A Brief Introduction to the Very Basics of Mass Spectrometry-Based Proteomics

Steven R. Shuken[1,2,3,4]

[1]*Department of Chemistry, Stanford University, Stanford, CA, USA.*
[2]*Department of Neurology and Neurological Sciences, Stanford University School of Medicine, Stanford, CA, USA.*
[3]*Wu Tsai Neurosciences Institute, Stanford University, Stanford, CA, USA.*
[4]*Department of Cell Biology, Harvard Medical School, Boston, MA, USA.*

Updated 31 August, 2022 and published to https://www.stevenshuken.com/intro-to-proteomics.

Introduction

A mass spectrometry (MS)-based proteomic experiment typically provides answers to two questions:

1. ***What*** are the proteins in my samples?
2. What is the ***abundance*** of each protein in each sample?

Sometimes there are additional questions about, e.g., protein structure or interactions with binding partners. The technologies used to answer these additional questions are extensions of the same technology used to answer the first two.

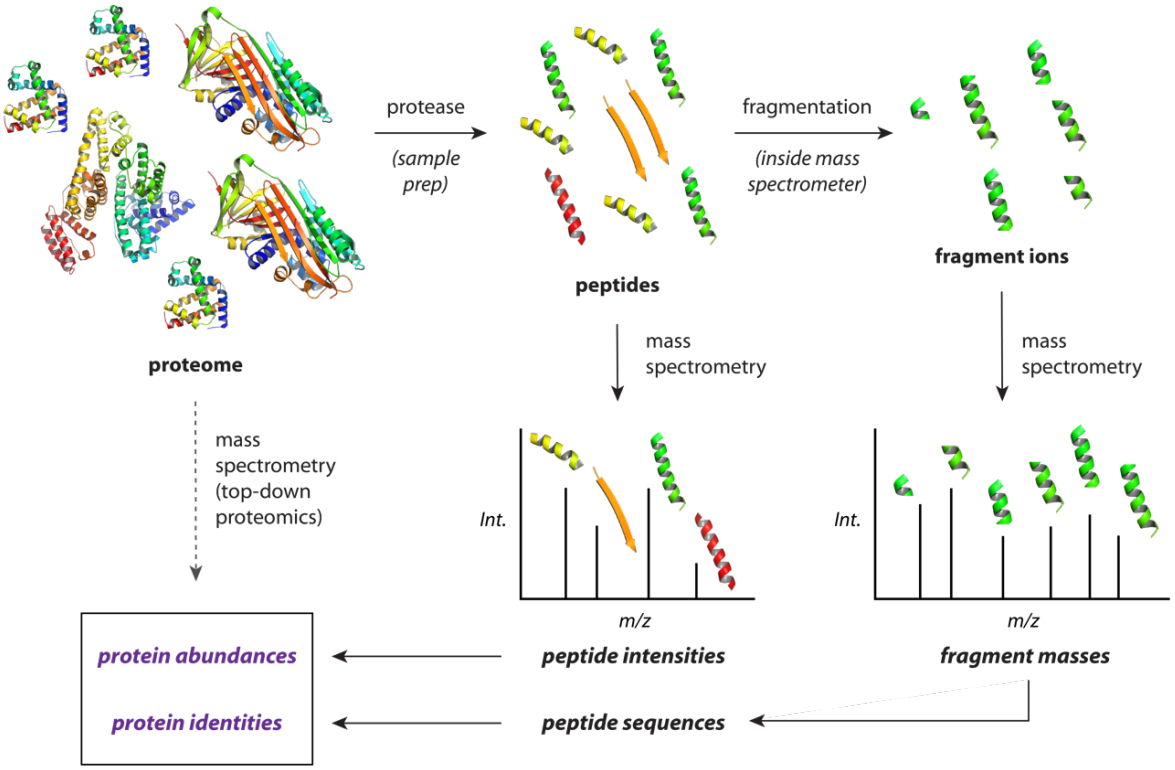


Figure 1. Conceptual diagram of the typical shotgun mass spectrometry-based proteomic experiment. Proteins are digested by a protease into peptides, which are injected into a mass spectrometer. The mass spectrometer measures the peptides' mass-to-charge

ratios (m/z) and electrical signals that the peptides produce as they interact with detectors. The intensity (Int.) of each electrical signal is plotted on the y-axis of each spectrum. Each peptide is then fragmented inside the mass spectrometer, and the fragments are measured. The resulting data are analyzed to infer the amino acid sequences and abundances of the peptides, which are used to infer the identities and abundances of the proteins.

This tutorial will focus on shotgun proteomics. The word "shotgun" is used to describe non-targeted proteomics, in which the only bias in the identification of proteins comes from the physics of protein detection. This is in contrast to targeted proteomics, in which a specific protein or set of proteins is detected, and proteomics methods that assay a predetermined set of proteins, such as antibody or aptamer arrays. Although a shotgun proteomic experiment will often include additional steps to reveal information such as the states of post-translational modifications (PTMs),[1,2,3] ligand/drug binding,[4,5,6] or three-dimensional structure,[7,8,9] we present here a description of the simplest experiment, designed only for protein identities and quantities. This tutorial is focused on the experimental aspects of a simple experiment from a biologist's perspective, heavily relying on visual aids, rather than providing a deep and broad description of mass spectrometry and its uses in proteomics, which can be found elsewhere.[10]

The most common approach to answering these 2 questions is called bottom-up proteomics, in which pieces of proteins called peptides (typically ~7–20 amino acids in length) are analyzed, and then inferences are made about the proteins from which the peptides originated (Figure 1). The alternatives—including middle-down, top-down, and native proteomics—are promising and powerful approaches that are still under development; recent advances are discussed elsewhere.[11,12,13]

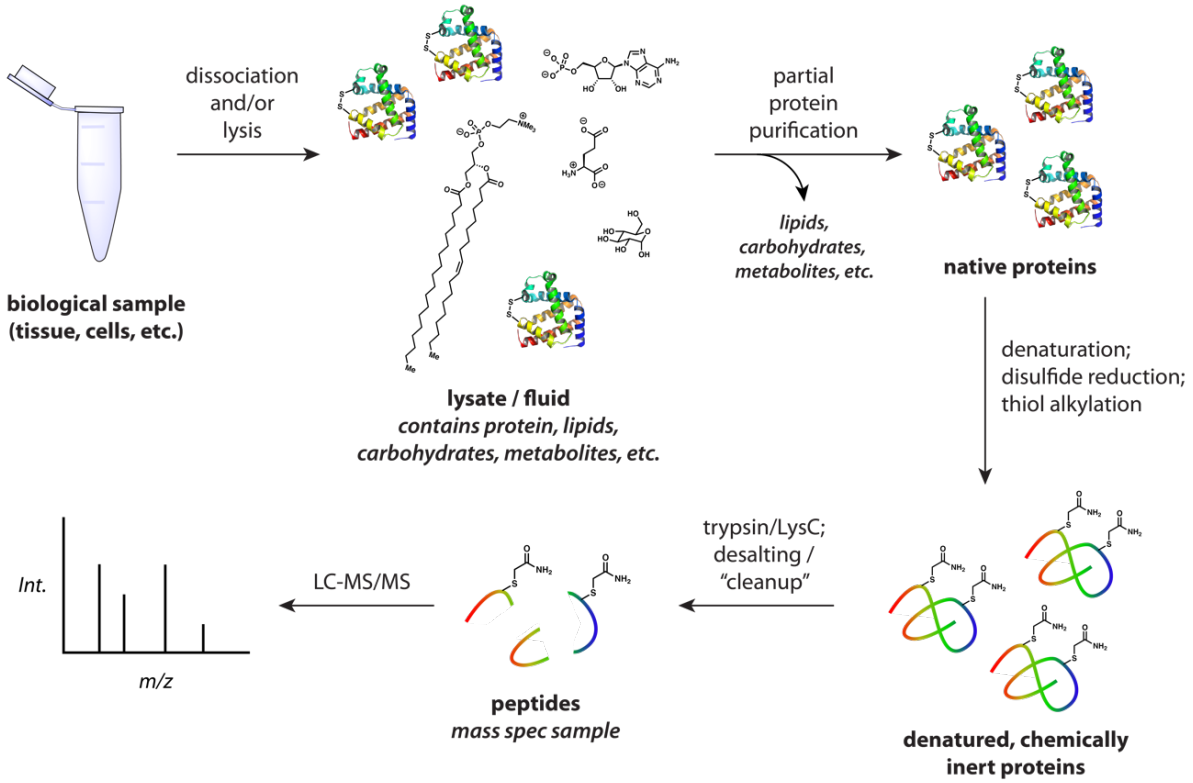


Figure 2. Generic sample preparation workflow for mass spectrometry-based proteomics. These steps may be rearranged, removed, or added to the workflow depending on the experiment. Reagents used for each step also vary.

Experimental Workflow

Sample preparation

Every proteomic experiment begins with sample preparation. A typical sample preparation is diagrammed in Figure 2. First, if tissues or plant samples are being analyzed, they are homogenized/dissociated, and if cells are intact then they are lysed. The proteins are partially purified, e.g., by precipitation in cold acetone followed by washing and then resuspension in aqueous buffer. After this step (or before this step, during lysis), the partially proteins are denatured (unfolded), typically with urea or a detergent such as sodium deoxycholate (DOC). The proteins are then rendered chemically inert by reduction and alkylation, usually with dithiothreitol (DTT) or tris-(2-carboxyethyl)phosphine (TCEP) followed by iodoacetamide (IAA). The reduced and alkylated proteins are then digested, most commonly with a combination of trypsin and LysC at 37 °C overnight, and the resulting

peptides are purified once more, e.g., by centrifugation or vacuum suction through a C18-coated solid phase using alternating mobile phases for trapping and elution. The eluent is then evaporated and the peptides are finally resuspended in a solvent compatible with liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis, such as 0.1% formic acid in high-pressure liquid chromatography- (HPLC-)grade water.

Various approaches to the details of these workflows have been published. Some effective methods with which the author is familiar are cited here,[14,7,15] and studies have been done comparing various methods.[16] The reader is encouraged to follow sample preparation protocols used in studies that best match their own experimental plan.

Liquid chromatography-tandem mass spectrometry (LC-MS/MS)

Once the autosampler takes up a specified volume of the peptide suspension and pushes it onto the HPLC column, three things begin to happen in concert (though a fourth step, MS3, is sometimes included as discussed below): liquid chromatography, mass spectrometry of peptides (MS1), and mass spectrometry of fragments (MS2). In LC-MS/MS, "LC," "MS," and "MS" signify these three steps, respectively.

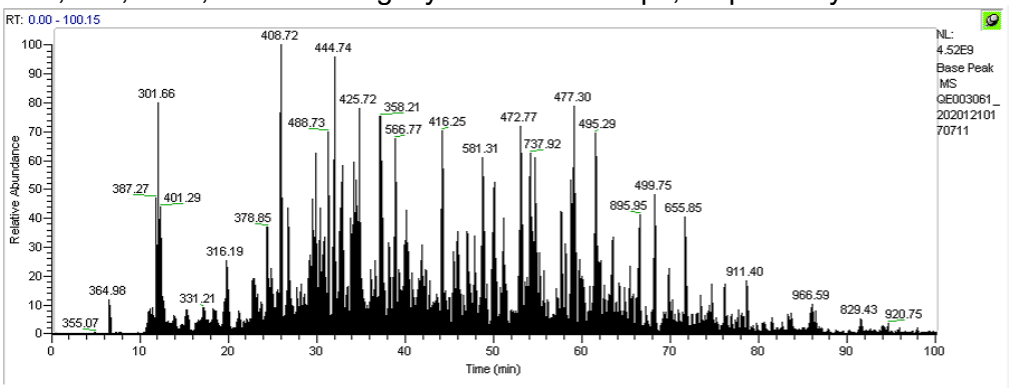


Figure 3. A typical chromatogram from an LC-MS/MS run. The sample is a commercially available mixture of peptides from a digested HeLa cell proteome (Thermo Fisher Scientific). The sample was run on a Q Exactive HF-X spectrometer (Thermo Fisher Scientific) with an UltiMate 3000 RSLCnano LC (Thermo Fisher Scientific). Along the y-axis is plotted the intensity of the tallest MS1 peak of the scan at that point in time (the "base peak"). Each peak is labeled with the m/z value of the base peak from the scan occurring at the peak. RT = retention time. NL = normalization level (the intensity of the tallest peak in the chromatogram).

1. High-pressure liquid chromatography (HPLC or "LC")

   The sample is eluted through an HPLC column using a pre-programmed solvent gradient, i.e., a mixture of solvents that changes in composition throughout the typically 1–3-h run. HPLC is almost always done in reversed-phase mode, meaning that the column has a hydrophobic stationary phase (typically C18-coated silica) and the mobile phase is a mixture of aqueous and polar organic solvents. The gradient can be altered to optimize peak shape and separation and allow greater time to sequence more peptides, or run for a shorter time to allow more samples to be run in the experiment or to increase signal by sharpening peaks. An example of a chromatogram resulting from an LC gradient is shown in Figure 3.

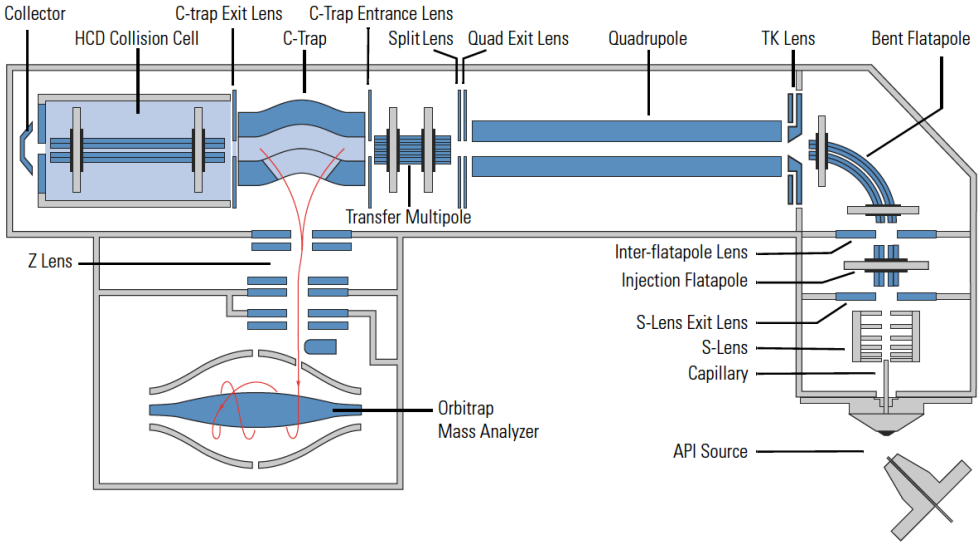


Figure 4. Schematic of the Q Exactive. The atmospheric pressure ionization (API) source can be replaced by the electrospray ionization (ESI) source, which applies a voltage across the gap between the HPLC column tip and the Capillary. The Capillary is heated to a high

temperature to facilitate vaporization of spray droplets, allowing peptides to enter the gas phase. All lenses are ion optical lenses; these, and the flatapoles and the multipole, are made of electromagnets. Generally, their purpose is to direct and focus the ion beam (peptides/fragments). The quadrupole filters out all but one m/z value during MS2 scans (or rather, a small range of m/z values centered around the desired peptide m/z). The higher-energy collision-induced dissociation (HCD) Collision Cell is where high-energy N_2 molecules collide with peptides, generating fragments. Peptides or fragments are collected in the C-Trap for some time before an applied voltage injects them into the Orbitrap for mass analysis and detection (m/z and intensity measurements, i.e., a scan). Image from Ref. 20.

2. Peptide MS scans (MS1)

As the sample is being continuously sprayed from the tip of the HPLC column into the spectrometer, the spectrometer repeatedly and quickly acquires spectra in order to detect peptides. Several effective mass analyzers and detectors exist;[17,18] one example of a popular mass analyzer/detector is called the Orbitrap (Thermo Fisher Scientific).[19] A relatively simple example of a mass spectrometer with an Orbitrap is the Q Exactive (Thermo Fisher Scientific), the interior of which is depicted in Figure 4.[20] In the Q Exactive, high- and low-m/z debris is removed as peptides follow a curved path, the "Bent Flatapole." During MS1, the Quadrupole is inactive. The peptides are trapped in the C-Trap, which is an electromagnet that applies a field that functions as a "well" for the peptides to electrostatically "fall into" with the help of nitrogen (N_2) molecules that absorb kinetic energy.[21]

After an adjustable ion accumulation time ("Injection Time" in the instrument software) of up to ~200 ms,[21] the C-Trap feeds its ions simultaneously and instantaneously into the Orbitrap, which produces a signal from which the peptide m/z values can be deconvoluted.[19] The ions orbit the Orbitrap for an adjustable period (adjusted by altering the "Resolution" parameter in the instrument software) of typically tens of milliseconds, up to ~200 ms or longer.[22,23] At this point, an MS1 spectrum such as that shown in Figure 5 is generated. As the peptides are sprayed into the mass spectrometer, MS1 spectra are acquired repeatedly in this manner, many times per minute—or multiple times per second, if few or no peptides are observed.

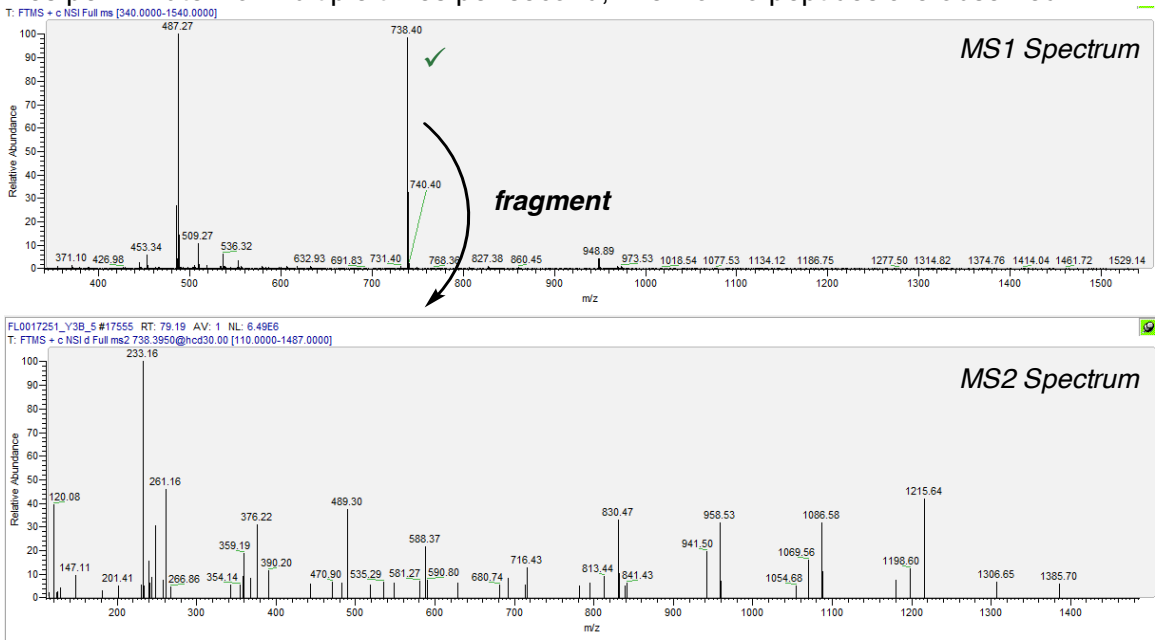


Figure 5. Example of MS1 and MS2 scans. Here, a peptide with m/z = 738.40 Th was fragmented (Th is the m/z unit). FTMS = fourier transform mass spectrum. c = centroid. NSI = nanospray ionization. AV = number of averaged scans. hcd30.00 = HCD at normalized collision energy (NCE) 30.

3. Fragment MS scans (MS2)

After each MS1 scan, the spectrometer, while running, selects the most intense (usually 10–20) peptides in the MS1 spectrum for fragmentation, one at a time. Fragmentation enables sequencing of the peptides (Figure 1). During an MS2 scan on a Q Exactive, all of the elements of the spectrometer are active at some point (Figure 4). The quadrupole acts as a filter, ensuring that only ions with the selected m/z (within some small adjustable tolerance such as ±0.5 Th[24]) pass into the transfer multipole. The selected ions pass through the C-Trap and enter the higher-energy collision-induced dissociation (HCD) chamber, where high-energy N_2 molecules cause fragmentation by collision. The fragments are then passed back into the C-Trap to be injected into the Orbitrap for analysis, producing an MS2 spectrum such as that shown in Figure 5.

In summary: throughout the ~1–3-h gradient, the machine continuously accumulates peptides; the peptides are analyzed (MS1) and selected for fragmentation; each selected peptide is re-accumulated and fragmented,

and then the fragments are analyzed (MS2). Then the cycle repeats. If an MS1 scan took a relatively long 250 ms and induced 10 MS2 scans each taking 250 ms, then the cycle would be complete in about 3 seconds. 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 MS2 scans are dependent upon the MS1 data. An increasingly popular acquisition mode is data-independent acquisition (DIA), in which all MS1 m/z values are included in fragmentation every cycle,[25] and other innovations in raw data acquisition continue to emerge.[26]

Data Analysis: Peptide Identification

The raw data produced by a single LC-MS/MS run is a massive collection of spectra, each with a retention time, other metadata (e.g., identity as an MS1 or MS2 spectrum), m/z values, and intensities. Software packages such as MaxQuant or Proteome Discoverer (Thermo Fisher Scientific) process these data to generate peptide and/or protein lists with statistical metrics of the identifications. The various software options for peptide searching are compared elsewhere,[27,28,29] as are the algorithms for peptide identification that those software packages implement.[30,31] Peptide-spectrum matching will be discussed conceptually here, but the details of the algorithms are beyond the scope of this work.

In most shotgun MS data analyses, the user provides a known reference proteome as a text file, usually in FASTA format[32] and most commonly retrieved from Uniprot.org.[33] 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 search algorithms, which do not require a reference proteome, lack this limitation, though they have their own disadvantages.[34]) The data processing software predicts all peptides that could arise from these protein by enzymatic cleavage (most commonly with trypsin and LysC) and compares these predicted peptides with the spectra to make matches. For example, the amino acid sequence of mouse albumin is shown in Figure 6 with four tryptic cleavage sites highlighted with orange asterisks. Underlined green regions represent peptides that were sequenced in a shotgun LC-MS/MS analysis of cerebrospinal fluid. Note that all sequenced regions end with an arginine (R) or lysine (K) residue and are preceded by an R or K because the trypsin/LysC combination cleaves at the C-terminal side of R and K.

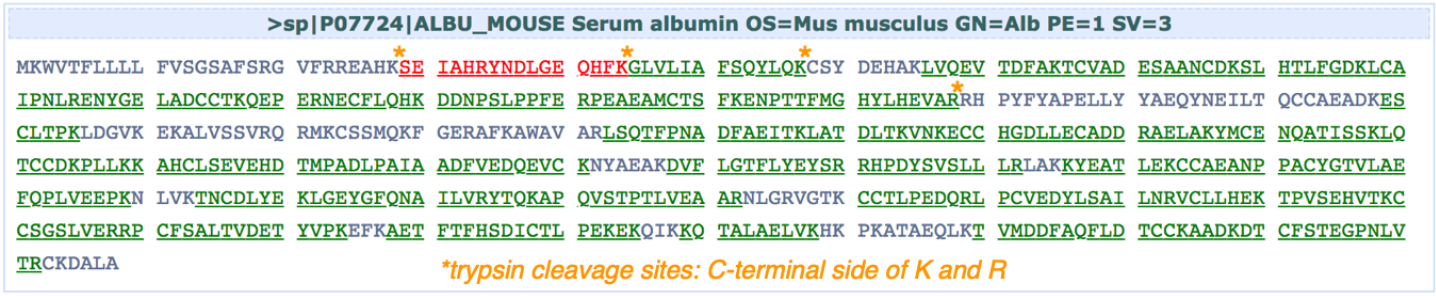
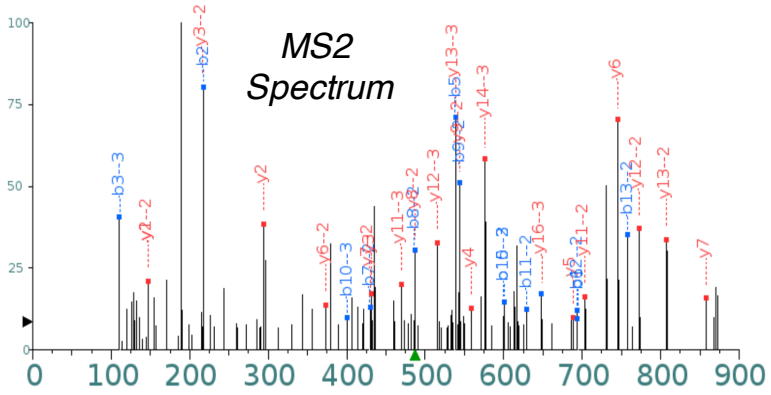


Figure 6. Amino acid sequence of mouse albumin with one tryptic peptide (red) and four tryptic cleavage sites (orange asterisks) indicated. Underlined green sequences represent regions that were sequenced by peptide LC-MS/MS.

The MS2 spectrum for the peptide highlighted in red in Figure 6 is shown in Figure 7. Peaks that have been identified as matches with predicted fragment ions are labeled with a blue b or a red y. This corresponds to the type of fragment generated: b-ions contain the N terminus of the peptide while y-ions contain the C terminus.[35] With HCD, these are the predominant ions formed, and only b- and y-ions are searched for in the analysis.

This spectrum is a good example of how a single spectrum can provide strong evidence for the sequence of a peptide in the sample. The quality of the peptide-spectrum match (PSM) is incorporated into a statistical score; for the peptide in Figure 7, the experimenter can be highly confident that this peptide was eluted from the column at this time point because of the high statistical score (>99.9%). All peptide searching algorithms apply scores to such peptide-spectrum matches (PSMs) and include manipulable cutoffs for peptide inclusion in the data output.[36] An estimation of false discovery rates (FDRs) in the PSM set[37] then produces a final list of peptides at a low FDR (typically 1%).



Seq	#	+1 b	+1 y	+2 b	+2 y	+3 b	+3 y
S	1	88.039	---	44.523	---	30.018	---
E	2	**217.082**	1856.904	109.045	928.956	73.032	619.639
I	3	330.166	1727.861	165.587	864.434	**110.727**	**576.625**
A	4	401.203	1614.777	201.105	**807.892**	134.406	**538.931**
H	5	**538.262**	1543.740	269.635	**772.374**	180.092	**515.252**
R	6	**694.363**	1406.681	347.685	**703.844**	232.126	**469.565**
Y	7	857.426	1250.580	**429.217**	625.794	286.480	417.532
N	8	971.469	1087.517	**486.238**	**544.262**	324.495	363.177
D	9	1086.496	973.474	**543.752**	**487.241**	362.837	325.163
L	10	1199.580	**858.447**	**600.294**	**429.727**	**400.532**	286.820
G	11	1256.602	**745.363**	**628.805**	**373.185**	419.539	249.126
E	12	1385.644	**688.341**	**693.326**	344.674	462.553	230.119
Q	13	1513.703	**559.299**	**757.355**	280.153	505.239	187.104
H	14	1650.762	**431.240**	825.885	**216.124**	550.925	144.418
F	15	1797.830	**294.181**	899.419	**147.594**	**599.948**	98.732
K	16	---	**147.113**	---	74.060	---	49.709

```
Peptide:
SEIAHRYNDLGEQHFK
Detected (b ions):
SE
SEI
SEIAH
SEIAHR
SEIAHRY
SEIAHRYN
SEIAHRYND
SEIAHRYNDL
SEIAHRYNDLG
SEIAHRYNDLGE
SEIAHRYNDLGEQ
SEIAHRYNDLGEQHF
```

Figure 7. MS2 spectrum, sequence, fragment m/z values, and b-ion sequences for the peptide highlighted in Figure 6. *Top*: MS2 spectrum. *Bottom left*: The theoretical m/z ratios of all b- and y-ions for this peptide are listed for each charge state (z value) from +1 to +3. A b-ion includes the N terminus of the peptide while a y-ion includes the C terminus (Ref. 35). Bold and colored m/z values represent peaks observed in the MS2 spectrum. *Bottom right*: The sequences of all b-ion fragments observed in this spectrum.

The "Protein Level:" Inferring Protein Identities and Abundances

Peptide quantification

As described above, the LC continuously feeds peptides into the mass spectrometer throughout the run. At least once every 3 seconds or so, the spectrometer produces an MS1 spectrum with ion peaks and intensities (Figure 5). Although each peptide is only fragmented once or a few times, the MS1 peak is observed every cycle during its elution from the HPLC column. Together, over a period of seconds or minutes, these MS1 data points form a chromatographic peak (Figure 3). A cartoon of twelve such peaks is illustrated in Figure 8. A feature of the chromatographic peak, such as its maximum height or the area under the curve (AUC), is used to measure the relative abundance of the peptide. In targeted proteomics, absolute abundance can be measured by the usage of an internal standard.[38] However, in shotgun proteomics, only relative comparisons between sample groups can be made accurately, because the relationship between peptide abundance and MS intensity differs between peptides. The relative abundances of peptides are most commonly calculated in one of three ways:[39]

- **Label-free quantification (LFQ):** chromatographic peaks are compared between runs (Figure 8, red vs. blue). No reagents required.
- **Metabolic labeling** (e.g. stable isotope labeling with amino acids in cell culture [SILAC]):[40] the biological system is utilized to label proteins with amino acids containing heavy isotopes. The protein samples are combined before sample prep and the comparison is performed in MS1 spectra (Figure 8, red vs. red).
- **Isobaric labeling** (e.g., tandem mass tags [TMT]):[41] proteins or peptides are labeled during sample prep and then combined before LC-MS/MS. Comparison is performed in MS2 or MS3 spectra (Figure 8, bottom-right).

There are cases in which it is appropriate for the analysis to stop here, at the "peptide level," for example, if covalent binding sites or post-translational modification (PTM) sites are under investigation.[1,5] In such cases,

hypotheses about protein chemistry can be made case-by-case based on the data and then tested in a targeted manner.[7]

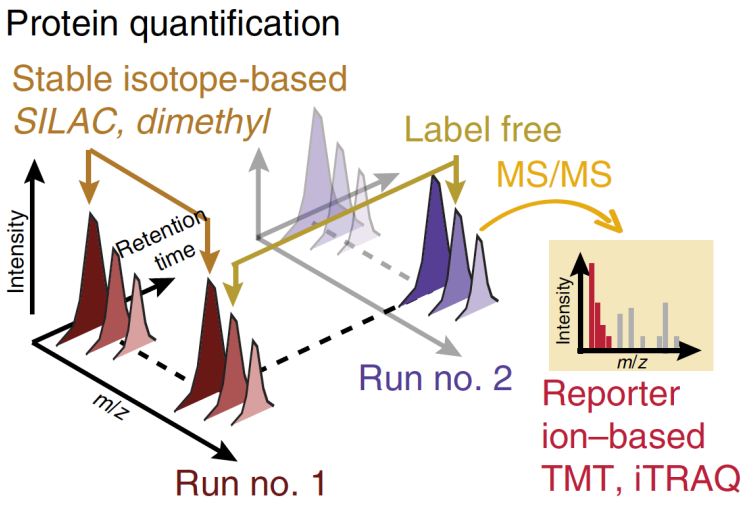


Figure 8. Diagram of peptide quantification methods. All 12 peaks represent the same peptide. Different chromatographic peak colors represent different LC-MS/MS runs; shading and small differences in m/z represent naturally occurring isotopes; large differences in m/z represent stable isotope labels, e.g., stable isotope labeling with amino acids in cell culture (SILAC). In isobaric labeling methods such as tandem mass tags (TMT), relative peak heights in MS2 or MS3 spectra are used to measure relative quantities instead of a chromatographic peak. From Ref. 39.

Protein inference and grouping

Although some biology papers using bottom-up proteomics refer to their results as "proteins," they usually refer to protein groups, which are virtual approximations of proteins.[42] Sequenced peptides can be ambiguous with regard to which protein and/or gene they originated from (Figure 9). In the case of proteins for which there is equal evidence that they existed in the sample (e.g., proteins V and VI in Figure 9), these are grouped together in a protein group (PG) so that the number of PGs is a good approximation of the number of proteins sequenced in the experiment. 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 9), are excluded by some algorithms but retained in others.[43,39] A more systematic, detailed discussion is had in Ref.s 39 and 43, but popular data processing software packages tend to deal with these issues appropriately, especially in a typical differential protein abundance or sequencing experiment.[44]

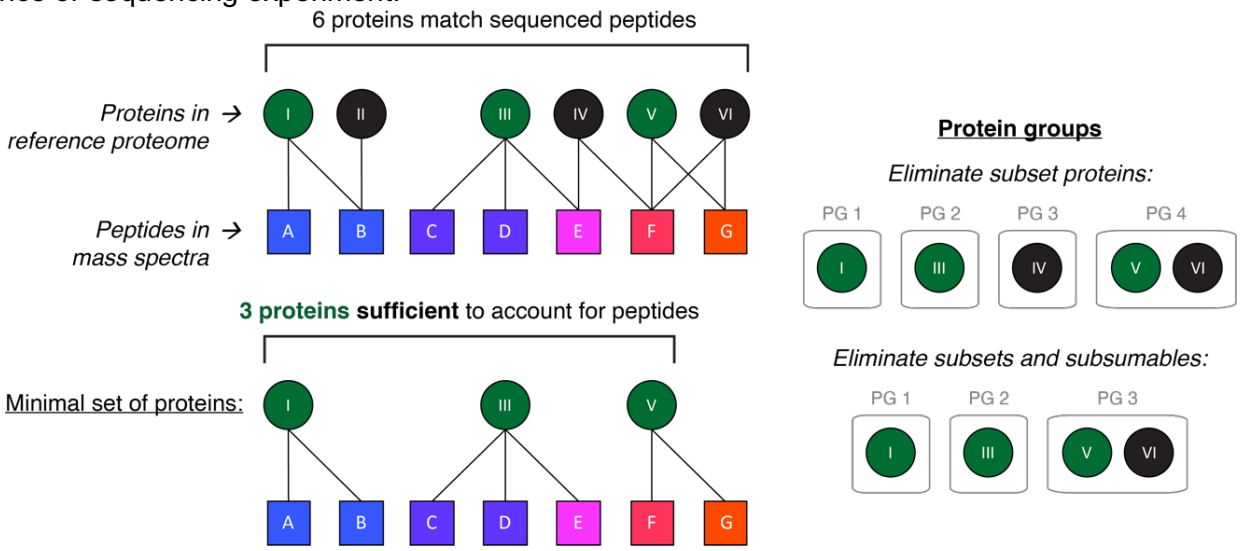


Figure 9. Diagram of protein inference and grouping strategies in a hypothetical example. In this example, 7 peptides were sequenced that match the sequences of 6 proteins. Matches are illustrated by black lines. 3 proteins (green) are sufficient to explain the existence of all peptides in the sample, so at least 3 proteins were sequenced. Since protein V and VI match the same peptides, they are equally likely to have existed in the sample(s). So, proteomic data processing software will report them as a protein group (PG) that contains both of them. Some algorithms will eliminate protein IV, which is called a subsumable protein (Ref. 43). Most algorithms will eliminate protein II, whose peptides are a subset of protein I's peptides.

Protein group quantification

PG quantification is a more complicated task than protein identification, because PGs are quantified using peptide intensities, and peptides in a PG may differ from each other in the proteins they match.[44] Some peptides may be unique to a PG while some may not; some may match a specific protein and no others (e.g., peptides A, C, and D in Figure 9), allowing quantification of a singular protein (assuming the FASTA file completely describes the proteome). To complicate matters further, a peptide may arise from an atypical protein in the PG or from a proteoform[45] with post-translational modifications that affect the peptide's relative intensity.

The most popular data processing software packages present several options to the user to address these issues: some allow quantification to be performed only with PG-unique or protein-unique ("proteotypic") peptides, and some allow the user to decide whether to use the median, sum, arithmetic mean, or geometric mean of peptide intensities to calculate the PG intensity. With labeling approaches (Figure 8), ratios of intensities of peptides with different labels are used to calculate relative PG intensities. For label-free quantification, the popular MaxLFQ algorithm implemented in the MaxQuant software package is of note. MaxLFQ uses a sophisticated treatment of label-free peptide intensity ratios to estimate relative PG intensities, using median ratios to avoid the influence of outlier peptides.[46]

The output of a data processing software package that performs PG quantification is a list of PGs with relative abundances, or absolute abundances if standards are used.[38] As mentioned above, the quantities of different PGs within a sample are not comparable, because of varying relationships between abundances and intensities. However, the measurements of each PG between samples serve as estimates of relative protein abundances.

Summary: The Workflow from Beginning to End

In total, the workflow for a label-free shotgun mass spectrometry experiment is as follows. Detailed experimental protocols as well as other more specialized workflows can be found in the References.

1. **Dissociate tissues/plants, lyse cells, denature proteins, reduce cystines** with DTT, **alkylate cysteines** with IAA.
2. **Partially purify proteins** with, e.g., cold acetone-mediated protein precipitation.
3. **Digest proteins** with a protease, e.g., 1:1 trypsin:LysC.
4. **Purify peptides** with, e.g., a C18-coated filter pipette tip using centrifugation. **Evaporate solvent** with, e.g., a vacuum centrifuge.
5. **Resuspend peptides** in an LCMS-compatible solvent, e.g., 0.1% formic acid in water, in an LCMS-compatible vial.
6. **Submit runs** to an LC-MS/MS system using the appropriate instrument method.
7. **Process raw data** using a suitable software package, such as the popular MaxQuant or Proteome Discoverer packages, or a combination of softwares. Perform peptide searching, FDR estimation, quantification, and protein inference.
8. **Interpret the results.** Perseus[47] and MSStats[48] are popular packages designed for further processing and statistical analysis of proteomic data. Many more exist After statistical analysis is complete, the biologist can appropriately interpret the results. Databases such as Gene Ontology[49] or the search tool for recurring instances of neighboring genes (STRING)[50,51] as well as correlational approaches such as weighted gene correlation network analysis (WGCNA)[52] allow pathway- and network-level analysis of results.

Outlook

The future of mass spectrometry in protein sequencing and quantification is made uncertain by non-mass-spectrometric technologies, such as antibody-[53] or aptamer-based[54] protein detection, which have already been fully commercialized, and emerging "next-generation protein sequencing" technologies based on, e.g., amino acid-wise side chain identification or nanopore-type sequencing.[55,56] While these technologies may bring a welcome decrease in the price and difficulty of proteomic studies in the future, mass spectrometry will always have unique advantages in the study of proteins. The flexibility of mass spectrometry allows relatively easy switching between studies of different species, whereas other methods may require development of a new set of probes to study each additional species. Mass spectrometers easily detect covalent tags that reveal protein chemistry or structure, whereas the same probe may not fit through a pore or may preclude fluorescence-based side chain identification. Such advantages make mass spectrometry the ideal tool to study spatiotemporal dynamics of proteins, post-translational modifications, ligand-host binding, protein-protein interactions, protein aggregation, or other aspects of protein chemistry that may or may not be known. The usage of mass spectrometry in combination with other non-mass-spectrometric technologies is likely to be a powerful paradigm for unprecedented understanding of proteomes across biology.

Acknowledgments

The author thanks members of the Elias Laboratory (Chan-Zuckerberg Biohub), Wyss-Coray Laboratory (Stanford University), and Gygi Laboratory (Harvard Medical School) for feedback and support. The Stanford Center for Molecular Analysis and Design (CMAD) is thanked for funding and support. Figure 6 and Figure 7 were generated using the Elias Laboratory's in-house mass spectrometry analysis software.

References

[1] Engholm-Keller, K.; Larsen, M. R. "Technologies and challenges in large-scale phosphoproteomics." *Proteomics* **2013**, *13*, 910–931.
[2] Bekker-Jensen, D. B.; Olsen, J. V.; et al. "Rapid and site-specific deep phosphoproteome profiling by data-independent acquisition without the need for spectral libraries." *Nature Comm.* **2020**, *11*, 787.
[3] Thomas, D. R.; Scott, N. E. "Glycoproteomics: growing up fast." *Curr. Op. Struc. Biol.* **2021**, *68*, 18–25.
[4] Chen, G.; et al. "Advances in MS Based Strategies for Probing Ligand-Target Interactions: Focus on Soft Ionization Mass Spectrometric Techniques." *Front. Chem.* **2019**, *7*, 703.
[5] Drewes, G.; Knapp, S. "Chemoproteomics and Chemical Probes for Target Discovery." *Trends Biotechnol.* **2018**, *36*, 1275–1286.
[6] Piazza, I.; Picotti, P.; et al. "A Map of Protein-Metabolite Interactions Reveals Principles of Chemical Communication." *Cell* **2018**, *172*, 358–372.
[7] Schopper, S.; Picotti, P.; et al. "Measuring protein structural changes on a proteome-wide scale using limited proteolysis-coupled mass spectrometry." *Nature Protocols* **2017**, *12*, 2391–2410.
[8] 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.
[9] Kaur, U.; Jones, L. M.; et al. "Evolution of Structural Biology through the Lens of Mass Spectrometry." *Anal. Chem.* **2019**, *91*, 142–155.
[10] Sinha, A.; Mann, M. "A beginner's guide to mass spectrometry-based proteomics." *Biochem. (Lond.)* **2020**, *42*, 64–69.
[11] Cristobal, A.; Heck, A. J. R.; et al. "Toward an Optimized Workflow for Middle-Down Proteomics." *Anal. Chem.* **2017**, *89*, 3318–3325.
[12] Cupp-Sutton, K. A.; Wu, S. "High-throughput quantitative top-down proteomics." *Mol. Omics* **2020**, *16*, 91–99.
[13] 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.
[14] Hughes, C. S.; Krijgsveld, J.; et al. "Ultrasensitive proteome analysis using paramagnetic bead technology." *Mol. Sys. Biol.* **2014**, *10*, 757.
[15] HaileMariam, M.; et al. "S-Trap, an Ultrafast Sample-Preparation Approach for Shotgun Proteomics." *J. Proteome Res.* **2018**, *17*, 2917–2924.
[16] Wang, W.-Q.; et al. "Evaluation of sample preparation methods for mass spectrometry-based proteomic analysis of barley leaves." *Plant Methods* **2018**, *14*, 72.
[17] Eliuk, S.; Makarov, A. "Evolution of Orbitrap Mass Spectrometry Instrumentation." *Annu. Rev. Anal. Chem.* **2015**, *8*, 61–80.
[18] 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.
[19] Makarov, A. "Electrostatic Axially Harmonic Orbital Trapping: A High-Performance Technique of Mass Analysis." *Anal. Chem.* **2000**, *72*, 1156–1162.
[20] Thermo Fisher Scientific. "Exactive Series Operating Manual." Revision A, April 2017. Document ID: BRE0012255.
[21] "Orbitrap." *Mass Spec Pro.* Accessed March 4, 2021. URL: http://www.massspecpro.com/mass-analyzers/orbitrap.
[22] "Q Exactive Plus (Q)." *UWPR.* Accessed March 4, 2021. URL: https://proteomicsresource.washington.edu/instruments/qexactiveplus.php.
[23] 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.
[24] 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.
[25] 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*, e8126.
[26] Meier, F.; Geyer, P. E.; Winter, S. V.; Cox, J.; Mann, M. "BoxCar acquisition method enables single-shot proteomics at a depth of 10,000 proteins in 100 minutes." *Nature Methods* **2018**, *15*, 440–448.
[27] Gonzalez-Galarza, F. F.; et al. "A Critical Appraisal of Techniques, Software Packages, and Standards for Quantitative Proteomic Analysis." *OMICS* **2012**, *16*, 431–442.
[28] Al Shweiki, MHD R. A.; 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.

[29] 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.
[30] 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.
[31] Tabb, D. L. “The SEQUEST Family Tree.” *J. Am. Soc. Mass Spectrom.* **2015**, *26*, 1814–1819.
[32] Pearson, W. R.; Lipman, D. J. “Improved tools for biological sequence comparison.” *Proc. Natl. Acad. Sci.* **1988**, *85*, 2444–2448.
[33] “Downloads release 2021_01.” *UniProt.* Accessed March 6, 2021. URL: https://www.uniprot.org/downloads.
[34] Blank-Landeshammer, B.; Sickmann, A.; et al. “Combining De Novo Peptide Sequencing Algorithms, A Synergistic Approach to Boost Both Identifications and Confidence in Bottom-up Proteomics.” *J. Proteome Res.* **2017**, *16*, 3209–3218.
[35] Roepstorff, P. “Proposal for a Common Nomenclature for Sequence Ions in Mass Spectra of Peptides.” *Biomed. Mass Spectrom.* **1984**, *11*, 601.
[36] Cox, J.; Mann, M.; et al. “Andromeda: A Peptide Search Engine Integrated into the MaxQuant Environment.” *J. Proteome Res.* **2011**, *10*, 1794–1805.
[37] Elias, J. E.; Gygi, S. P. “Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry.” *Nature Methods* **2007**, *4*, 207–214.
[38] Schwanhäusser, B.; Selbach, M.; et al. “Global quantification of mammalian gene expression control.” *Nature* **2011**, *473*, 337–342.
[39] Tyanova, S.; Temu, T.; Cox, J. “The MaxQuant computational platform for mass spectrometry–based shotgun proteomics.” *Nature Protocols* **2016**, *11*, 2301–2319.
[40] Krüger, M.; Moser, M.; Mann, M.; et al. “SILAC Mouse for Quantitative Proteomics Uncovers Kindlin-3 as an Essential Factor for Red Blood Cell Function.” **2008**, *134*, 353–364.
[41] Thermo Fisher Scientific. “User Guide: TMT10plex Mass Tag Labeling Kits and Reagents.” URL: https://www.thermofisher.com/order/catalog/product/90110.
[42] Nesvizhskii, A. I.; Aebersold, R. “Interpretation of Shotgun Proteomic Data: The protein inference problem.” *Mol. Cell. Proteomics* **2005**, *4*, 1419–1440.
[43] 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.
[44] 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.
[45] Smith, L. M.; Kelleher, N. L.; The Consortium for Top Down Proteomics. “Proteoform: a single term describing protein complexity.” *Nature Methods* **2013**, *10*, 186–187.
[46] 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.
[47] Tyanova, S.; Cox, J. “Perseus: A Bioinformatics Platform for Integrative Analysis of Proteomics Data in Cancer Research.” *Cancer systems Biology: Methods and Protocols*, Methods in Molecular Biology, vol. 1711. Louise von Stechow (ed.). **2018**.
[48] Choi, M. “MSstats: an R package for statistical analysis of quantitative mass spectrometry-based proteomic experiments.” *Bioinformatics* **2014**, *30*, 2524–2526.
[49] “The Gene Ontology Resource.” *Gene Ontology: Unifying Biology.* Accessed March 6, 2021. URL: http://geneontology.org/.
[50] 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.
[51] “STRING.” *STRING.* Accessed March 6, 2021. URL: https://string-db.org/.
[52] Langfelder, P.; Horvath, S. “WGCNA: an R package for weighted correlation network analysis.” *BMC Bioinformatics* **2008**, *9*, 559.
[53] Suhre, K.; McCarthy, M. I.; Schwenk, J. M. “Genetics meets proteomics: perspectives for large population-based studies.” *Nature Reviews Genetics* **2021**, *22*, 19–37.
[54] Gold, L.; et al. “Aptamer-Based Multiplexed Proteomic Technology for Biomarker Discovery.” *PLoS One* **2010**, *5*, 12, e15004.
[55] Alfaro, J. A.; et al. “The emerging landscape of single-molecule protein sequencing technologies.” *Nature Methods* **2021**, *18*, 604–617.
[56] Cardozo, N.; et al. “Multiplexed direct detection of barcoded protein reporters on a nanopore array.” *Nature Biotechnol.* **2022**, *40*, 42–46.