally efficient and effective, (2) its high selectivity for K and R result in few potential proteolytic products to consider during database search, (3) the common occurrence of K and R in proteomes result in peptides with suitable lengths for mass spectrometry, and (4) it produces peptides that are likely to carry positive charge because most of the cleavage products contain K or R, both of which have high pK_a values.

Any remaining detergents, which suppress the signal in MS, must be removed before the next step. Some are usually partially removed by protein precipitation before digestion (see above);[5,7] conversely, DOC may be removed after digestion by centrifugation at a low pH.[6] The digested peptides are then cleaned, e.g., by solid-phase extraction (SPE): centrifugation or vacuum suction through a C18-coated solid phase using alternating mobile phases for trapping, washing, and elution. This step is sometimes referred to as "desalting" because buffer salts, which can adversely affect the ionization/desolvation process and/or suppress signal, are removed.[15] The eluent is then evaporated and the peptides are finally resuspended in an aqueous buffer compatible with liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis, such as 0.1% formic acid in high-performance liquid chromatography(HPLC)-grade water with a small amount of acetonitrile (e.g., 3−5%).

At this point, the peptides are ready for LC-MS/MS analysis. The sample is put into the autosampler of an HPLC instrument, which is attached to a mass spectrometer, and an acquisition method is run. The acquisition method executes the uptake and analysis of the peptides. In the type of experiment described here, LC-MS/MS achieves three things: it partially separates and concentrates the peptides into detectable chromatographic peaks, acquires data that enable assignment of the peptides' amino acid sequences, and acquires data that enable relative quantification of the peptides.

2.2. Liquid Chromatography-Tandem Mass Spectrometry (LC-MS/MS)

The first step that occurs after the LC-MS/MS run begins is sample pickup and loading onto the HPLC column. This process can happen in many different ways depending on the instrument; a simple example is illustrated in Figure 3. In untargeted proteomic methods using HPLC, this usually involves taking up a specified amount of sample (e.g., 2 μL containing 0.1−1 μg of peptide) by applying negative pressure (pulling) and then applying positive pressure (pushing) to load the sample onto the HPLC column. Once the peptides are loaded onto the column, four things begin to happen in concert: liquid chromatography, the formation of gas-phase peptide ions (ionization), mass spectrometry of peptides (MS1), and mass spectrometry of fragments (MS2). In LC-MS/MS, "LC," "MS," and "MS" signify the first, third, and fourth steps, respectively.

2.2.1. High-Performance Liquid Chromatography (HPLC or "LC"). The sample is eluted through an HPLC column using a preprogrammed solvent gradient, i.e., a mixture of solvents that changes in composition throughout the run, whose length is typically 30−180 min. In proteomics, HPLC is almost always conducted in reversed-phase mode, meaning that the column is packed with a hydrophobic stationary phase (typically silica coated with linear hydrocarbon chains 18 carbons in length, referred to as C18). The solvents (mobile phase) are usually a mixture of two solutions, each pumped by its own gradient pump: an aqueous buffer (often 0.1% formic acid

https://doi.org/10.1021/acs.jproteome.2c00838
J. Proteome Res. 2023, 22, 2151−2171

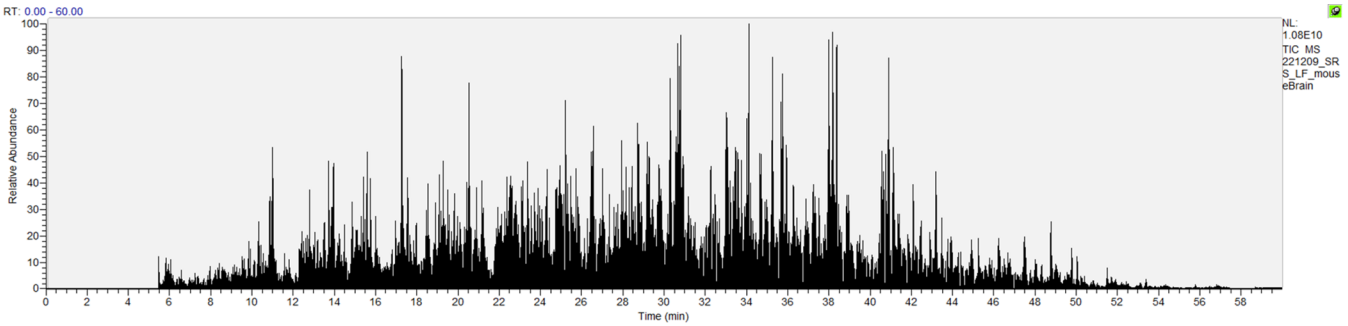


Figure 4. Typical chromatogram from an LC-MS/MS run. Each point on the plot represents an acquisition of a spectrum. The *x*-axis is the time (retention time, RT) at which the spectrum was acquired. The *y*-axis is the total electrical signal from all ions in the spectrum (total ion current, TIC), scaled to the TIC of the highest point in the chromatogram (normalized level, NL). Sample: mouse brain proteins digested with trypsin and LysC. Peptides (0.5 μg) analyzed on an Orbitrap Eclipse mass spectrometer (Thermo Fisher Scientific) with an EASY-nLC 1200 (Thermo Fisher Scientific) using a 60-minute method. Peptides were separated using a linear gradient from 4% Buffer B (95% acetonitrile + 5% water + 0.1% formic acid) in Buffer A (95% water + 5% acetonitrile + 0.1% formic acid) to 33% Buffer B in Buffer A. Plot generated in XCalibur Qual Browser (Thermo Fisher Scientific).

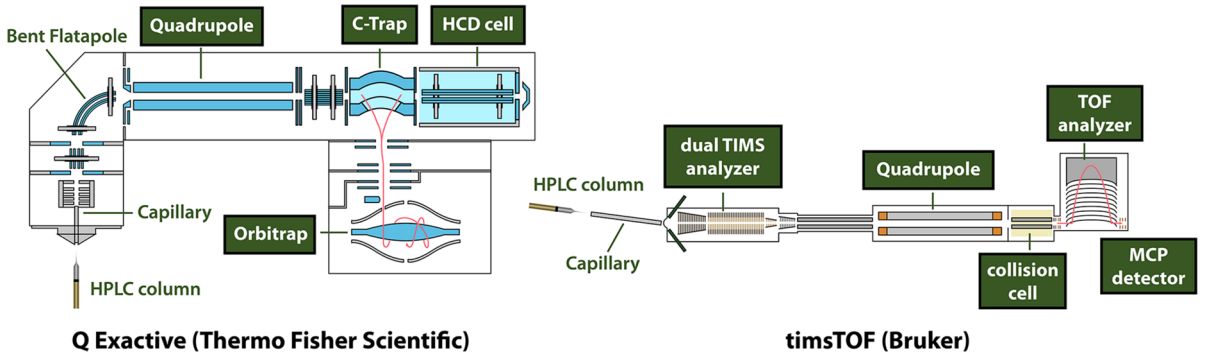


Figure 5. Schematics of two popular mass spectrometers in proteomics: top view of Q Exactive (Thermo Fisher Scientific) and side view of timsTOF (Bruker). The tip of the HPLC column is pointed at the front of the spectrometer, positioned some small distance away (e.g., 0.5−5 mm). A voltage is applied across the gap between the HPLC column and the front of the spectrometer by placing an electrode, which is connected to the spectrometer, in the liquid path somewhere before the column tip. Liquid droplets containing peptides are sprayed from the column tip; peptides become mostly desolvated before entering the capillary, which is heated to a high temperature to help complete desolvation. Applied electromagnetic fields direct and focus the ion beam. During MS2 spectrum acquisitions (scans) in a Q Exactive or timsTOF, the quadrupole filters out all but a small range of m/z values centered around the desired precursor m/z. The higher-energy collision-induced dissociation (HCD) cell and collision cell are where high-velocity precursors collide with gas particles, generating fragments. In orbitrap-based mass spectrometers such as the Q Exactive, peptides or fragments are collected in the C-trap for some time ("injection time") before an applied voltage injects them into the orbitrap for mass analysis and detection (m/z and intensity measurements, i.e., a scan). In the timsTOF, ions are collected in the first region of the dual trapped ion mobility spectrometry (TIMS) analyzer for some time and then separated in the second part of the dual TIMS analyzer. MS1 and MS2 scans are then performed on different populations of peptides stored in the dual TIMS analyzer. Scans are performed by time-of-flight (TOF) mass analysis followed by detection by the MCP detector. The end of the ion flight path is indicated as a pink line. For more detailed diagrams and descriptions, see refs 21−26.

in water) and an organic solvent (often acetonitrile with small amounts of formic acid, e.g., 0.1%, and water, e.g., 5%). Peptides are partially separated according to their hydrophobicities, eluting at different times (retention times, RTs) during the gradient. The gradient can be altered to optimize separation and chromatographic peak shape: a longer gradient allows more time to detect more peptides, whereas a shorter gradient allows more runs during the experiment and can increase signal by sharpening chromatographic peaks. The optimal gradient depends on several factors, including the column, sample complexity, instrument, and the goal of the experiment. Figure 4 shows an example of peptide separation using an HPLC gradient: each peak in the chromatogram represents a distinct set of ions.

2.2.2. Peptide Ionization. As the gradient proceeds, the partially separated peptides are continuously sprayed into the spectrometer. Charged liquid droplets containing peptides are sprayed from the HPLC column tip under the influence of an applied voltage; charged peptides become desolvated (enter the gas phase), enter the spectrometer, and are propelled forward by electromagnetic fields within the spectrometer (Figure 5).[16,17] Ionizability (the propensity to become a gas-phase ion) varies widely among peptides;[18] those with insufficient ionizability are not detected. Charge-neutral species are not guided by electromagnetic fields and therefore fly astray, e.g., upon desolvation or at the first bend in the flight path inside the spectrometer. A given peptide can typically adopt different charge states upon ionization; each charge state usually corresponds to a protonation state (how many protons are bound to the peptide). Different charge states result in different m/z values: for example, an ionized peptide with $z = 1$ will have $m/z = [M + H]$, where M is the mass of the neutral peptide and H is the mass of a proton, whereas the same peptide with $z = 2$ will have $m/z = [M + 2H]/2$. Because of this, different charge state-specific versions of a peptide are isolated separately and serve separately as precursors in fragmentation reactions (see below); thus, they are referred to as different "precursors" corresponding to that peptide.

https://doi.org/10.1021/acs.jproteome.2c00838
J. Proteome Res. 2023, 22, 2151−2171

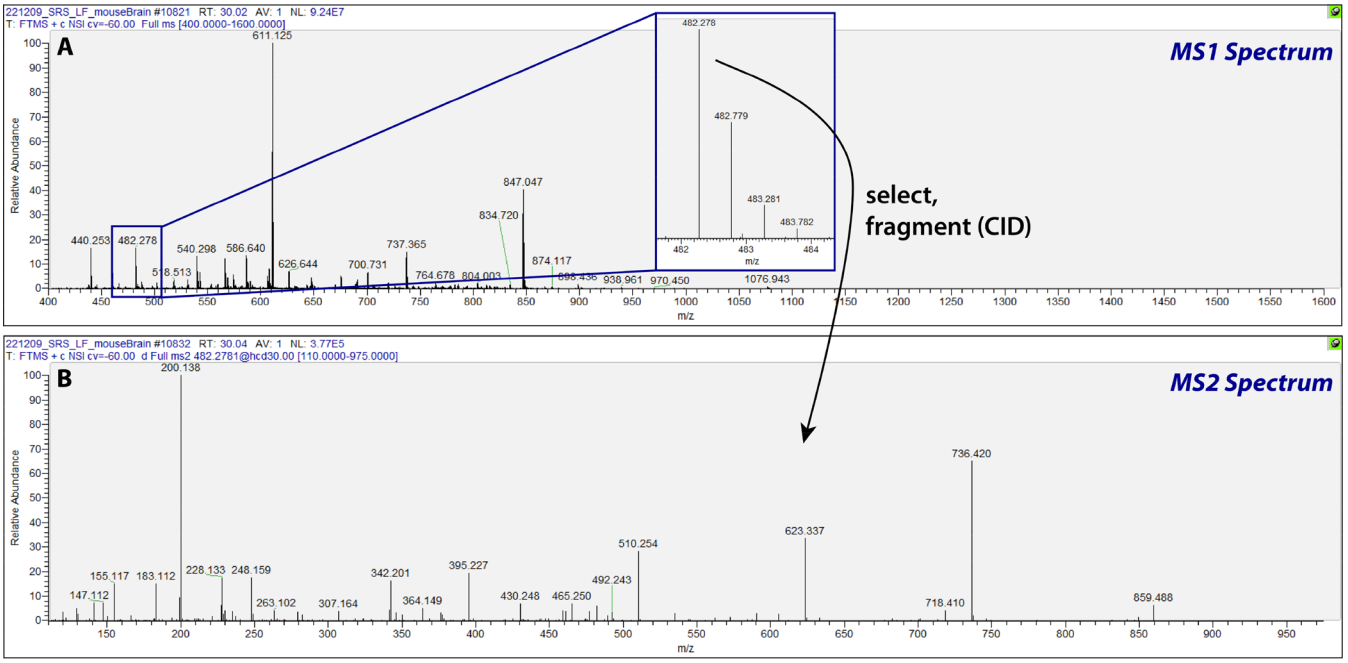


Figure 6. Examples of MS1 and MS2 spectra. The *y*-axis is signal intensity normalized (divided) by injection time and then scaled to the tallest peak (normalized level, NL). Here, a peptide with m/z = 482.278 Th was detected by MS1, selected, and fragmented using higher-energy CID (HCD); the fragments were mass-analyzed and detected, producing an MS2 spectrum. A. MS1 spectrum. Inset: peptide isotopic pattern. B. MS2 spectrum. AV = number of averaged scans. FTMS = Fourier transform mass spectrum (produced by Fourier transform-based mass analyzer, an orbitrap in this case). c = centroid mode, i.e., peaks are reported as 1-dimensional spectral lines (centroids) rather than the 2-dimensional peaks produced by raw Fourier transform. NSI = nanospray ionization (i.e., nanoliter-flow-rate HPLC ion source). cv = FAIMS compensation voltage. hcd30.00 = HCD at normalized collision energy (NCE) 30%. Spectra acquired on an Orbitrap Eclipse (Thermo Fisher Scientific) and visualized in XCalibur Qual Browser (Thermo Fisher Scientific).

2.2.3. Peptide MS Scans (MS1). The spectrometer repeatedly and quickly acquires spectra to detect ionized peptides. These spectrum acquisition events are called "MS1 scans." The acquisition of any mass spectrum requires a mass analyzer, which resolves ions according to their mass-to-charge (m/z) values, and a detector, which measures an electrical signal. The signal generated by a peptide is related to the peptide's ionizability, charge state, and abundance, and the abundance-signal relationship is generally linear over some range of values.[19] Several mass analyzers and detectors are effective for proteomics; two popular examples of mass analyzers are the time-of-flight (TOF) analyzer[20−22] and the orbitrap (Figure 5).[23,24] The orbitrap also contains electrodes that detect ions,[25] whereas a TOF analyzer must be coupled to a detector, typically a secondary electron emission-based detector such as a multichannel plate (MCP).[21,22]

A relatively simple orbitrap mass spectrometer is the Q Exactive (Thermo Fisher Scientific), the interior of which is depicted in Figure 5.[26] For an MS1 scan, all ions within a wide m/z range, such as 400−1600 Th, are allowed to pass through the quadrupole (Thomson, Th, is the m/z unit[27]). The peptides are trapped in the C-trap, which applies an electromagnetic field that functions as a "well" for the peptides to "fall into" with the help of nitrogen (N_2) molecules that absorb kinetic energy.[28] After some ion accumulation time ("injection time"), often <0.2 ms but ranging up to ∼50 ms for MS1, the C-trap feeds its ions simultaneously and instantaneously into the orbitrap. Once inside the orbitrap, ions precess around the central spindle and oscillate along the spindle at frequencies proportional to their m/z values. As they move, they induce a fluctuating current, which is processed via Fourier transform to derive m/z values and intensities.[23,25] The current is measured for an adjustable period, typically tens to hundreds of milliseconds.[29−31] This method of mass analysis is useful for its high m/z resolution and accuracy. When the measurement is complete, an MS1 spectrum is generated such as that shown in Figure 6A. The quantity plotted on the *y*-axis is intensity normalized to (divided by) injection time. As the peptides are sprayed into the mass spectrometer, MS1 spectra are acquired repeatedly in this manner, many times per minute and sometimes multiple times per second. Each MS1 spectrum has the potential to trigger MS2 acquisition events, which occur before the next MS1 spectrum is acquired.

The timsTOF (Bruker) also performs mass analysis and detection of peptides to acquire MS1 spectra. However, one major conceptual difference between the timsTOF and the Q Exactive is the added trapped ion mobility spectrometry (TIMS) element (Figure 5). As peptides enter the timsTOF, they accumulate in the first region of the dual TIMS analyzer for 25−200 ms and then are positioned at different positions along the second part of the TIMS analyzer by a gas flow opposed by an increasing longitudinal electric field gradient.[22] Position in the second part of the dual TIMS analyzer relates to an ion's ability to move through a gas ("mobility," which is inversely related to collision cross section). They are then "eluted" gradually from the dual TIMS analyzer over the course of 25−200 ms, thereby performing an additional separation of the different precursors. Because mass analysis by TOF and detection by MCP take a very short time (∼0.1 ms per scan), many MS1 spectra are acquired during the emptying of the dual TIMS analyzer, with each MS1 acquisition potentially triggering MS2 spectra before the next MS1 as with the Q Exactive. Some of the advantages of TIMS,

https://doi.org/10.1021/acs.jproteome.2c00838
J. Proteome Res. 2023, 22, 2151−2171

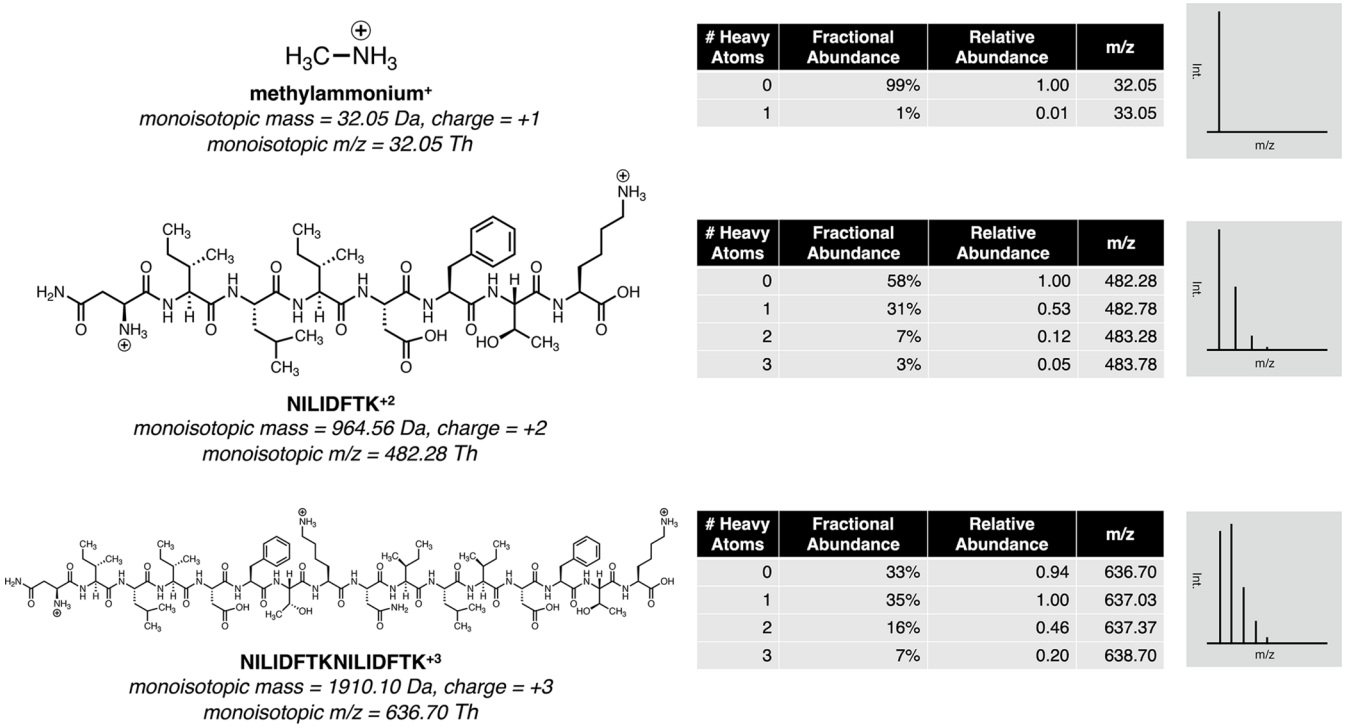


Figure 7. Simulated MS1 spectra of cations of increasing mass. Spectra were generated in ChemDraw (PerkinElmer). Monoisotopic mass is the sum of atomic masses using the most abundant naturally occurring isotope for each atom. Da = daltons.

such as a reduced probability of fragmenting multiple different precursors at once, can be gained on an orbitrap-based spectrometer by equipping a high-field asymmetric waveform ion mobility spectrometry (FAIMS) device.[32−34]

MS1 acquisition measures m/z values and intensities of ionized peptides but lacks the information necessary to determine peptide identities/sequences. Although m/z values are usually determined with high accuracy (at least three decimal places), there are usually too many possible peptides within the allowed range of m/z errors to identify peptides based on m/z alone. To gain the information necessary for identification, peptides are fragmented (typically at peptide bonds) and the fragments are measured with MS. The resulting spectra are called "MS2 spectra."

2.2.4. Fragment MS Scans (MS2). After each MS1 scan, the spectrometer, while running, may select some ion(s) observed in the MS1 spectrum to be reaccumulated and fragmented. The highest-intensity precursors in the spectrum are usually selected, though multiple strategies for precursor selection exist.[35] To avoid redundantly selecting previously fragmented peptides, recently fragmented m/z values are often excluded from selection. The length of time during which a previously fragmented m/z value is excluded from selection (called "dynamic exclusion" in the context of Thermo Fisher Scientific mass spectrometers) is usually chosen to be 90 s or shorter;[36,37] however, the time it takes for a peptide to elute (chromatographic peak width) varies widely. (Chromatographic peak width depends on various features of the HPLC, including the column, gradient, and flow rate, as well as on features of the peptide, including abundance and hydrophobicity. Therefore, dynamic exclusion duration is best selected on the basis of empirical chromatographic peak widths.) Because of dynamic exclusion, many precursors are fragmented only once, early in their elution from the column (though some may be fragmented more than once). Therefore, major MS1 peaks in a spectrum such as in Figure 6A have often already been fragmented some seconds before.

To determine whether an MS peak likely represents a peptide (as opposed to a contaminant ion, e.g., from the sample or the surrounding air), the spectrometer looks for neighboring peaks that indicate the presence of heavy isotopes such as ^{13}C and ^{15}N. Although ^{13}C makes up only ∼1% of carbon atoms in nature (∼99% are ^{12}C), the probability that a molecule contains a ^{13}C atom increases with the number of carbon atoms; this rule applies to other elements as well, including nitrogen (^{15}N comprises ∼0.4% of nitrogen atoms in nature whereas ∼99.6% are ^{14}N). Peptides tend to have enough atoms that ^{13}C-containing (and, to a lesser extent, ^{15}N-containing) ions become a significant or even dominant contribution to the peaks observed (Figure 7). An ion whose carbons are all ^{12}C will have a mass (m) 1.003 Da lower than the same ion with one ^{13}C instead of ^{12}C. The corresponding MS peak of the ^{13}C-containing ion therefore will have an m/z value that is about $1/z$ higher: 1.0 Th higher for $z = 1$, 0.5 for $z = 2$, ∼0.33 for $z = 3$, and so on. This is true for each additional ^{13}C, which is reflected in the m/z spacings of the isotopic peaks in Figure 7. (The spacings induced by ^{15}N are very similar; an additional neutron on N adds only 0.006 Da less than an additional neutron on C.) Thus, given any observed peak, the charge of the ion can be putatively assigned if neighboring peaks are found at these m/z distances away. The relative intensities of these peaks also give an estimate of how many C and N atoms are in the ion and therefore how likely the ion is to be a peptide; at a given m/z, the typical ratio of isotopic peak intensities for a peptide is different than for other molecules such as detergents or carbohydrates.[38,39] Because many nonpeptide contaminant ions have a charge of +1 and because fragmentation of peptides with $z = +1$ is often less informative,

https://doi.org/10.1021/acs.jproteome.2c00838
J. Proteome Res. 2023, 22, 2151−2171

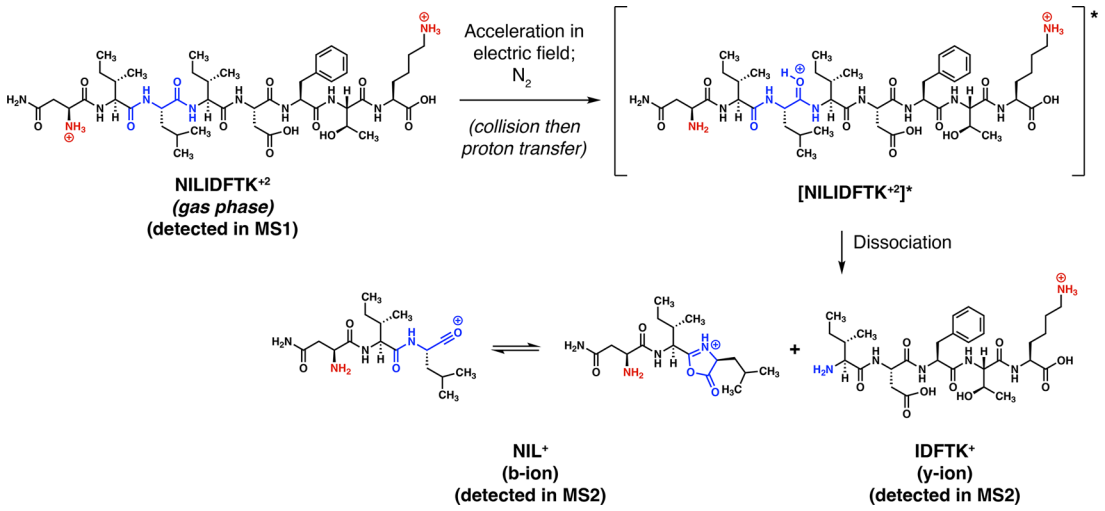


Figure 8. Possible mechanism of a hypothetical example of a collision-induced dissociation (CID)-based fragmentation event based on the mobile proton model (refs 44, 45). Brackets indicate unstable intermediate; asterisk indicates collision-induced excited state.

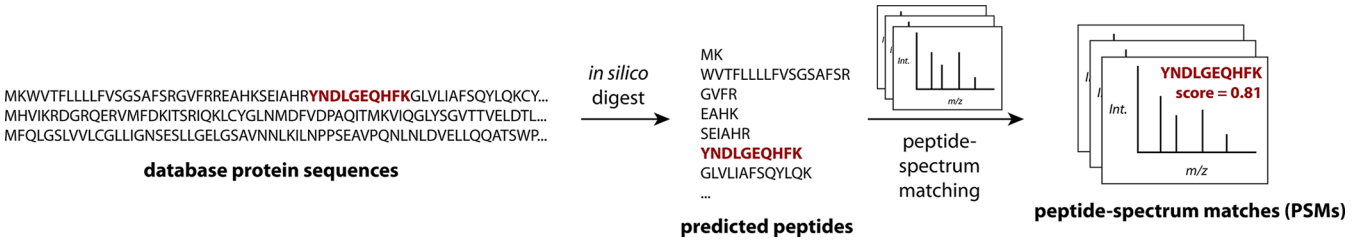


Figure 9. Simplified overview of database search. *In silico* digest = computational prediction of digestion products.

ions of charge +2 or greater are typically selected for fragmentation.

One at a time, each selected precursor is then fragmented using collision-induced dissociation (CID) and an MS2 spectrum of the fragments is acquired. These MS2 spectra are used to identify the peptides.[40] To produce an MS2 spectrum with a Q Exactive or timsTOF, all of the elements of the spectrometer are utilized (Figure 5). The quadrupole acts as a filter, ensuring that only ions with the selected m/z (within some small adjustable tolerance, usually between ±0.2 and ±1.0 Th[41]) pass through, and the ions are accumulated (in the C-trap in orbitrap-based mass spectrometers or in the dual TIMS analyzer in the timsTOF). The ions are then accelerated into the higher-energy CID (HCD) cell or collision cell, where they collide with gas molecules (usually N_2).[21,42,43] Although the mechanisms of CID reactions are incompletely understood and may vary depending on the amino acid sequence and protonation state,[44] CID is thought to proceed, at least sometimes, according to the mobile proton model, in which collision induces a proton to be transferred from somewhere on the peptide to one of the peptide backbone amide bonds before dissociation (Figure 8).[45] CID tends to preferentially cleave C−N bonds between carbonyl carbons and amide nitrogens (i.e., peptide bonds), resulting in fragments whose m/z values are predictable, called b- and y-ions:[46] b-ions contain the C terminus of the peptide, whereas y-ions, which tend to predominate for peptides generated by trypsin and/or LysC, contain the N terminus of the peptide.[47] The resulting fragments then undergo a mass analysis and detection process similar to MS1, producing an MS2 spectrum such as that shown in Figure 6B.

Because the range of protein abundances (dynamic range) is vast in most complex biological samples, and because the peptides producing the most intense signals are usually selected for fragmentation, it can be difficult to detect less abundant peptides. The HPLC partially addresses this by separating peptides (see above), but this can be further addressed by sample preparation techniques such as enrichment[48] or fractionation,[50] by separation by other methods such as ion mobility-based approaches (see above), or by alternative acquisition modes such as data-independent acquisition or targeted MS (see the Advanced Topics and Further Reading section).

2.2.5. Summary. Throughout the gradient (typically 30−180 min), the spectrometer continually accumulates gas-phase ions. The ions are analyzed (MS1) and each ion is assessed for peptide-like properties. Each putative peptide is reaccumulated and fragmented, and the fragments are analyzed (MS2). The cycle then repeats. In a Q Exactive, if an MS1 scan took 50 ms and triggered 10 MS2 scans each taking 200 ms, then the cycle would be complete in about 2 s. In a timsTOF, a typical cycle takes 25−200 ms.[21] This cycle repeats throughout the course of the gradient hundreds to thousands of times, generating thousands of spectra. This acquisition mode is called data-dependent acquisition (DDA) because the decision to acquire an MS2 spectrum depends on the MS1 data; the alternative, data-independent acquisition (DIA), is discussed in the Advanced Topics and Further Reading section. As mentioned above, insufficiently ionizable peptides will not be detected, but a peptide must also fragment well—i.e., produce sufficiently informative fragments—to be identified.[49]

https://doi.org/10.1021/acs.jproteome.2c00838
J. Proteome Res. 2023, 22, 2151−2171

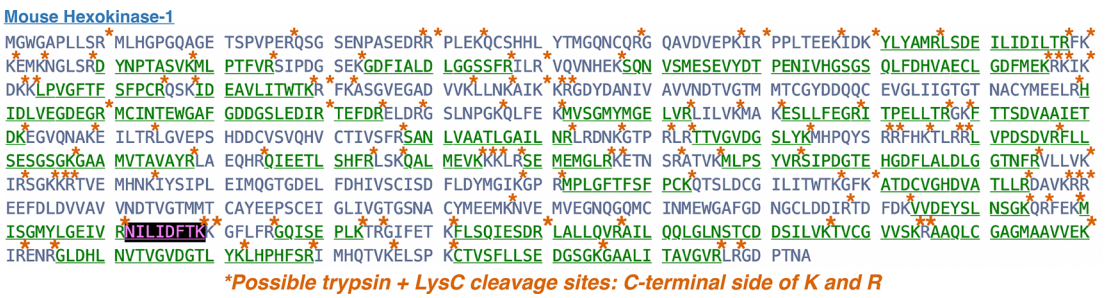


Figure 10. Amino acid sequence of mouse hexokinase-1. Underlined green regions are composed of peptide sequences identified in a real LC-MS/MS analysis of mouse brain proteins. Possible tryptic cleavage sites are highlighted with orange asterisks (certain sites such as R-P are less likely to be cleaved, but all are potential cleavage sites). The peptide fragmented in Figure 6 is highlighted with inverted color.

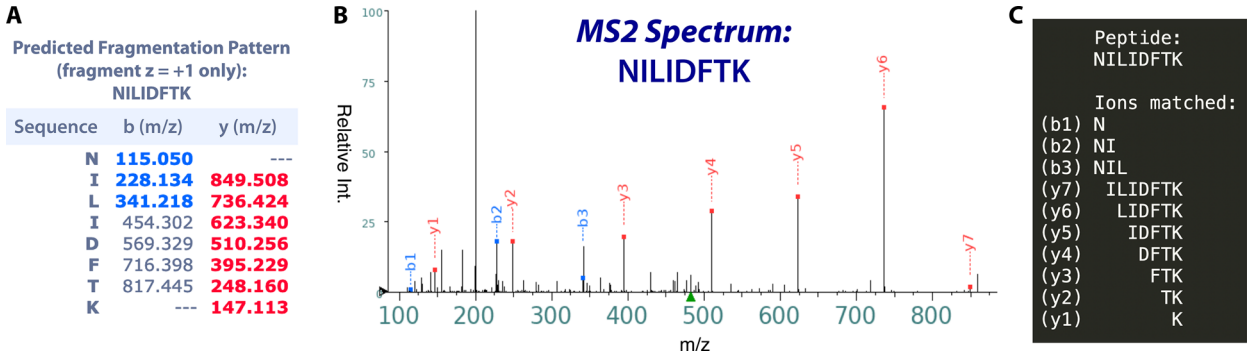


Figure 11. Fragment m/z values, annotated MS2 spectrum, and matched fragment sequences for the peptide fragmented in Figure 6 and highlighted in Figure 10 (NILIDFTK). A. Theoretical m/z values of b- and y-ions. A b-ion includes the N terminus of the peptide whereas a y-ion includes the C terminus (ref 46). Bold and colored m/z values represent peaks observed in the MS2 spectrum. B. Annotated MS2 spectrum. Number after "b" or "y" signifies amino acid length of fragment. C. Sequences of fragments matched to spectrum peaks.

3. DATA ANALYSIS: PEPTIDE IDENTIFICATION

The raw data set produced by a single LC-MS/MS run is a large collection of spectra, each with a retention time, m/z values, intensities, and metadata (e.g., information in scan headers in Figure 6). Software packages such as MaxQuant[51] or Proteome Discoverer (Thermo Fisher Scientific) process these data to generate a peptide and/or protein list with a score for each identification. The various software options for peptide searching are compared elsewhere,[52−54] as are the algorithms for peptide identification that those software packages implement.[55−57] The most common peptide identification approach, used in most untargeted bottom-up DDA proteomic experiments, is called database search (Figure 9). A general description of database search is given in this section, and in the next section, one database search algorithm, SEQUEST, is described in more detail.

In practice, database search begins with the user loading raw data files into the software along with a known reference proteome/database as a text file, usually in FASTA[58] format. Uniprot.org has reference proteomes for many species and is widely used for human and mouse;[59] other databases are preferred for certain species.[60,61] If a protein is not in the user-supplied FASTA file, it will not appear in the final list, even if it was present in the sample. *De novo* sequencing algorithms do not require a reference proteome and therefore lack this limitation; however, they have their own disadvantages and are usually only used for specialized applications in which the peptides in the sample are not encoded by a species' genome, such as antigen presentation profiling or synthetic peptide library analysis (see the Advanced Topics and Further Reading section).[62−64] With database search, the data processing software predicts all peptides that could arise from the proteins in the database by enzymatic cleavage (most commonly with trypsin and LysC) and predicts the MS2 spectra of the corresponding charge-specific peptide ions (precursors), which serve as "fingerprints" for the peptides. These predicted peptides and their predicted spectra are compared with the experimental spectra to make peptide-spectrum matches (PSMs) (Figure 9). After false discovery rate (FDR) control (see below), peptide identification is complete.

As an example, the amino acid sequence of mouse hexokinase-1 is shown in Figure 10 with cleavage sites potentially targeted by trypsin and LysC ("tryptic" cleavage sites) highlighted with orange asterisks. Underlined green regions are composed of sequences of peptides that were identified in an LC-MS/MS analysis of mouse brain tissue. Note that all underlined regions end with an arginine (R) or lysine (K) residue and are preceded by an R or K. The MS2 spectrum for the peptide highlighted in Figure 10 (NILIDFTK) is shown in Figure 6B and annotated in Figure 11. Peaks that match predicted fragment ions are labeled with a blue "b" or a red "y" and an integer signifying the amino acid length of the fragment. The letter signifies the type of fragment (b-ion or y-ion; see above).[46,47] Because these are the ions predominantly formed with CID, only b- and y-ions are usually searched for in the analysis.

The spectrum in Figure 11 is a good example of how a single spectrum can provide strong evidence for the sequence of a peptide in the sample. The high quality of the PSM, qualitatively evident in Figure 11, is represented by numerical score(s) generated by the search algorithm based on how well the observed spectrum matches the peptide. Software packages use these scores to estimate the likelihood that a PSM is a false

https://doi.org/10.1021/acs.jproteome.2c00838
J. Proteome Res. 2023, 22, 2151−2171

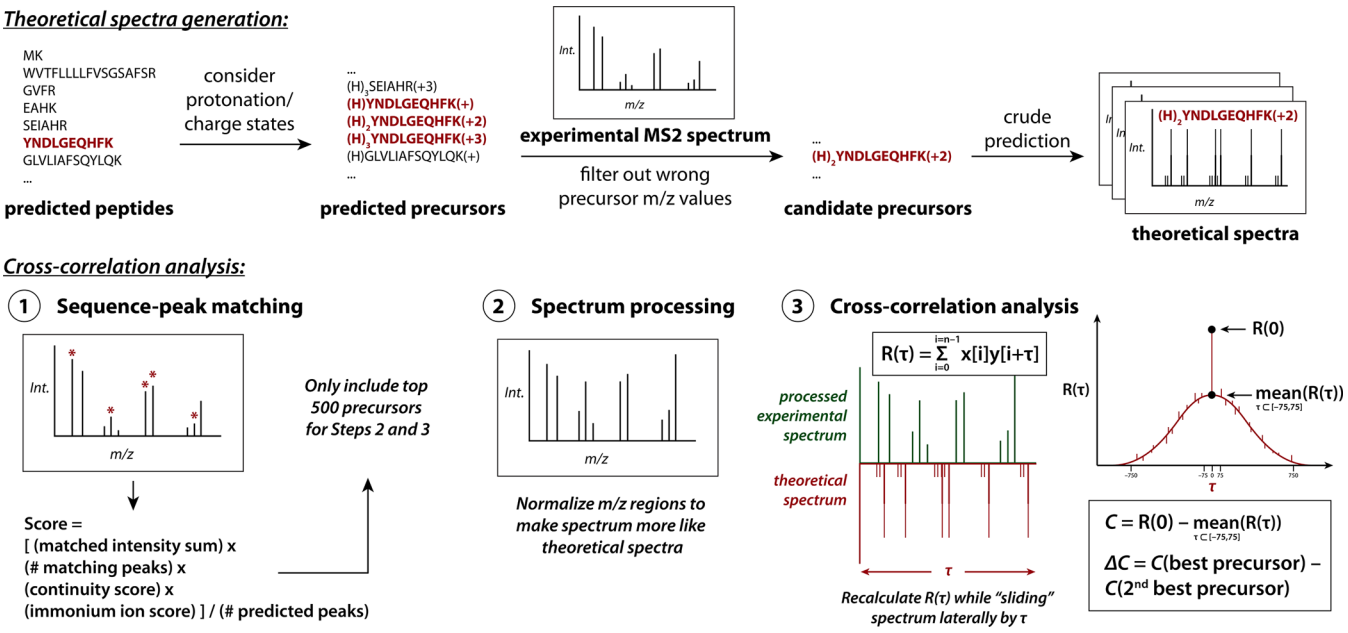


Figure 12. Simplified scheme of the original SEQUEST database search algorithm. This algorithm is executed for each experimentally acquired MS2 spectrum. Top: Filtering of predicted peptides into a set of candidate precursors (i.e., charge state-specific peptide ions) relevant to the spectrum and generation of theoretical spectra to which to compare the experimental spectrum. Bottom: the peptide-spectrum matching algorithm. Note: "step" numbers are chosen for simplicity and do not match the original publication. In Step 1, the sequence is crudely compared to the spectrum to quickly rank the theoretical spectra. In Step 2, the spectrum is adjusted to reduce variability in intensity values. In Step 3, the theoretical spectrum is compared to the experimental spectrum using a cross-correlation function $R(\tau)$. The theoretical spectrum is laterally translated by τ m/z units; the laterally translated cross-correlation values, averaged between $\tau = -75$ and $\tau = +75$, are compared to the original cross-correlation value ($R(0)$) and the difference between the two is the cross-correlation score C for that theoretical spectrum. C(best precursor) refers to the highest C value calculated among candidate precursors, whereas $C(2^{nd}$ best precursor) is the second highest.

match (the local false discovery rate, FDR; see below).[65] For the peptide in Figure 11, which has an estimated local FDR well below 1%, the experimenter can be highly confident that this peptide was eluted from the column at this time point. Accepting all peptides above a chosen score threshold then produces a final list of peptides with an estimated "global FDR" at a desired level, e.g., 1%.[66,67]

A simplified description of one of the most relevant peptide-spectrum matching algorithms, SEQUEST, is presented below.[68] Versions of SEQUEST continue to be among the most popular in the field, alongside Andromeda,[67] Mascot,[69] and others.[56,70] A simplified description of the original version is given here and the reader is referred to the original disclosure for more details;[68] however, the original algorithm is no longer in use, as improved versions have taken its place.[71] Particularly noteworthy is Comet, a free and open-source descendant of SEQUEST that is currently well maintained and relatively easy to use via the Crux platform which is also free and open-source.[72,73]

3.1. Database Search Algorithm Example: A Simplified Description of SEQUEST

SEQUEST, disclosed in a landmark publication by Eng, McCormack, and Yates in 1994, was the first fully automated peptide identification software.[56,68] Figure 12 illustrates a simplified conceptualization of the SEQUEST algorithm. This procedure is performed for each MS2 spectrum in the experimental data set. First, the peptides predicted from the reference proteome are filtered so that only precursors (i.e., charge state-specific peptide ions) with m/z values similar to that of the ion isolated for fragmentation are considered, and then theoretical spectra are generated (Figure 12, Top). Note that high-accuracy mass analyzers such as the TOF analyzer and orbitrap give a tremendous computational benefit at this step: previously, spectrometers were only accurate to within ±1 Th, leaving numerous theoretical spectra to consider.[68] Now, spectrometers are accurate to within ~20 millionths of analyte m/z or less (i.e., 20 ppm or less), generally ranging from 0.04 Th to 0.008 Th or less for peptides, so that many fewer precursors need be considered.

Theoretical spectra are quickly filtered by crudely comparing the experimental spectrum to the predicted fragments' m/z values in a procedure inspired by classical manual spectrum interpretation:[40] the numbers of matching fragments, the MS2 intensities, and other relevant features are combined into a score (Figure 12, Step 1). Only the top 500 theoretical spectra based on this score are retained. The experimental spectrum is adjusted by eliminating the precursor peak, dividing the spectrum into ten equal regions, and normalizing the intensities in each region to the same value (Figure 12, Step 2). This step makes the experimental spectrum more similar to the theoretical spectra, reducing the effects of variable fragmentation efficiencies that are neglected when constructing the theoretical spectra.

A cross-correlation is computed between the experimental spectrum and each theoretical spectrum by multiplying the intensities at each m/z value and summing the products. (The m/z values are binned/discretized in a previous step.) This correlation value, $R(0)$, is compared to the average correlation value when the spectrum is translated or "slid" 75 Th to the left and to the right (Figure 12, Step 3). The difference between $R(0)$ and this average background correlation is the final score C (now commonly referred to as XCorr), which is computed for each theoretical spectrum (precursor). ΔC (now commonly

https://doi.org/10.1021/acs.jproteome.2c00838
J. Proteome Res. 2023, 22, 2151−2171

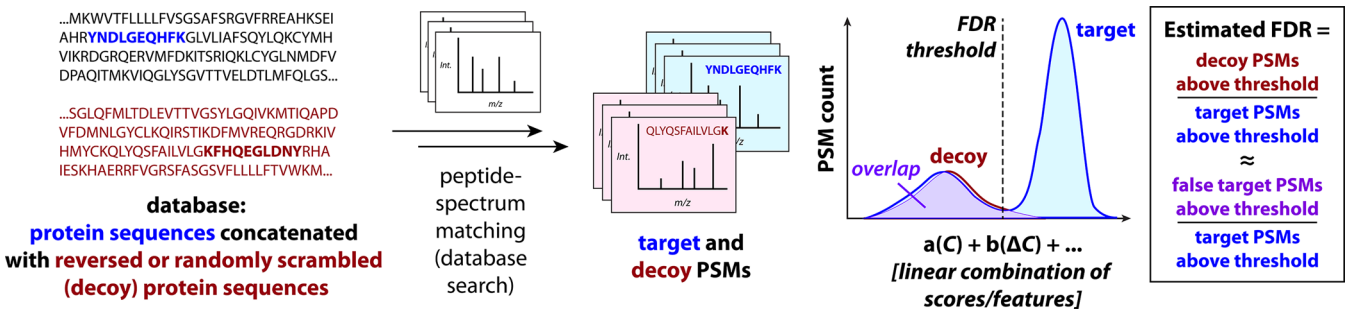


Figure 13. Target-decoy search. PSM = peptide-spectrum match. FDR = false discovery rate.

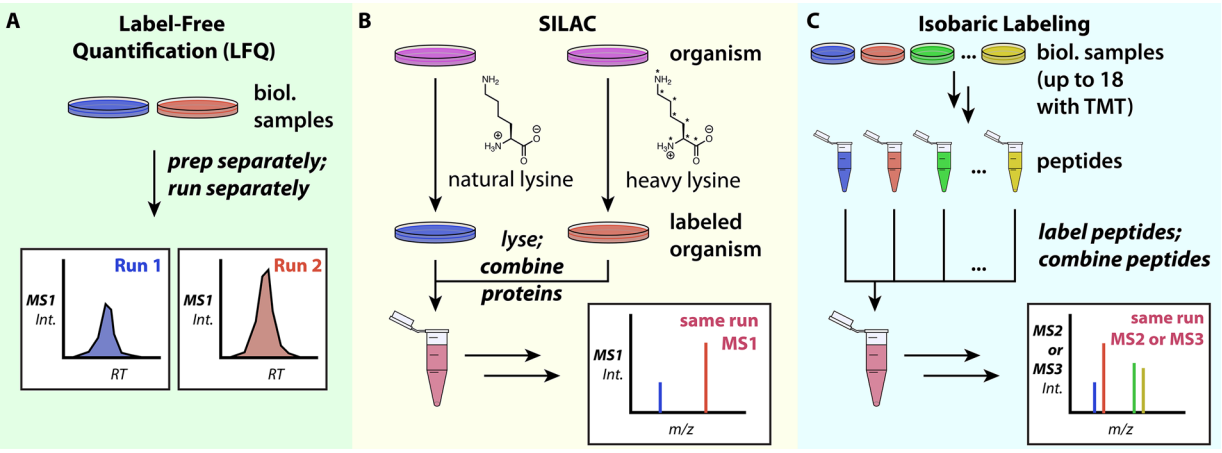


Figure 14. Peptide quantification methods. A. Label-free quantification is performed at the MS1 level (except with DIA; see the Advanced Topics and Further Reading section) between separate LC-MS/MS runs. RT = retention time. B. Stable isotope labeling with amino acids in cell culture (SILAC). An organism is fed with natural ("light") amino acids or with amino acids containing heavy isotopes before the biological experiment is performed. Once cells are lysed, the proteomes are combined, and then the comparison is performed at the MS1 level within the same run. Note: SILAC can be performed in various species, including mice (ref 79). C. Isobaric labeling. Each peptide sample (up to 18 at once with TMTs) is labeled with a tag. All tags have the same total mass, but HCD fragmentation releases a different reporter ion from each tag; these reporter ions are simultaneously observed at the MS2 or MS3 level, allowing within-run comparisons.

referred to as dCn), the difference between the highest and second-highest C value—i.e., C for the best and second-best precursor—is also calculated, as it is a helpful additional metric for thresholding/statistical scoring of the PSMs.

3.2. FDR Control in Peptide Identification

It is intuitive that higher-scoring matches are more likely to represent authentic peptide identifications. How do we decide which scores are high enough to signify a true ID? By far, the most popular approach is target-decoy search (Figure 13).[66,74] Before database search, the reference proteome is concatenated with a "decoy" proteome of false proteins, which can be generated by randomly scrambling the sequences or by reversing them (reversal has the advantages of being deterministic and preserving certain properties of the predicted tryptic peptides, such as peptide length distribution). The predicted peptides from the decoy portion of the database are not produced by the organism in real life, so any match between a spectrum and one of these decoys is assumed to occur randomly and correspond to a false ID. Importantly, it is assumed that a random nonpeptide spectrum is equally likely to match a decoy or non-decoy ("target") peptide. Therefore, given any score threshold, there are an equal number of false target PSMs above the threshold as decoy PSMs (Figure 13). The proportion of target PSMs above a score threshold that are false can therefore be estimated as the number of decoy PSMs above the threshold divided by the number of target PSMs above the threshold. Based on this estimate, a score threshold is chosen, above which target PSMs are accepted as true.

This procedure controls the global false discovery rate (FDR), i.e., the average proportion of accepted PSMs that are false. It accomplishes this regardless of what score is used; C, ΔC, or a crude sequence-peak match score such as that described in Step 1 of Figure 12 could work. However, these different scores have different abilities to detect true peptides, i.e., different abilities to segregate target PSMs from decoy PSMs. To maximize the detection of true peptides while controlling the FDR, a linear combination of search scores and other features of the PSM can be used (Figure 13).[74,75] One example of an often strongly contributing feature is precursor mass error: the difference in m/z between the observed and theoretical ion divided by theoretical ion m/z (often multiplied by 10^6 for convenience and then referred to as "ppm", parts per million). The coefficients used in the linear combination {a, b, ...} are optimized to maximize the detection of true peptides, often using a machine learning algorithm such as that used by Percolator.[74,76] After this optimization is complete, all target matches beyond the threshold corresponding to the desired FDR (e.g., 1%) are retained. The FDR for this filtered data set is called the "global" FDR. The target and decoy distributions can also be used to calculate the "local FDR" for each PSM: the estimated proportion of false target PSMs (based on decoy PSMs) among target PSMs with similar scores.[65]

https://doi.org/10.1021/acs.jproteome.2c00838
J. Proteome Res. 2023, 22, 2151−2171

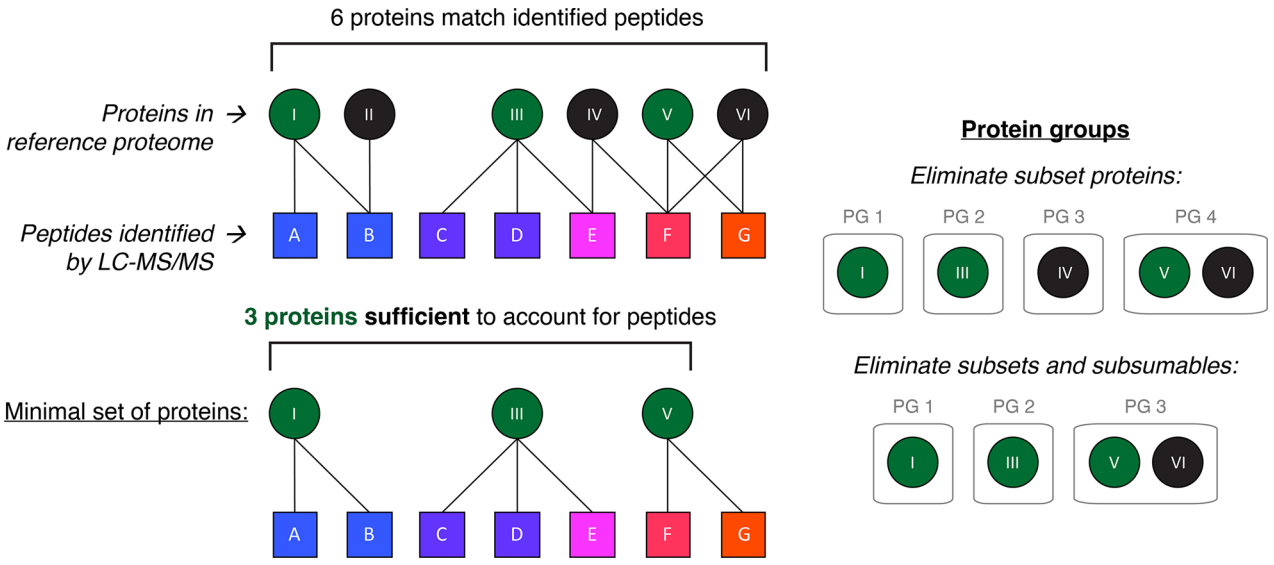


Figure 15. Protein inference and grouping strategies in a hypothetical example. In this example, 7 peptides were identified that match the sequences of 6 proteins. Black lines represent matches. Three proteins (green) are sufficient to explain the existence of all peptides in the sample, so at least 3 proteins were detected. Since proteins V and VI match the same peptides, they are equally likely to have existed in the sample(s) based on this information. So, proteomic data processing software will report them as a protein group (PG). Some algorithms, such as IDPicker, will eliminate protein IV, which is called a subsumable protein (ref 91), whereas some, such as MaxQuant, will not (ref 51). Most algorithms will eliminate protein II, whose peptide constitutes a subset of protein I's peptides.

4. PEPTIDE QUANTIFICATION

As described above, the HPLC continuously sprays peptides into the mass spectrometer throughout the run. Once per scan cycle, which usually takes 3 s or less (see above), the spectrometer produces an MS1 spectrum; a precursor's MS1 peak is therefore often observed multiple times during its elution from the HPLC column, even if the precursor is fragmented only once during the whole run. Together, over a period of seconds or minutes, these MS1 data points form a chromatographic peak (Figure 14A). A feature of the chromatographic peak, such as its height or the area under the curve (AUC), is used to measure the relative abundance of the peptide. This is called label-free quantification (LFQ).[77] Given a particular peptide, the quantities in different samples can be used to compare the samples; however, because different peptides have different ionizabilities (see above), these are only relative quantities, and MS-derived quantities of different peptides cannot be accurately compared.

To improve quantitative reproducibility, sample throughput, and/or data completeness, labeling reagents can be used to quantify multiple biological samples in the same mass spectrum. In these approaches, peptides are labeled with groups of atoms that are identical except for different isotopes that result in different m/z values, either of the peptides or of their fragments, while conserving chemical properties such as retention time, ionizability, and fragmentation pattern. In metabolic labeling methods such as stable isotope labeling with amino acids in cell culture (SILAC), the biological system is utilized to label the whole proteome with amino acids containing heavy isotopes.[78−80] The natural "light" proteins are combined with the heavy proteins at the start of sample preparation and the comparison is then performed with MS1 spectra (Figure 14B). In isobaric (meaning equal mass, because the different labels have the same total mass) labeling methods,[81,82] such as tandem mass tags (TMT),[10,83] peptides are labeled after digestion and then combined before LC-MS/MS. Higher-energy CID (HCD) releases a different reporter ion for each biological sample; the reporter ion intensities, from which relative abundances can be deduced, are measured with MS2 or MS3 scans (Figure 14C). In MS3, CID fragments the peptides without releasing reporter ions, and then fragment(s) are further fragmented to release reporter ions; more detailed descriptions as well as demonstrations of the advantages and disadvantages of MS3 compared to MS2 can be found elsewhere[84−86] and are briefly summarized in the Advanced Topics and Further Reading section. With LFQ, each run analyzes one biological sample; with SILAC, each run usually analyzes two to three biological samples; with TMT, one run can analyze up to 18 different biological samples.[87]

Many software packages that perform database search and FDR control, such as MaxQuant and Proteome Discoverer, also perform peptide quantification. Relative peptide quantities are generated by combining charge-specific precursor quantities in some way, such as taking the sum, median, mean, or geometric mean of the precursor quantities, or generating between-sample precursor quantity ratios first and then combining them.[77] The software produces a table of peptides and their relative quantities. In the relatively simple experiment described here, as discussed above, absolute abundance is not measured and intensities of different peptides are not comparable.[18] Absolute abundance can be measured using internal standards or by spiking external standards into the digest and constructing a calibration curve; targeted proteomics is best suited for this (see the Advanced Topics and Further Reading section).[88,89]

There are cases in which it is appropriate for further analysis to be performed on these "peptide-level" results before proteins are considered, e.g., if covalent binding sites, proteolytic cleavage sites, or other post-translational modification (PTM) sites are under investigation. The subsequent process of inferring information about proteins may then differ from what is described below (see the Advanced Topics and Further Reading section). The discussion here focuses on the identification and relative quantification of protein groups.

https://doi.org/10.1021/acs.jproteome.2c00838
J. Proteome Res. 2023, 22, 2151−2171

5. THE "PROTEIN LEVEL": INFERRING PROTEIN IDENTITIES AND ABUNDANCES

5.1. Protein Inference and Grouping

Although some publications using bottom-up proteomics refer to "protein" identities and quantities, they have usually identified and/or quantified protein groups, which are virtual approximations of proteins.[90] Identified peptides can be ambiguous regarding the proteins or genes from which they originated (Figure 15); the most common solutions to this ambiguity use each PSM as a piece of evidence that the corresponding protein(s) existed in the sample(s). If there is equal evidence among a group of proteins (i.e., the protein sequences match the same set of identified peptide sequences, e.g., proteins V and VI in Figure 15), these are grouped together in a protein group (PG). If a protein's matched peptides are a subset of those of another protein, this protein is usually excluded from the report, while proteins that lack a peptide unique to that protein but whose peptides are not a subset—called subsumable proteins (see protein IV in Figure 15)—are excluded by some algorithms, such as IDPicker,[91] but retained in others, such as MaxQuant.[51,92] The final number of PGs thus serves as a conservative estimate of the number of proteins identified in the experiment.

Another less common category of approaches uses the PSM probabilities to calculate a probability for each protein. These probabilistic approaches perhaps more accurately represent the nonbinary evidence for the proteins and do not necessarily require protein grouping, except in cases of identically connected proteins,[93,94] and they can also be combined with more sophisticated protein grouping strategies.[95]

5.2. Protein-Level FDR Control

Most popular software packages that perform protein grouping, such as MaxQuant and Proteome Discoverer, can also calculate an estimate of the protein-level identification FDR and readjust it to an acceptable level. There are multiple approaches to protein-level FDR control.[96,97] One straightforward strategy makes use of decoy PGs, i.e., PGs consisting entirely of decoy PSMs, which are automatically retained and subjected to the same protein inference process as target PSMs. A score is constructed for each target and decoy PG, allowing the protein-level FDR to be controlled in a similar way as the PSM FDR is controlled (Figure 13).[96−98] One approach to constructing the PG score is to use PSM "q-values" (in this context, the PSM q-value is defined as the minimum estimated global FDR at which the PSM will appear in the filtered output list): by taking a negative logarithm of each q-value to create a "Q-score," the maximum PSM Q-score within the PG can be used as the PG score.[96] Alternatively, MaxQuant multiplies the posterior error probabilities (PEPs) of the peptides in the PG (the PEP is conceptually related to the local FDR described above).[98] Another approach is to find the lowest PSM local FDR for each peptide in the PG, multiply together all these peptide-level minimal local FDRs to estimate the probability that the PG is false, and then subtract that probability from 1.[65,75,99]

Unlike with PSMs, the number of false target PGs is not necessarily similar to the number of decoy PGs above a given score threshold, because a true target PG can contain both true and false PSMs whereas a decoy PG only contains decoy (false) PSMs;[100] assuming this equality, therefore, results in an inflated estimated protein-level FDR, especially in large data sets where a significant portion of the proteome is covered by identified PGs.[96] There are multiple strategies for addressing this.[96,100] One simple and scalable method pairs each target PG with its decoy version, and the version with the lower PG score is discarded.[96] This "picked protein FDR" approach allows the assumption of equal numbers of decoy PGs and false target PGs by eliminating most decoy PGs paired with true target PGs while maintaining the balance between identified decoy PGs and falsely identified target PGs. By adjusting the PG score threshold and accepting PGs with scores above the threshold (as with PSMs in Figure 13), the estimated protein-level FDR can then be adjusted to the desired level (e.g., 1%).

5.3. Protein Group Quantification

Relative PG quantities are generated using peptide intensities; multiple aspects of the relationship between peptides and proteins complicate this process. Peptides within a PG may differ from each other in the proteins they match, which can affect their abundances (Figure 15).[101] It is helpful if peptides in a PG are unique to that PG; some may even match a specific protein and no others ("proteotypic," e.g., peptides A, C, and D in Figure 15), allowing quantification of a singular protein (assuming the FASTA file completely describes the proteome). To complicate matters further, however, a peptide may arise from a proteoform,[102] e.g., containing a post-translational modification, whose abundance affects the peptide's intensity.[101] In addition, some peptides may be observed in one sample but not another because of the inherent stochasticity of DDA.

Popular data processing software packages present options to the user to address these issues. Most, including MaxQuant, allow quantification to be performed only with PG-unique peptides. Some, such as Spectronaut (Biognosys), allow more specifications, such as restricting quantification to protein-specific (proteotypic) peptides or choosing whether to use the median, sum, arithmetic mean, or geometric mean of peptide intensities to calculate the PG quantity. For LFQ, the popular MaxLFQ algorithm implemented in MaxQuant is of note. When different sets of precursors are identified in different samples, MaxLFQ deals with the missing values by taking available pairwise comparisons and using median ratios to compare PGs. The reader is directed to the original publication for more details.[77] Isobaric labeling and data-independent acquisition (DIA) are other ways of reducing or eliminating missing values (see the Advanced Topics and Further Reading section). A recent method that groups peptides by quantitative changes in addition to sequences may help address the issue of proteoform-specific peptides.[101]

The output of a data processing software package that performs PG quantification is a table containing PGs and their relative quantities. The quantities of different PGs within a sample are not comparable because of varying relationships between peptide abundances and precursor intensities (see above). However, the relative quantities of a PG can often be accurately compared between samples to test a biological hypothesis.

5.4. Statistical Analysis and Biological Interpretation

Once relative PG quantities are calculated, various tools can be used to statistically analyze and interpret the results depending on the biology of the experiment. Perseus[103] and MSStats[104] are popular packages designed for processing and statistical analysis of proteomic quantities. Data can also be analyzed manually using a programming language such as python or R. In a generic comparative proteomic experiment, this will consist of removing

https://doi.org/10.1021/acs.jproteome.2c00838
J. Proteome Res. 2023, 22, 2151−2171

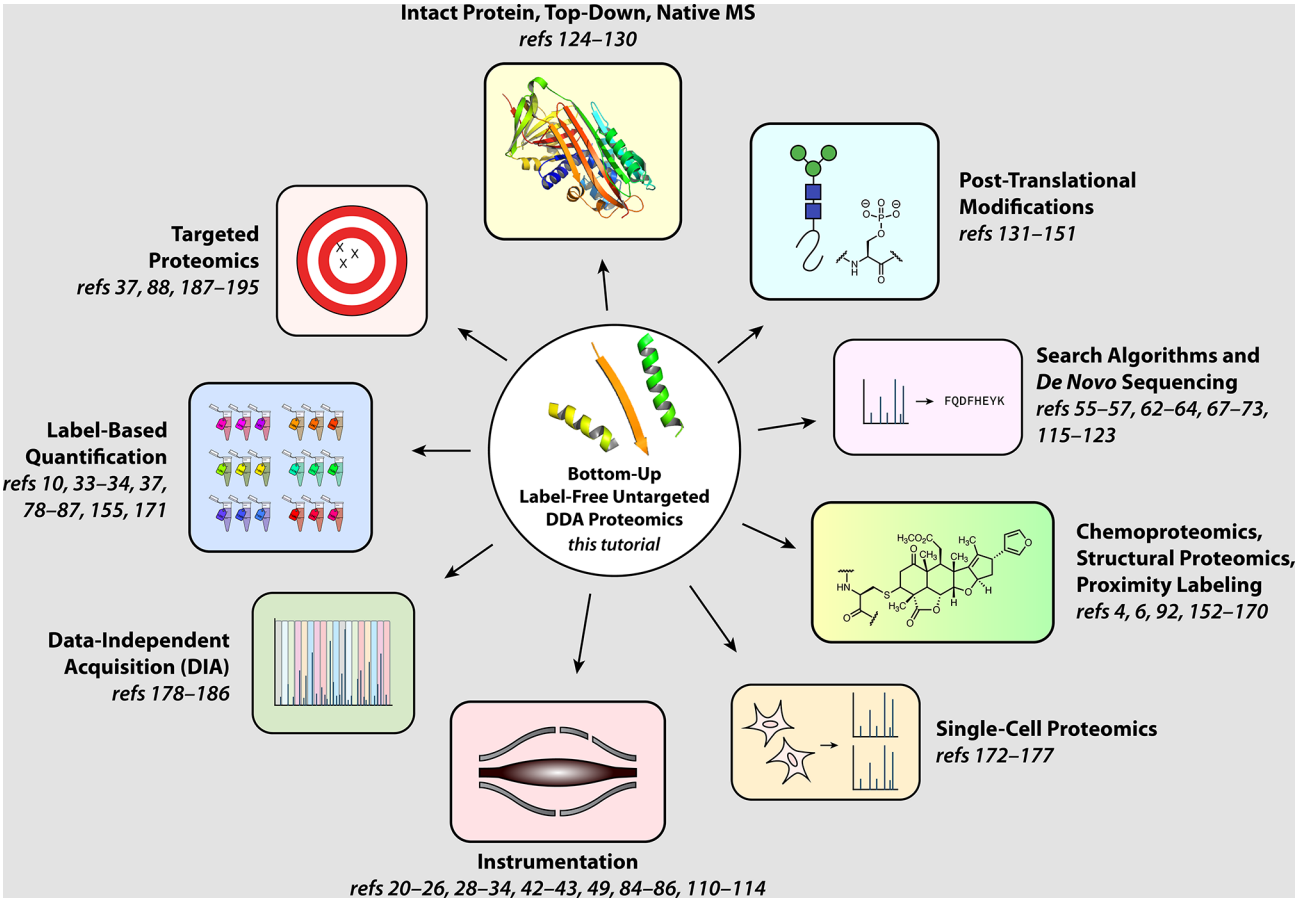


Figure 16. Selected subfields of mass spectrometry-based proteomics. In many cases, the technologies used in these subfields are extensions of the same technologies detailed in this tutorial, used in bottom-up label-free untargeted DDA.

contaminants and/or decoys from the list, applying a logarithm to give the quantities an approximately normal distribution, normalizing the data to correct for inter-run technical variability, testing for changes using statistical tests such as *t* test/ANOVA/linear regression, and correcting the resulting *p*-values for multiple hypothesis testing.[105]

After statistical analysis is complete, the results can be interpreted biologically. The Gene Ontology database is a popular resource for studying biological pathways that are altered in the experiment.[106] The search tool for recurring instances of neighboring genes (STRING)[107,108] gives information about known relationships or interactions between proteins of interest. Correlational analyses such as weighted gene correlation network analysis (WGCNA)[109] construct a network of proteins based on how similarly their abundances change in the experiment, which can give further information about how proteins in the experiment relate to each other. By interpreting the results of these analyses together, biological hypotheses can be supported or refuted, and new hypotheses can be generated.

6. SUMMARY: THE WORKFLOW FROM BEGINNING TO END

In total, the workflow for an untargeted label-free bottom-up mass spectrometry experiment is as follows. Detailed experimental protocols can be found in the References, as can more specialized workflows, which are discussed in the Advanced Topics and Further Reading section.

1. **Dissociate** tissues/plants; **lyse cells**; **denature proteins** with, e.g., urea and/or detergents.
2. **Reduce disulfides** with, e.g., DTT or TCEP; **alkylate cysteines** with, e.g., IAA.
3. **Partially purify proteins** with, e.g., cold acetone, chloroform−methanol precipitation, or magnetic bead-based precipitation. (Some workflows skip this because they do not use harmful detergents such as SDS.)
4. **Digest proteins** with a protease or combination of proteases, e.g., trypsin and LysC.
5. **Clean/desalt peptides** with, e.g., a C18-coated filter pipet tip using centrifugation; **evaporate solvent** with, e.g., a vacuum centrifuge.
6. **Resuspend peptides** in an LCMS-compatible solvent, e.g., 0.1% formic acid in water, in an LCMS-compatible vial; **run LC-MS/MS** using the appropriate instrument method. Retrieve the raw data after the run(s) are complete.
7. **Process raw data** using a suitable software package, such as MaxQuant or Proteome Discoverer, or a combination of packages. Perform database search, FDR control, relative quantification, and protein inference.
8. **Perform statistical analysis** using a programming language such as python or R or using a software package such as Perseus or MSStats.
9. **Interpret the results** according to the biology of the experiment. This can be aided by resources for pathway- or network-level analysis such as Gene Ontology, STRING, or WGCNA.

https://doi.org/10.1021/acs.jproteome.2c00838
J. Proteome Res. 2023, 22, 2151−2171

7. ADVANCED TOPICS AND FURTHER READING

This tutorial has described the workflow for an untargeted bottom-up label-free proteomic experiment using DDA. There are many more MS-based approaches to studying the proteome, and several subfields of MS-based proteomics are advancing in order to better answer a wider array of biological questions. A selection of these is illustrated in Figure 16 and discussed below. A deeper and broader grounding in several of these topics, and others, can also be gained via a 2013 review by Zhang, Yates et al.,[3] while a 2016 review by Aebersold and Mann focuses more on specialized applications, including most of those discussed below.[4]

7.1. Instrumentation

Publications describing the inner workings of orbitrap-[23−26,28,29] and TOF-based[20−22] instruments were referenced in the Experimental Workflow section. The Q Exactive and timsTOF use beam-type CID, briefly described above, in which a beam of ions (peptides) is accelerated into a chamber of gas (often N_2);[42,43] Xia et al. give a more detailed description.[42] Ion trap CID, in which trapped ions are accelerated by an oscillating voltage in the presence of a gas such as helium,[110] is also useful, e.g., for isobaric label analysis using MS3 (see above).[84−86] Macias, Stantos, and Brodbelt give a thorough review of additional fragmentation methods, including electron transfer dissociation (ETD), electron capture dissociation (ECD), ultraviolet photodissociation (UVPD), infrared multiphoton dissociation (IRMPD), and surface induced dissociation (SID).[111] Ion mobility, an increasingly relevant concept in instrumentation, is briefly described above for the timsTOF; more details, including descriptions of FAIMS and its uses, can also be found elsewhere.[32−34,112,113] Fundamental instrumentation concepts can be further explored in a tutorial by Savaryn, Toby, and Kelleher.[114]

7.2. Search Algorithms and *De Novo* Sequencing

The above description of SEQUEST included citations of works describing several other SEQUEST-related search algorithms;[56,68,71,72] other popular search algorithms, such as Andromeda,[67] Mascot,[69] and Morpheus,[70] may also be of interest. Eng et al. give a discussion of various database search algorithms,[55] while Hoopmann and Moritz survey non-SEQUEST-related approaches.[57] In general, popular modern database search algorithms have similar levels of performance.[115] *De novo* sequencing algorithms,[116] especially PEAKS,[117] have proven impressively useful in experiments for which there is no reference proteome, such as in immunopeptidomics,[62] antibody characterization,[118] identification of novel modifications,[63] or studies using peptide libraries generated by split-pool synthesis.[64] Other MS2 interpretation approaches are designed for specialized applications, such as analyzing data sets with different types of spectra[119] or considering many different peptide modifications at once, using either *de novo* sequencing with PTM-optimized probabilistic scoring (the TagGraph method)[63] or "open search," in which very wide precursor m/z tolerances are used (e.g., MSFragger).[120,121] Increasingly, database search is being augmented with the use of deep learning-based models that predict MS2 spectra and retention times; Wen et al. give an overview of these, including prominent implementations such as Prosit and DeepRescore,[122] whereas Mann et al. discuss the use of artificial intelligence in proteomics more broadly.[123]

7.3. Intact Protein, Top-Down, and Native Mass Spectrometry

Alternatives to bottom-up proteomics include intact protein mass spectrometry, top-down proteomics, and native mass spectrometry. These are typically applied to lower-complexity samples in which a target protein or complex of interest has been isolated. In its simplest form, intact protein analysis is a technique used to accurately measure the mass of a purified protein.[124] More information about a protein, such as the positions of PTMs, can be gained by fragmentation; intact protein fragmentation, and the subsequent extraction of information from the resulting fragment spectra, is called top-down proteomics. Top-down proteomics has been powerfully applied in several studies, and detailed experimental guidelines for its application are available.[124] Intact protein MS and top-down proteomics are often applied to denatured proteins, but they can also be performed on a protein or complex in the native state; this technique, called native mass spectrometry, is reviewed by Jooß, McGee, and Kelleher.[125] The single-molecule sensitivity of Fourier transform-based mass analyzers/detectors[126] has recently enabled discrimination between individual ion particles in the orbitrap, allowing detailed characterization of viral particles, whole nucleosomes, and single native protein molecules with masses up to 466 kDa.[127−129] This has been applied to the imaging of tissues, enabling the spatial mapping of particular proteoforms' relative abundances.[130]

7.4. Post-Translational Modifications

Post-translational modifications (PTMs) are essential to the function and dysfunction of numerous biological systems. Because each type of PTM is studied by its own field of biologists, reviews do not typically cover all PTMs, but a 2003 review by Mann and Jensen is an effective starting point from the perspective of mass spectrometry.[131] A thorough review of PTM chemistry published in 2005 may also be of interest.[132] However, novel PTMs continue to be discovered within various fields.[133] One of the most well-studied PTMs is phosphorylation, which has critical relevance in cancer and many other fields of biology, largely due to its importance in intracellular signaling. Most relevant technical aspects of phosphoproteomics are discussed by Riley and Coon,[48] though improvements continue to be made.[9,134] Mass spectrometric analyses of ubiquitylation, another widely relevant PTM especially important for protein degradation, are also under development; Udeshi et al. have published protocols for these types of studies.[135,136]

One intensely studied PTM is glycosylation. Glycobiology is its own vast field with relevance to cancer, immunology, and many other biological subfields. Glycosylation can occur on a variety of amino acid residues, and glycans can assume a diverse array of varyingly branched structures.[137] Some fundamentals of glycoproteomics have now been established:[138] enrichment of glycopeptides or glycoproteins is often performed[139−143] and relatively gentle electron transfer dissociation (ETD) is frequently employed, often in conjunction with HCD.[139,141,144,145] Recently, digestion enzymes have also been developed to enable the detection of challenging glycans such as *O*-linked glycans on mucins, which are resistant to tryptic digestion.[146−148] Whereas other PTMs with constant atomic composition, such as phosphorylation, require the consideration of few additional possible fragments in database search, peptide-spectrum matching in glycoproteomics is greatly complicated by the various structures of glycans; search algorithms continue to

https://doi.org/10.1021/acs.jproteome.2c00838
J. Proteome Res. 2023, 22, 2151−2171

be developed and improved.[121,149,150] All steps in the glycoproteomic workflow, from digestion and enrichment to raw data acquisition and processing, continue to undergo improvements to further illuminate glycobiology.[151]

7.5. Structural Proteomics, Chemoproteomics, and Proximity Labeling

Using chemical, enzymatic, or thermal techniques to probe proteins opens a large swath of opportunities for studying protein structure, chemistry, localization, protein−ligand interactions, or protein−protein interactions. Aebersold and Mann give introductions to some of these techniques.[4] A 2018 review by Leitner is a thorough introduction to techniques that use chemical probes specifically.[152]

Chemical and enzymatic probes can be used to reveal structural biology of proteins. Compared to X-ray crystallography and cryoelectron microscopy, mass spectrometry tends to give less complete and/or lower-resolution structural information, but it has several advantages: (1) it is performed in solution, in the native state, or even in cells or tissues; (2) a readout is obtained whether or not the protein(s) can be crystallized or vitrified in an orderly set of states; and (3) it can be performed on a proteome-wide scale, yielding information about thousands of proteins at once.[153] Liu, Zhang, and Gross give a thorough review of hydrogen−deuterium exchange and covalent labeling-based footprinting, in which the solvent-exposed surfaces of native proteins are mapped;[154] new approaches continue to emerge.[155] O'Reilly and Rappsilber review cross-linking mass spectrometry (CLMS/XLMS), which uses bifunctional bioconjugation reagents to create covalent links between proximal amino acids.[156] Limited proteolysis-mass spectrometry (LiP-MS) is a relatively robust and easy-to-use method for probing protein structure as well as small molecule-protein binding, using proteolytic enzymes to map the proteolytic vulnerabilities of peptide bonds.[6,92,157−160]

Mass spectrometry is increasingly being used to study compound−protein interactions for small-molecule drug discovery and development, a field concisely referred to as chemoproteomics.[161] The majority of studies in this field use one of two tools: (1) compounds that react with proteins, forming a covalent bond that can be detected by MS, often via either electrophilic functional groups or photoreactive carbene or nitrene precursors; or (2) thermal profiling, in which compound-induced changes in protein melting (unfolding) temperature can be measured by MS via centrifugation of protein aggregates. Drewes and Knapp efficiently reviewed both approaches in 2018.[162] More recent state-of-the-art examples of the use of electrophiles, often called activity-based protein profiling (ABPP), have come out of the laboratories of Nomura,[163] Gygi,[164] and Cravatt.[165] The use of photoreactive groups, called photoaffinity labeling (PAL), is also an active area of research; current approaches are covered in a recent review by West and Woo.[166] Thermal proteome profiling (TPP) is also currently evolving into higher-throughput methods such as proteome integral solubility alteration (PISA).[167,168]

In structural proteomics and chemoproteomics, chemical reactions are usually performed during sample preparation to yield additional information. In a separate category, the biological system itself adds chemical labels to proteins (*in situ* or *in vivo*), usually to yield information about localization or transport. A recent review by Brewer, Shi, and Wyss-Coray surveys different methods of protein labeling and tracking *in vivo* with emphasis on the brain.[169] The most popular approach in this category is proximity labeling, in which a genetically encoded enzyme labels nearby proteins. Proximity labeling has been developed using several different enzymatic constructs and has been used in many different biological systems; a thorough recent review by Qin et al. concisely surveys these approaches.[170]

7.6. Label-Based Quantification and Single-Cell Proteomics

The Peptide Quantification section discusses some label-based quantification methods and cites works that give more details, including reviews/comparisons[80−82] and state-of-the-art methods.[10,79,86,87] One approach not discussed above is the use of chemical (as opposed to metabolic) non-isobaric stable isotope labeling reactions such as lysine dimethylation. This is a low-cost alternative to isobaric labeling especially helpful in pairwise comparisons where SILAC is not feasible.[171] A recent report demonstrated the use of such a label as a footprinting reagent for structural biology, fruitfully combining these concepts.[155]

One popular isobaric labeling technology, tandem mass tag (TMT) labeling, has undergone major developments since its inception[83] and is still being improved. MS3 greatly increases quantitative accuracy by reducing interference,[84] but the original MS3 implementation only selected one CID fragment for MS3-level fragmentation, reducing the signal-to-noise ratio of TMT reporter ions relative to MS2; also, the addition of a third scan type in the cycle slowed acquisition, reducing proteomic depth. The former disadvantage was addressed by synchronous precursor selection (SPS), which allows multiple CID fragments to be selected and undergo HCD together, combining TMT reporter signals for greater signal-to-noise ratios.[85] The second issue (time efficiency) is partially ameliorated with real-time search (RTS), which excludes non-peptide and contaminant ions from MS3.[86] The acquisition mode combining these innovations, "RTS-SPS-MS3," has superior quantitative accuracy to MS2 while achieving comparable proteomic depth.[86] In both MS2 and MS3 methods, quantitative accuracy can be improved with FAIMS.[33,34]

One application of TMT is in single-cell proteomics: the first publication describing the SCoPE-MS workflow for single-cell MS used the signal-boosting effect of TMT sample pooling to cross the single-cell threshold of sensitivity.[172] The original disclosure of the nanoPOTS approach to single-cell MS, published the same year (2018),[173] was quickly followed by considerable improvements in proteomic depth using TMT.[174] An overview of the concepts was given in a 2020 review by Kelly,[175] but the field continues to evolve: recent work by Brunner et al. describes the use of the highly sensitive timsTOF instead of TMT for single-cell MS.[176] Kelly details the conditions necessary for successful single-cell MS, including miniaturized sample preparation and decreased HPLC column inner diameter and flow rate.[175] A recent review by Vistain and Tay discusses single-cell MS alongside other non-MS techniques for single-cell proteomics.[177]

7.7. Data-Independent Acquisition (DIA)

As discussed above, this tutorial has detailed a DDA approach, in which the fragmentation of peptides is dependent upon their detection in MS1 spectra. Data-independent acquisition (DIA), which excludes such data-dependent scans, is an increasingly popular category of acquisition methods. The most popular DIA methods are based on SWATH-MS, in which all m/z values within the whole MS1 range are included in fragmentation every scan cycle using wide isolation windows;[178] for LFQ, this greatly improves data completeness and can increase proteomic depth.

https://doi.org/10.1021/acs.jproteome.2c00838
J. Proteome Res. 2023, 22, 2151−2171

Quantification can also be improved over DDA-based LFQ, because repeated MS2 scans allow each fragment to form its own chromatographic peak; combining these fragment chromatographic peaks gives richer quantitative information for each precursor. Initially, a disadvantage of SWATH-MS was the requirement of a DDA-based spectral library, but library-free approaches are increasingly used.[179] Ludwig et al. give a thorough tutorial on the subject.[180] A recent overview by Zhang et al. describes several DIA methods,[181] including non-SWATH-based concepts such as BoxCar,[182] in which no MS2 spectra are acquired, and PAcIFIC,[183] in which DDA-like ion isolation and MS2 are performed without MS1. Both of these methods dramatically increase proteomic depth, at the expense of identification confidence for BoxCar and time and material for PAcIFIC. As with DDA, DIA data processing increasingly benefits from deep learning-based models such as those included in the DIA-NN software, a free, open-source, and easy-to-use software suite for DIA data analysis (NN stands for neural networks).[184,185] Recently, DIA-NN has been used in a pipeline that combines DIA with three-plex non-isobaric labeling, increasing the sample throughput of DIA.[186]

7.8. Targeted Proteomics

In targeted proteomics, a specific protein or set of proteins is targeted for analysis.[187] As an alternative to untargeted proteomics, the goals of targeted proteomics are to ensure detection of the target protein(s) every time the method is run (high data completeness), to maximize sensitivity and dynamic range, and to maximize quantitative accuracy and precision. Traditionally, targeted proteomics is performed using single-reaction monitoring (SRM),[188] multiple reaction monitoring (MRM),[189] or parallel reaction monitoring (PRM).[190] Here, "reaction" refers to fragmentation, i.e., as the peptide(s) elute, they are repeatedly isolated and fragmented and MS2 fragment peak(s) are "monitored" (intensities measured). The resulting MS2 peaks are used for both identification (detection) and quantification.

Because FDR control such as that described in the Peptide Identification section is not generally feasible with targeted proteomics, evaluating targeted peptide detection usually requires a comparison to a known fragmentation pattern. In addition, to target many peptides efficiently in the same run, MS2 acquisition must be limited to a specific time (RT) window; the peptide RT must therefore be known beforehand (known RT is also helpful for identification confidence). Known fragmentation pattern and RT often come from synthetic peptides, especially peptides containing heavy stable isotopes, that can be used for assay development and/or as standards.[88,191] The RT window can become inaccurate throughout an experiment as, e.g., the HPLC column wears; recent real-time retention time adjustment methods can dynamically correct RT windows.[192,193] To avoid the initial manual development of assays, a method has also been developed to detect synthetic peptides on the fly, triggering MS2 in real time (and MS3 for TMT analysis).[194] The recent "GoDig" TMT method does not require synthetic peptides; instead, it uses RTS with an elution library to trigger monitoring in real time and uses a spectral library to identify targets and trigger MS3 scans (quantification events).[37] As the number of feasible targets in targeted proteomics grows to over 1000 peptides per run,[37,192,193] automatic detection/identification FDR control may become necessary; this has been achieved with decoy SRM transitions,[195] though this type of strategy has not yet been widely adopted.

8. OUTLOOK

For large-scale protein identification and quantification, mass spectrometry is currently the most popular method, but alternative technologies are steadily advancing. Antibody-[196] and aptamer-based[197] protein detection have been fully commercialized,[198,199] and "next-generation protein sequencing" technologies based on, e.g., amino acid-wise side chain identification or nanopore-type sequencing are making considerable progress.[200−205] Each of these technologies has advantages, such as dynamic range, proteoform specificity, or proteomic depth; some will likely bring welcome decreases in the difficulty and cost of proteomic studies. However, mass spectrometry will probably continue to have advantages in the study of proteins. In contrast to single-molecule technologies, mass spectrometers can identify millions of copies of a molecule in milliseconds or less, quickly clearing the way for the next peptide to be identified.[206] Among bottom-up approaches, varying ionizabilities also increase the protein throughput of MS, because each protein has a limited number of observable peptides; conversely, on a peptide array, every digestion product molecule could potentially occupy a site.[204] Protein-specific probe-based methods may require development of a new set of probes to study each additional species, whereas the flexibility of mass spectrometry allows relatively easy switching between studies of different organisms (where reference proteomes are available).

Perhaps the most lasting advantages of mass spectrometers are in studies that go beyond protein abundance. MS easily detects an ever-growing set of covalent tags that reveal protein chemistry, structure, or function, whereas a tag may not fit through a pore or may preclude fluorescence-based side chain identification. Flexible and specific mass spectrometric detection of modifications allows studies of spatiotemporal dynamics, PTMs, ligand−host binding, protein−protein interactions, protein aggregation, and many other aspects of protein chemistry and biology. In the future, the usage of mass spectrometry in combination with other non-MS technologies is likely to be a powerful framework for deeper illumination of proteomes across biology.

AUTHOR INFORMATION

Corresponding Author

Steven R. Shuken − *Department of Cell Biology, Harvard Medical School, Boston, Massachusetts 02115, United States; Department of Chemistry, Stanford University, Stanford, California 94305, United States; Department of Neurology and Neurological Sciences, Stanford University School of Medicine, Palo Alto, California 94304, United States; Wu Tsai Neurosciences Institute, Stanford University, Stanford, California 94305, United States;* orcid.org/0000-0002-2782-2165; Email: steven_shuken@hms.harvard.edu

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jproteome.2c00838

Notes

The author declares no competing financial interest.

https://doi.org/10.1021/acs.jproteome.2c00838
J. Proteome Res. 2023, 22, 2151−2171

■ ACKNOWLEDGMENTS

The author thanks researchers at the Gygi Laboratory (Harvard Medical School), Wyss-Coray Laboratory (Stanford University), Wernig Laboratory (Stanford University), Cole Laboratory (Harvard Medical School), Bristol Myers Squibb, and Elias Laboratory (Chan-Zuckerberg Biohub), as well as Susan Weintraub (University of Texas Health Science Center, San Antonio), for feedback and support. The Stanford Center for Molecular Analysis and Design (CMAD) is thanked for funding and support. The example analysis shown in Figure 4, Figure 6, Figure 10, and Figure 11 was performed using Gygi Laboratory instruments on a sample prepared in the Gygi Laboratory by Divya Damacharla. Figure 10 and Figure 11 were generated using the Gygi Laboratory's in-house mass spectrometry data analysis software.

■ REFERENCES

(1) Mallick, P.; Kuster, B. Proteomics: a pragmatic perspective. *Nat. Biotechnol.* **2010**, *28*, 695−709.
(2) Sinha, A.; Mann, M. A beginner's guide to mass spectrometry-based proteomics. *Biochem. (London)* **2020**, *42*, 64−69.
(3) Zhang, Y.; Fonslow, B. R.; Shan, B.; Baek, M.-C.; Yates, J. R., III Protein Analysis by Shotgun/Bottom-up Proteomics. *Chem. Rev.* **2013**, *113*, 2343−2394.
(4) Aebersold, R.; Mann, M. Mass-spectrometric exploration of proteome structure and function. *Nature* **2016**, *537*, 347−355.
(5) Hughes, C. S.; Krijgsveld, J.; et al. Ultrasensitive proteome analysis using paramagnetic bead technology. *Mol. Sys. Biol.* **2014**, *10*, 757.
(6) Schopper, S.; Picotti, P.; et al. Measuring protein structural changes on a proteome-wide scale using limited proteolysis-coupled mass spectrometry. *Nat. Protoc.* **2017**, *12*, 2391−2410.
(7) HaileMariam, M.; et al. S-Trap, an Ultrafast Sample-Preparation Approach for Shotgun Proteomics. *J. Proteome Res.* **2018**, *17*, 2917−2924.
(8) Smith, P. K.; et al. Measurement of Protein Using Bicinchoninic Acid. *Anal. Biochem.* **1985**, *150*, 76−85.
(9) Bekker-Jensen, D. B.; Bernhardt, O. M.; Hogrebe, A.; Martinez-Val, A.; Verbeke, L.; Gandhi, T.; Kelstrup, C. D.; Reiter, L.; Olsen, J. V. Rapid and site-specific deep phosphoproteome profiling by data-independent acquisition without the need for spectral libraries. *Nature Comm.* **2020**, *11*, 787.
(10) Navarrete-Perea, J.; Yu, Q.; Gygi, S. P.; Paulo, J. A. Streamlined Tandem Mass Tag (SL-TMT) Protocol: An Efficient Strategy for Quantitative (Phospho)proteome Profiling Using Tandem Mass Tag-Synchronous Precursor Selection-MS3. *J. Proteome Res.* **2018**, *17*, 2226−2236.
(11) Griffin, P. R.; et al. The Amino Acid Sequence of the Sex Steroid-binding Protein of Rabbit Serum. *J. Biol. Chem.* **1989**, *264*, 19066−19075.
(12) Yen, C.-Y.; et al. Improving Sensitivity in Shotgun Proteomics Using a Peptide-Centric Database with Reduced Complexity: Protease Cleavage and SCX Elution Rules from Data Mining of MS/MS Spectra. *Anal. Chem.* **2006**, *78*, 1071−1084.
(13) Glatter, T.; Ludwig, C.; Ahrné, E.; Aebersold, R.; Heck, A. J. R.; Schmidt, A. Large-Scale Quantitative Assessment of Different In-Solution Protein Digestion Protocols Reveals Superior Cleavage Efficiency of Tandem Lys-C/Trypsin Proteolysis over Trypsin Digestion. *J. Proteome Res.* **2012**, *11*, 5145−5156.
(14) Hakobyan, A.; Schneider, M. B.; Liesack, W.; Glatter, T. Efficient Tandem LysC/Trypsin Digestion in Detergent Conditions. *Proteomics* **2019**, *19*, No. 1900136.
(15) Cavanagh, J.; Benson, L. M.; Thompson, R.; Naylor, S. In-Line Desalting Mass Spectrometry for the Study of Noncovalent Biological Complexes. *Anal. Chem.* **2003**, *75*, 3281−3286.
(16) Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. Electrospray Ionization for Mass Spectrometry of Large Biomolecules. *Science* **1989**, *246*, 64−71.
(17) Liuni, P.; Wilson, D. J. Understanding and optimizing electrospray ionization techniques for proteomic analysis. *Expert Rev. Proteomics* **2011**, *8*, 197−209.
(18) Jarnuczak, A. F.; et al. Analysis of Intrinsic Peptide Detectability via Integrated Label-Free and SRM-Based Absolute Quantitative Proteomics. *J. Proteome Res.* **2016**, *15*, 2945−2959.
(19) Yuan, L.; et al. Systematic evaluation of the root cause of non-linearity in liquid chromatography/tandem mass spectrometry bioanalytical assays and strategy to predict and extend the linear standard curve range. *Rapid Commun. Mass Spectrom.* **2012**, *26*, 1465−1474.
(20) Boesl, U. Time-of-flight mass spectrometry: Introduction to the basics. *Mass Spectrom. Reviews* **2017**, *36*, 86−109.
(21) Meier, F.; Mann, M.; et al. Parallel Accumulation−Serial Fragmentation (PASEF): Multiplying Sequencing Speed and Sensitivity by Synchronized Scans in a Trapped Ion Mobility Device. *J. Proteome Res.* **2015**, *14*, 5378−5387.
(22) Meier, F.; Cox, J.; Mann, M.; et al. Online Parallel Accumulation-Serial Fragmentation (PASEF) with a Novel Trapped Ion Mobility Mass Spectrometer. *Mol. Cell. Proteomics* **2018**, *17*, 2534−2545.
(23) Hu, Q.; et al. The Orbitrap: a new mass spectrometer. *J. Mass Spectrom.* **2005**, *40*, 430−443.
(24) Eliuk, S.; Makarov, A. Evolution of Orbitrap Mass Spectrometry Instrumentation. *Annu. Rev. Anal. Chem.* **2015**, *8*, 61−80.
(25) Makarov, A. Electrostatic Axially Harmonic Orbital Trapping: A High-Performance Technique of Mass Analysis. *Anal. Chem.* **2000**, *72*, 1156−1162.
(26) *Exactive Series Operating Manual*, Revision A; Document ID: BRE0012255; Thermo Fisher Scientific, April 2017.
(27) Cooks, R. G.; Rockwood, A. L. The 'Thomson'. A Suggested Unit for Mass Spectroscopists. *Rapid Commun. Mass Spectrom.* **1991**, *5*, 92−93.
(28) "Orbitrap." Mass Spec Pro. URL: http://www.massspecpro.com/mass-analyzers/orbitrap (accessed March 19, 2023).
(29) "Q Exactive Plus (Q)." UWPR. URL: https://proteomicsresource.washington.edu/instruments/qexactiveplus.php (accessed March 19, 2023).
(30) Kelstrup, C. D.; Olsen, J. V.; et al. Performance Evaluation of the Q Exactive HF-X for Shotgun Proteomics. *J. Proteome Res.* **2018**, *17*, 727−738.
(31) "Orbitrap Exploris 480." UWPR. URL: https://proteomicsresource.washington.edu/instruments/orbitrapexploris480.php (accessed March 19, 2023).
(32) Kolakowski, B. M.; Mester, Z. Review of applications of high-field asymmetric waveform ion mobiligy spectrometry (FAIMS) and differential mobility spectrometry (DMS). *Analyst* **2007**, *132*, 842−864.
(33) Pfammatter, S.; Bonneil, E.; Thibault, P. Improvement of Quantitative Measurements in Multiplex Proteomics Using High-Field Asymmetric Waveform Spectrometry. *J. Proteome Res.* **2016**, *15*, 4653−4665.
(34) Schweppe, D. K.; Rusin, S. F.; Gygi, S. P.; Paulo, J. A. Optimized Workflow for Multiplexed Phosphorylation Analysis of TMT-Labeled Peptides Using High-Field Asymmetric Waveform Ion Mobility Spectrometry. *J. Proteome Res.* **2020**, *19*, 554−560.
(35) Kreimer, S.; et al. Advanced Precursor Ion Selection Algorithms for Increased Depth of Bottom-Up Proteomic Profiling. *J. Proteome Res.* **2016**, *15*, 3563−3573.
(36) Zhang, Y.; Wen, Z.; Washburn, M. P.; Florens, L. Effect of Dynamic Exclusion Duration on Spectral Count Based Quantitative Proteomics. *Anal. Chem.* **2009**, *81*, 6317−6326.
(37) Yu, Q.; Liu, X.; Keller, M. P.; Navarrete-Perea, J.; Zhang, T.; Fu, S.; Vaites, L. P.; Shuken, S. R.; et al. Sample multiplexing-based targeted pathway proteomics with real-time analytics reveals the impact of genetic variation on protein expression. *Nature Comm.* **2023**, *14*, 555.
(38) Horn, D. M.; Zubarev, R. A.; McLafferty, F. W. Automated Reduction and Interpretation of High Resolution Electrospray Mass Spectra of Large Molecules. *J. Am. Soc. Mass Spectrom.* **2000**, *11*, 320−332.

https://doi.org/10.1021/acs.jproteome.2c00838
J. Proteome Res. 2023, 22, 2151−2171

(39) Rad, R.; Li, J.; Mintseris, J.; O'Connell, J.; Gygi, S. P.; Schweppe, D. K. Improved Monoisotopic Mass Estimation for Deeper Proteome Coverage. *J. Proteome Res.* **2021**, *20*, 591−598.
(40) Hunt, D. F.; Yates, J. R., III; Shabanowitz, J.; Winston, S.; Hauer, C. R. Protein sequencing by tandem mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* **1986**, *83*, 6233−6237.
(41) Scheltema, R. A.; Mann, M.; et al. The Q Exactive HF, a Benchtop Mass Spectrometer with a Pre-filter, High-performance Quadrupole and an Ultra-high-field Orbitrap Analyzer. *Mol. Cell. Proteomics* **2014**, *13*, 3698−3708.
(42) Xia, Y.; Liang, X.; McLuckey, S. A. Ion Trap versus Low-Energy Beam-Type Collision-Induced Dissociation of Protonated Ubiquitin Ions. *Anal. Chem.* **2006**, *78*, 1218−1227.
(43) Olsen, J. V.; et al. Higher-energy C-trap dissociation for peptide modification analysis. *Nat. Methods* **2007**, *4*, 709−712.
(44) Paizs, B.; Suhai, S. Fragmentation pathways of protonated peptides. *Mass Spectrom. Rev.* **2005**, *24*, 508−548.
(45) Wysocki, V. H.; Tsaprailis, G.; Smith, L. L.; Breci, L. A. Mobile and localized protons: a framework for understanding peptide dissociation. *J. Mass Spectrom.* **2000**, *35*, 1399−1406.
(46) Roepstorff, P.; Fohlman, J. Proposal for a Common Nomenclature for Sequence Ions in Mass Spectra of Peptides. *Biomed. Mass Spectrom.* **1984**, *11*, 601.
(47) Steen, H.; Mann, M. The abc's (and xyz's) of peptide sequencing. *Nature Rev. Mol. Cell Biol.* **2004**, *5*, 699−711.
(48) Riley, N. M.; Coon, J. J. Phosphoproteomics in the Age of Rapid and Deep Proteome Profiling. *Anal. Chem.* **2016**, *88*, 74−94.
(49) Brodbelt, J. S. Ion Activation Methods for Peptides and Proteins. *Anal. Chem.* **2016**, *88*, 30−51.
(50) Wang, H.; et al. An off-line high pH reversed-phase fractionation and nano-liquid chromatography-mass spectrometry method for global proteomic profiling of cell lines. *J. Chromatography B* **2015**, *974*, 90−95.
(51) Tyanova, S.; Temu, T.; Cox, J. The MaxQuant computational platform for mass spectrometry−based shotgun proteomics. *Nat. Protoc.* **2016**, *11*, 2301−2319.
(52) Gonzalez-Galarza, F. F.; et al. A Critical Appraisal of Techniques, Software Packages, and Standards for Quantitative Proteomic Analysis. *OMICS* **2012**, *16*, 431−442.
(53) Al Shweiki, MHD R.; et al. Assessment of Label-Free Quantification in Discovery Proteomics and Impact of Technological Factors and Natural Variability of Protein Abundance. *J. Proteome Res.* **2017**, *16*, 1410−1424.
(54) Perez-Riverol, Y.; et al. Open source libraries and frameworks for mass spectrometry based proteomics: a developer's perspective. *Biochim. Biophys. Acta* **2014**, *1844*, 63−76.
(55) Eng, J. K.; Searle, B. C.; Clauser, K. R.; Tabb, D. L. A Face in the Crowd: Recognizing Peptides Through Database Search. *Mol. Cell. Proteomics* **2011**, *10*, 1−9.
(56) Tabb, D. L. The SEQUEST Family Tree. *J. Am. Soc. Mass Spectrom.* **2015**, *26*, 1814−1819.
(57) Hoopmann, M. R.; Moritz, R. L. Current algorithmic solutions for peptide-based proteomics data generation and identification. *Curr. Op. Biotechnol.* **2013**, *24*, 31−38.
(58) Pearson, W. R.; Lipman, D. J. Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. U. S. A.* **1988**, *85*, 2444−2448.
(59) "Downloads." UniProt. URL: https://www.uniprot.org/help/downloads (accessed March 19, 2023).
(60) Huala, E.; et al. The Arabidopsis Information Resource (TAIR): a comprehensive database and web-based information retrieval, analysis, and visualization system for a model plant. *Nucleic Acids Res.* **2001**, *29*, 102−105.
(61) Cherry, J. M.; et al. Saccharomyces Genome Database: the genomiocs resource of budding yeast. *Nucleic Acids Res.* **2012**, *40*, D700−705.
(62) Khodadoust, M. S.; Olsson, N.; Elias, J. E.; Alizadeh, A. A.; et al. Antigen presentation profiling reveals recognition of lymphoma immunoglobulin neoantigens. *Nature* **2017**, *543*, 723−727.
(63) Devabhaktuni, A.; Elias, J. E.; et al. TagGraph reveals vast protein modification landscapes from large tandem mass spectrometry datasets. *Nat. Biotechnol.* **2019**, *37*, 469−479.
(64) Quartararo, A. J.; et al. Ultra-large chemical libraries for the discovery of high-affinity peptide binders. *Nature Comm.* **2020**, *11*, 3183.
(65) Tang, W. H.; Shilov, I. V.; Seymour, S. L. Nonlinear Fitting Method for Determining Local False Discovery Rates from Decoy Database Searches. *J. Proteome Res.* **2008**, *7*, 3661−3667.
(66) Elias, J. E.; Gygi, S. P. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat. Methods* **2007**, *4*, 207−214.
(67) Cox, J.; Mann, M.; et al. Andromeda: A Peptide Search Engine Integrated into the MaxQuant Environment. *J. Proteome Res.* **2011**, *10*, 1794−1805.
(68) Eng, J. K.; McCormack, A. L.; Yates, J. R., III An Approach to Correlate Tandem Mass Spectral Data of Peptides with Amino Acid Sequences in a Protein Database. *J. Am. Soc. Mass Spectrom.* **1994**, *5*, 976−989.
(69) Perkins, D. N.; Pappin, D. J. C.; Creasy, D. M.; Cottrell, J. S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **1999**, *20*, 3551−3567.
(70) Wenger, C. D.; Coon, J. J. A Proteomics Search Algorithm Specifically Designed for High-Resolution Tandem Mass Spectra. *J. Proteome Res.* **2013**, *12*, 1377−1386.
(71) Eng, J. K.; Fischer, B.; Grossmann, J.; MacCoss, M. J. A Fast SEQUEST Cross Correlation Algorithm. *J. Proteome Res.* **2008**, *7*, 4598−4602.
(72) Eng, J. K.; Jahan, T. A.; Hoopmann, M. R. Comet: An open-source MS/MS sequence database search tool. *Proteomics* **2013**, *13*, 22−24.
(73) McIlwain, S.; Hoopmann, M. R.; Käll, L.; Eng, J. K.; MacCoss, M. J.; Noble, W. S.; et al. Crux: Rapid Open Source Protein Tandem Mass Spectrometry Analysis. *J. Proteome Res.* **2014**, *13*, 4488−4491.
(74) Käll, L.; Canterbury, J. D.; Weston, J.; Noble, W. S.; MacCoss, M. J. Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nat. Methods* **2007**, *4*, 923−925.
(75) Huttlin, E. L.; et al. A Tissue-Specific Atlas of Mouse Protein Phosphorylation and Expression. *Cell* **2010**, *143*, 1174−1189.
(76) Halloran, J. T.; et al. Speeding Up Percolator. *J. Proteome Res.* **2019**, *18*, 3353−3359.
(77) Cox, J.; Hein, M. Y.; Luber, C. A.; Paron, I.; Nagaraj, N.; Mann, M. Accurate Proteome-wide Label-free Quantification by Delayed Normalization and Maximal Peptide Ratio Extraction, Termed MaxLFQ. *Mol. Cell. Proteomics* **2014**, *13*, 2513−2526.
(78) Ong, S.-E.; et al. Stable Isotope Labeling by Amino Acids in Cell Culture, SILAC, as a Simple and Accurate Approach to Expression Proteomics. *Mol. Cell. Proteomics* **2002**, *1*, 376−386.
(79) Krüger, M.; et al. SILAC Mouse for Quantitative Proteomics Uncovers Kindlin-3 as an Essential Factor for Red Blood Cell Function. *Cell* **2008**, *134*, 353−364.
(80) Ong, S.-E. The expanding field of SILAC. *Anal. Bioanal. Chem.* **2012**, *404*, 967−976.
(81) Rauniyar, N.; Yates, J. R., III Isobaric Labeling-Based Relative Quantification in Shotgun Proteomics. *J. Proteome Res.* **2014**, *13*, 5293−5309.
(82) Sivanich, M. K.; Gu, T.-J.; Tabang, D. N.; Li, L. Recent advances in isobaric labeling and applications in quantitative proteomics. *Proteomics* **2022**, *22*, 2100256.
(83) Thompson, A.; et al. Tandem Mass Tags: A Novel Quantification Strategy for Comparative Analysis of Complex Protein Mixtures by MS/MS. *Anal. Chem.* **2003**, *75*, 1895−1904.
(84) Ting, L.; Rad, R.; Gygi, S. P.; Haas, W. MS3 eliminates ratio distortion in isobaric multiplexed quantitative proteomics. *Nat. Methods* **2011**, *8*, 937−940.
(85) McAlister, G. C.; et al. MultiNotch MS3 Enables Accurate, Sensitive, and Multiplexed Detection of Differential Expression across Cancer Cell Line Proteomes. *Anal. Chem.* **2014**, *86*, 7150−7158.

https://doi.org/10.1021/acs.jproteome.2c00838
J. Proteome Res. 2023, 22, 2151−2171

(86) Schweppe, D. K.; et al. Full-Featured, Real-Time Database Searching Platform Enables Fast and Accurate Multiplexed Quantitative Proteomics. *J. Proteome Res.* **2020**, *19*, 2026−2034.
(87) Li, J.; Gygi, S. P.; Paulo, J. A.; et al. TMTpro-18plex: The Expanded and Complete Set of TMTpro Reagents for Sample Multiplexing. *J. Proteome Res.* **2021**, *20*, 2964−2972.
(88) Kettenbach, A. N.; Rush, J.; Gerber, S. A. Absolute quantification of protein and post-translational modification abundance with stable isotope−labeled synthetic peptides. *Nat. Protoc.* **2011**, *6*, 175−186.
(89) Pino, L. K.; et al. Matrix-Matched Calibration Curves for Assessing Analytical Figures of Merit in Quantitative Proteomics. *J. Proteome Res.* **2020**, *19*, 1147−1153.
(90) Nesvizhskii, A. I.; Aebersold, R. Interpretation of Shotgun Proteomic Data: The protein inference problem. *Mol. Cell. Proteomics* **2005**, *4*, 1419−1440.
(91) Zhang, B.; Chambers, M. C.; Tabb, D. L. Proteomic Parsimony through Bipartite Graph Analysis Improves Accuracy and Transparency. *J. Proteome Res.* **2007**, *6*, 3549−3557.
(92) Shuken, S. R.; et al. Limited proteolysis-mass spectrometry reveals aging-associated changes in cerebrospinal fluid protein abundances and structures. *Nature Aging* **2022**, *2*, 379−388.
(93) Serang, O.; MacCoss, M. J.; Noble, W. S. Efficient marginalization to compute protein posterior probabilities from shotgun mass spectrometry data. *J. Proteome Res.* **2010**, *9*, 5346−5357.
(94) Serang, O.; Moruz, L.; Hoopmann, M. R.; Käll, L. Recognizing Uncertainty Increases Robustness and Reproducibility of Mass Spectrometry-based Protein Inferences. *J. Proteome Res.* **2012**, *11*, 5586−5591.
(95) Pfeuffer, J.; Sachsenberg, T.; Dijkstra, T. M. H.; Serang, O.; Reinert, K.; Kohlbacher, O. EPIFANY: A Method for Efficient High-Confidence Protein Inference. *J. Proteome Res.* **2020**, *19*, 1060−1072.
(96) Savitski, M. M.; Wilhelm, M.; Hahne, H.; Kuster, B.; Bantscheff, M. A Scalable Approach for Protein False Discovery Rate Estimation in Large Proteomic Data Sets. *Mol. Cell. Proteomics* **2015**, *14*, 2394−2404.
(97) The, M.; Tasnim, A.; Käll, L. How to talk about protein-level false discovery rates in shotgun proteomics. *Proteomics* **2016**, *16*, 2461−2469.
(98) Cox, J.; Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **2008**, *26*, 1367−1372.
(99) Nesvizhskii, A. I.; Keller, A.; Kolker, E.; Aebersold, R. A Statistical Model for Identifying Proteins by Tandem Mass Spectrometry. *Anal. Chem.* **2003**, *75*, 4646−4658.
(100) Reiter, L.; Claassen, M.; Schrimpf, S. P.; Jovanovic, M.; Schmidt, A.; Buhmann, J. M.; Hengartner, M. O.; Aebersold, R. Protein Identification False Discovery Rates for Very Large Proteomics Data Sets Generated by Tandem Mass Spectrometry. *Mol. Cell. Proteomics* **2009**, *8*, 2405−2417.
(101) Bludau, I.; et al. Systematic detection of functional proteoform groups from bottom-up proteomic datasets. *Nat. Commun.* **2021**, *12* (3810), 1−18.
(102) Smith, L. M.; Kelleher, N. L. The Consortium for Top Down Proteomics. "Proteoform: a single term describing protein complexity. *Nat. Methods* **2013**, *10*, 186−187.
(103) Tyanova, S.; Cox, J. Perseus: A Bioinformatics Platform for Integrative Analysis of Proteomics Data in Cancer Research. In *Cancer systems Biology: Methods and Protocols*; von Stechow, L., Ed.; Springer Nature: New York, 2018; pp 133−148.
(104) Choi, M. MSstats: an R package for statistical analysis of quantitative mass spectrometry-based proteomic experiments. *Bioinformatics* **2014**, *30*, 2524−2526.
(105) Shuken, S. R.; McNerney, M. W. Costs and Benefits of Popular P-Value Correction Methods in Three Models of Quantitative Omic Experiments. *Anal. Chem.* **2023**, *95*, 2732−2740.
(106) "The Gene Ontology Resource." Gene Ontology: Unifying Biology. URL: http://geneontology.org/ (accessed March 6, 2021).
(107) Snel, B.; Lehmann, G.; Bork, P.; Huynen, M. A. STRING: a web-server to retrieve and display the repeatedly occurring neighbourhood of a gene. *Nucleic Acids Res.* **2000**, *28*, 3442−3444.
(108) "STRING." STRING. URL: https://string-db.org/ (accessed March 6, 2021).
(109) Langfelder, P.; Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **2008**, *9*, 559.
(110) Louris, J. N.; Cooks, R. G.; et al. Instrumentation, Applications, and Energy Deposition in Quadrupole Ion-Trap Tandem Mass Spectrometry. *Anal. Chem.* **1987**, *59*, 1677−1685.
(111) Macias, L. A.; Santos, I. C.; Brodbelt, J. S. Ion Activation Methods for Peptides and Proteins. *Anal. Chem.* **2020**, *92*, 227−251.
(112) Kanu, A. B.; et al. Ion mobility-mass spectrometry. *J. Mass Spectrom.* **2008**, *43*, 1−22.
(113) Hebert, A. S.; et al. Comprehensive Single-Shot Proteomics with FAIMS on a Hybrid Orbitrap Mass Spectrometer. *Anal. Chem.* **2018**, *90*, 9529−9537.
(114) Savaryn, J. P.; Toby, T. K.; Kelleher, N. L. A researcher's guide to mass spectrometry-based proteomics. *Proteomics* **2016**, *16*, 2435−2443.
(115) Paulo, J. A. Practical and Efficient Searching in Proteomics: A Cross Engine Comparison. *Webmedcentral* **2013**, *4*, No. WMCPLS0052.
(116) Medzihradszky, K. F.; Chalkley, R. J. Lessons in *de novo* peptide sequencing by tandem mass spectrometry. *Mass Spectrom. Rev.* **2015**, *34*, 43−63.
(117) Ma, B.; et al. PEAKS: powerful software for peptide *de novo* sequencing by tandem mass spectrometry. *Rapid Commun. Mass Spectrom.* **2003**, *17*, 2337−2342.
(118) Peng, W.; Pronker, M. F.; Snijder, J. Mass Spectrometry-Based De Novo Sequencing of Monoclonal Antibodies Using Multiple Proteases and a Dual Fragmentation Scheme. *J. Proteome Res.* **2021**, *20*, 3559−3566.
(119) Kim, S.; Pevzner, P. A. MS-GF+ makes progress towards a universal database search tool for proteomics. *Nature Comm.* **2014**, *5* (5277), 1−10.
(120) Kong, A. T.; et al. MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry−based proteomics. *Nat. Methods* **2017**, *14*, 513−520.
(121) Polasky, D. A.; Yu, F.; Teo, G. C.; Nesvizhskii, A. I. Fast and comprehensive N- and O-glycoproteomics analysis with MSFragger-Glyco. *Nat. Methods* **2020**, *17*, 1125−1132.
(122) Wen, B.; et al. Deep Learning in Proteomics. *Proteomics* **2020**, *20*, 1900335.
(123) Mann, M.; Kumar, C.; Zeng, W.-F.; Strauss, M. T. Artificial intelligence for proteomics and biomarker discovery. *Cell Systems* **2021**, *12*, 759−770.
(124) Donnelly, D. P.; et al. Best practices and benchmarks for intact protein analysis for top-down mass spectrometry. *Nat. Methods* **2019**, *16*, 587−594.
(125) Jooß, K.; McGee, J. P.; Kelleher, N. L. Native Mass Spectrometry at the Convergence of Structural Biology and Compositional Proteomics. *Acc. Chem. Res.* **2022**, *55*, 1928−1937.
(126) Smith, R. D.; et al. Trapping, detection and reaction of very large single molecular ions by mass spectrometry. *Nature* **1994**, *369*, 137−139.
(127) Kafader, J. O.; et al. Measurement of Individual Ions Sharply Increases the Resolution of Orbitrap Mass Spectra of Proteins. *Anal. Chem.* **2019**, *91*, 2776−2783.
(128) Wörner, T. P.; et al. Resolving heterogeneous macromolecular assemblies by Orbitrap-based single-particle charge detection mass spectrometry. *Nat. Methods* **2020**, *17*, 395−398.
(129) McGee, J. P.; Kelleher, N. L.; et al. Isotopic Resolution of Protein Complexes up to 466 kDa Using Individual Ion Mass Spectrometry. *Anal. Chem.* **2021**, *93*, 2723−2727.
(130) Su, P.; et al. Highly multiplexed, label-free proteoform imaging of tissues by individual ion mass spectrometry. *Sci. Adv.* **2022**, *8* (32), eabp9929.
(131) Mann, M.; Jensen, O. N. Proteomic analysis of post-translational modifications. *Nat. Biotechnol.* **2003**, *21*, 255−261.

https://doi.org/10.1021/acs.jproteome.2c00838
J. Proteome Res. 2023, 22, 2151−2171

(132) Walsh, C. T.; Garneau-Tsodikova, S.; Gatto, G. J. Protein Posttranslational Modifications: The Chemistry of Proteome Diversifications. *Angew. Chem., Int. Ed.* **2005**, *44*, 7342−7372.
(133) Chan, J. C.; Maze, I. Nothing Is Yes Set in (Hi)stone: Novel Post-Translational Modifications Regulating Chromatin Function. *Trends Biochem. Sci.* **2020**, *45*, 829−844.
(134) Gassaway, B. M.; et al. A multi-purpose, regenerable, proteome-scale, human phosphoserine resource for phosphoproteomics. *Nat. Methods* **2022**, *19*, 1371−1375.
(135) Udeshi, N. D.; Mertins, P.; Svinkina, T.; Carr, S. A. Large-scale identification of ubiquitination sites by mass spectrometry. *Nat. Protoc.* **2013**, *8*, 1950−1960.
(136) Udeshi, N. D.; et al. Rapid and deep-scale ubiquitylation profiling for biology and translational research. *Nature Comm.* **2020**, *11*, 359.
(137) *Essentials of Glycobiology*; Varki, A., Ed.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor (NY), 2022; ISBN-13: 978-1-621824-22-0.
(138) Bagdonaite, I.; et al. Glycoproteomics. *Nature Rev. Methods Primers* **2022**, *2*, 48.
(139) Totten, S. M.; Feasley, C. L.; Bermudez, A.; Pitteri, S. J. Parallel Comparison of N-Linked Glycopeptide Enrichment Techniques Reveals Extensive Glycoproteomic Analysis of Plasma Enabled by SAX-ERLIC. *J. Proteome Res.* **2017**, *16*, 1249−1260.
(140) Xiao, H.; Chen, W.; Smeekens, J. M.; Wu, R. An enrichment method based on synergistic and reversible covalent interactions for large-scale analysis of glycoproteins. *Nat. Commun.* **2018**, *9* (1692), 1−12.
(141) Burt, R. A.; et al. Novel Antibodies for the Simple and Efficient Enrichment of Native O-GlcNAc Modified Peptides. *Mol. Cell. Proteomics* **2021**, *20*, 100167.
(142) Morgenstern, D.; et al. Optimized Glycopeptide Enrichment Method−It Is All About the Sauce. *Anal. Chem.* **2022**, *94*, 10308−10313.
(143) Malaker, S. A.; et al. Revealing the human mucinome. *Nat. Commun.* **2022**, *13*, 3542.
(144) Riley, N. M.; Coon, J. J. The Role of Electron Transfer Dissociation in Modern Proteomics. *Anal. Chem.* **2018**, *90*, 40−64.
(145) Riley, N. M.; Malaker, S. A.; Driessen, M. D.; Bertozzi, C. R. Optimal Dissociation Methods Differ for *N*- and *O*-Glycopeptides. *J. Proteome Res.* **2020**, *19*, 3286−3301.
(146) Malaker, S. A.; et al. The mucin-selective protease StcE enables molecular and functional analysis of human cancer-associated mucins. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116*, 7278−7287.
(147) Shon, D. J.; et al. An enzymatic toolkit for selective proteolysis, detection, and visualization of mucin-domain glycoproteins. *Proc. Natl. Acad. Sci. U. S. A.* **2020**, *117*, 21299−21307.
(148) Shon, D. J.; Kuo, A.; Ferracane, M. J.; Malaker, S. A. Classification, structural biology, and applications of mucin domain-targeting proteases. *Biochem. J.* **2021**, *478*, 1585−1603.
(149) Bern, M.; Kil, Y. J.; Becker, C. Byonic: advanced peptide and protein identification software. *Curr. Protoc. Bioinformatics* **2012**, *40*, 13.20.1−13.20.14.
(150) Riley, N. M.; Bertozzi, C. R. Deciphering *O*-glycoprotease substrate preferences with O-Pair Search. *Mol. Omics* **2022**, *18*, 908.
(151) Thomas, D. R.; Scott, N. E. Glycoproteomics: growing up fast. *Curr. Op. Struc. Biol.* **2021**, *68*, 18−25.
(152) Leitner, A. A review of the role of chemical modification methods in contemporary mass spectrometry-based proteomics research. *Anal. Chim. Acta* **2018**, *1000*, 2−19.
(153) Kaur, U.; Jones, L. M.; et al. Evolution of Structural Biology through the Lens of Mass Spectrometry. *Anal. Chem.* **2019**, *91*, 142−155.
(154) Liu, X. R.; Zhang, M. M.; Gross, M. L. Mass Spectrometry-Based Protein Footprinting for Higher-Order Structure Analysis: Fundamentals and Applications. *Chem. Rev.* **2020**, *120*, 4355−4454.
(155) Bamberger, C.; et al. Protein Footprinting via Covalent Protein Painting Reveals Structural Changes of the Proteome in Alzheimer's Disease. *J. Proteome Res.* **2021**, *20*, 2762−2771.
(156) O'Reilly, F. J.; Rappsilber, J. Cross-linking mass spectrometry: methods and applications in structural, molecular and systems biology. *Nature Struc. Mol. Biol.* **2018**, *25*, 1000−1008.
(157) Feng, Y.; et al. Global analysis of protein structural changes in complex proteomes. *Nat. Biotechnol.* **2014**, *32*, 1036−1044.
(158) Mackmull, M.-T.; et al. Global, in situ analysis of the structural proteome in individuals with Parkinson's disease to identify a new class of biomarker. *Nature Struc. Mol. Biol.* **2022**, *29*, 978−989.
(159) Piazza, I.; et al. A Map of Protein-Metabolite Interactions Reveals Principles of Chemical Communication. *Cell* **2018**, *172*, 358−372.e23.
(160) Malinovska, L.; et al. Proteome-wide structural changes measured with limited proteolysis-mass spectrometry: an advanced protocol for high-throughput applications. *Nat. Protoc.* **2023**, *18*, 659−682.
(161) Moellering, R. E.; Cravatt, B. F. How Chemoproteomics Can Enable Drug Discovery and Development. *Chem. & Biol.* **2012**, *19*, 11−22.
(162) Drewes, G.; Knapp, S. Chemoproteomics and Chemical Probes for Target Discovery. *Trends Biotechnol.* **2018**, *36*, 1275−1286.
(163) Luo, M.; et al. Chemoproteomics-enabled discovery of covalent RNF114-based degraders that mimic natural product function. *Cell Chem. Biol.* **2021**, *28*, 559−566.
(164) Kuljanin, M.; et al. Reimagining high-throughput profiling of reactive cysteines for cell-based screening of large electrophile libraries. *Nat. Biotechnol.* **2021**, *39*, 630−641.
(165) Kemper, E. K.; Zhang, Y.; Dix, M. M.; Cravatt, B. F. Global profiling of phosphorylation-dependent changes in cysteine reactivity. *Nat. Methods* **2022**, *19*, 341−352.
(166) West, A. V.; Woo, C. M. Photoaffinity Labeling Chemistries Used to Map Biomolecular Interactions. *Isr. J. Chem.* **2023**, *63*, No. e202200081.
(167) Gaetani, M.; et al. Proteome Integral Solubility Alteration: A High-Throughput Proteomics Assay for Target Deconvolution. *J. Proteome Res.* **2019**, *18*, 4027−4037.
(168) Li, J.; Van Vranken, J. G.; et al. Selection of Heating Temperatures Improves the Sensitivity of the Proteome Integral Solubility Alteration Assay. *J. Proteome Res.* **2020**, *19*, 2159−2166.
(169) Brewer, K. D.; Shi, S. M.; Wyss-Coray, T. Unraveling protein dynamics to understand the brain − the next molecular frontier. *Mol. Neurodeg.* **2022**, *17* (45), 1−10.
(170) Qin, W.; Cho, K. F.; Cavanagh, P. E.; Ting, A. Y. Deciphering molecular interactions by proximity labeling. *Nat. Methods* **2021**, *18*, 133−143.
(171) Boersema, P. J.; et al. Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics. *Nat. Protoc.* **2009**, *4*, 484−494.
(172) Budnik, B.; Levy, E.; Harmange, G.; Slavov, N. SCoPE-MS: mass spectrometry of single mammalian cells quantifies proteome heterogeneity during cell differentiation. *Genome Biol.* **2018**, *19* (1), 161.
(173) Zhu, Y.; et al. Proteomic Analysis of Single Mammalian Cells Enabled by Microfluidic Nanodroplet Sample Preparation and Ultrasensitive NanoLC-MS. *Angew. Chem.* **2018**, *130*, 12550−12554.
(174) Dou, M.; et al. High-Throughput Single Cell Proteomics Enabled by Multiplex Isobaric Labeling in a Nanodroplet Sample Preparation Platform. *Anal. Chem.* **2019**, *91*, 13119−13127.
(175) Kelly, R. T. Single-cell Proteomics: Progress and Prospects. *Mol. Cell. Proteomics* **2020**, *19*, 1739−1748.
(176) Brunner, A.-D.; et al. Ultra-high sensitivity mass spectrometry quantifies single-cell proteome changes upon perturbation. *Mol. Sys. Biol.* **2022**, *18* (3), e10798.
(177) Vistain, L. F.; Tay, S. Single-Cell Proteomics. *Trends Biochem. Sci.* **2021**, *46*, 661−672.
(178) Gillet, L. C.; et al. Targeted Data Extraction of the MS/MS Spectra Generated by Data-independent Acquisition: A New Concept for Consistent and Accurate Proteome Analysis. *Mol. Cell. Proteomics* **2012**, *11* (6), O111.016717.

https://doi.org/10.1021/acs.jproteome.2c00838
J. Proteome Res. 2023, 22, 2151−2171

(179) Muntel, J.; et al. Comparison of Protein Quantification in a Complex Background by DIA and TMT Workflows with Fixed Instrument Time. *J. Proteome Res.* **2019**, *18*, 1340−1351.
(180) Ludwig, C.; Gillet, L.; Rosenberger, G.; Amon, S.; Collins, B. C.; Aebersold, R. Data-independent acquisition-based SWATH-MS for quantitative proteomics: a tutorial. *Mol. Sys. Biol.* **2018**, *14*, No. e8126.
(181) Zhang, F.; et al. Data-Independent Acquisition Mass Spectrometry-Based Proteomics and Software Tools: A Glimpse in 2020. *Proteomics* **2020**, *20*, 1900276.
(182) Meier, F.; et al. BoxCar acquisition method enables single-shot proteomics at a depth of 10,000 proteins in 100 minutes. *Nat. Methods* **2018**, *15*, 440−448.
(183) Panchaud, A.; et al. Precursor Acquisition Independent From Ion Count: How to Dive Deeper into the Proteomics Ocean. *Anal. Chem.* **2009**, *81*, 6481−6488.
(184) Demichev, V.; Messner, C. B.; Vernardis, S. I.; Lilley, K. S.; Ralser, M. DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. *Nat. Methods* **2020**, *17*, 41−44.
(185) Demichev, V.; et al. dia-PASEF data analysis using FragPipe and DIA-NN for deep proteomics of low sample amounts. *Nat. Commun.* **2022**, *13*, 3944.
(186) Derks, J.; et al. Increasing the throughput of sensitive proteomics by plexDIA. *Nat. Biotechnol.* **2023**, *41*, 50−59.
(187) Marx, V. Targeted proteomics. *Nat. Methods* **2013**, *10*, 19−22.
(188) Picotti, P.; Aebersold, R. Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. *Nat. Methods* **2012**, *9*, 555−566.
(189) Zhang, H.; et al. Methods for Peptide and Protein Quantitation by Liquid Chromatography-Multiple Reaction Monitoring Mass Spectrometry. *Mol. Cell. Proteomics* **2011**, *10*, M110.006593.
(190) Peterson, A. C.; et al. Parallel Reaction Monitoring for High Resolution and High Mass Accuracy Quantitative, Targeted Proteomics. *Mol. Cell. Proteomics* **2012**, *11*, 1475−1488.
(191) Erickson, B. K.; et al. A Strategy to Combine Sample Multiplexing with Targeted Proteomics Assays for High-Throughput Protein Signature Characterization. *Mol. Cell* **2017**, *65*, 361−370.
(192) Remes, P. M.; Yip, P.; MacCoss, M. J. Highly Multiplex Targeted Proteomics Enabled by Real-Time Chromatographic Alignment. *Anal. Chem.* **2020**, *92*, 11809−11817.
(193) Zhu, H.; et al. PRM-LIVE with Trapped Ion Mobility Spectrometry and Its Application in Selectivity Profiling of Kinase Inhibitors. *Anal. Chem.* **2021**, *93*, 13791−13799.
(194) Yu, Q.; et al. Sample multiplexing for targeted pathway proteomics in aging mice. *Proc. Natl. Acad. Sci. U. S. A.* **2020**, *117*, 9723−9732.
(195) Reiter, L.; et al. mProphet: automated data processing and statistical validation for large-scale SRM experiments. *Nat. Methods* **2011**, *8*, 430−435.
(196) Suhre, K.; McCarthy, M. I.; Schwenk, J. M. Genetics meets proteomics: perspectives for large population-based studies. *Nat. Rev. Genet.* **2021**, *22*, 19−37.
(197) Gold, L.; et al. Aptamer-Based Multiplexed Proteomic Technology for Biomarker Discovery. *PLoS One* **2010**, *5* (12), No. e15004.
(198) Sun, B. B.; et al. Genomic atlas of the human plasma proteome. *Nature* **2018**, *558*, 73−79.
(199) Petrera, A.; et al. Multiplatform Approach for Plasma Proteomics: Complementarity of Olink Proximity Extension Assay Technology to Mass Spectrometry-Based Protein Profiling. *J. Proteome Res.* **2021**, *20*, 751−762.
(200) Swaminathan, J.; et al. Highly parallel single-molecule identification of proteins in zeptomole-scale mixtures. *Nat. Biotechnol.* **2018**, *36*, 1076−1082.
(201) Brinkerhoff, H.; Kang, A. S. W.; Liu, J.; Aksimentiev, A.; Dekker, C. Multiple rereads of single proteins at single-amino acid resolution using nanopores. *Science* **2021**, *374*, 1509−1513.
(202) Alfaro, J. A.; et al. The emerging landscape of single-molecule protein sequencing technologies. *Nat. Methods* **2021**, *18*, 604−617.
(203) Cardozo, N.; et al. Multiplexed direct detection of barcoded protein reporters on a nanopore array. *Nat. Biotechnol.* **2022**, *40*, 42−46.
(204) Reed, B. D.; et al. Real-time dynamic single-molecule protein sequencing on an integrated semiconductor device. *Science* **2022**, *378*, 186−192.
(205) Marx, V. Proteomics sets up single-cell and single-molecule solutions. *Nat. Methods* **2023**, *20*, 350−354.
(206) MacCoss, M. J.; Alfaro, J. A.; Faivre, D. A.; Wu, C. C.; Wanunu, M.; Slavov, N. Sampling the proteome by emerging single-molecule and mass spectrometry methods. *Nat. Methods* **2023**, *20*, 339−346.

https://doi.org/10.1021/acs.jproteome.2c00838
J. Proteome Res. 2023, 22, 2151−2171