proteomes

MDPI

Review

A Critical Review of Bottom-Up Proteomics: The Good, the Bad, and the Future of This Field

Emmalyn J. Dupree [1,†], **Madhuri Jayathirtha** [1,†], **Hannah Yorkey** [1], **Marius Mihasan** [2,*], **Brindusa Alina Petre** [3,4,*] and **Costel C. Darie** [1,*]

[1] Biochemistry & Proteomics Group, Department of Chemistry and Biomolecular Science, Clarkson University, 8 Clarkson Avenue, Potsdam, NY 13699-5810, USA; dupreeej@clarkson.edu (E.J.D.); jayathm@clarkson.edu (M.J.); yorkeyh@clarkson.edu (H.Y.)
[2] Laboratory of Biochemistry & Molecular Biology, Faculty of Biology, "Alexandru Ioan Cuza" University of Iasi, 700506 Iasi, Romania
[3] Laboratory of Biochemistry, Department of Chemistry, "Alexandru Ioan Cuza" University of Iasi, 700506 Iasi, Romania
[4] Center for Fundamental Research and Experimental Development in Translation Medicine–TRANSCEND, Regional Institute of Oncology, 700506 Iasi, Romania
* Correspondence: marius.mihasan@uaic.ro (M.M.); brindusa.petre@uaic.ro (B.A.P.); cdarie@clarkson.edu (C.C.D.)
† These authors contributed equally to this work.

Received: 1 June 2020; Accepted: 1 July 2020; Published: 6 July 2020

Abstract: Proteomics is the field of study that includes the analysis of proteins, from either a basic science prospective or a clinical one. Proteins can be investigated for their abundance, variety of proteoforms due to post-translational modifications (PTMs), and their stable or transient protein–protein interactions. This can be especially beneficial in the clinical setting when studying proteins involved in different diseases and conditions. Here, we aim to describe a bottom-up proteomics workflow from sample preparation to data analysis, including all of its benefits and pitfalls. We also describe potential improvements in this type of proteomics workflow for the future.

Keywords: mass spectrometry; proteomics; protein identification; protein characterization

1. Proteomics

The term proteome refers to all proteins that are produced or modified by an organism (e.g., human [1], animal [2], plant [3], bacteria [4]) or living system (e.g., organ, cell culture, complex community from an environmental sample)). The term "proteome" and the first dedicated proteomics laboratory were introduced in 1994 by Wilkins et al. to describe proteins as a complement to genomic data [5]. However, the "whole" proteome of a particular cell, tissue, organ, or organism is yet to be identified. This is particularly difficult due to the vast variety of proteins and their isoproteins/proteoforms/protein species, which are expressed at different levels—from very abundant proteins, such as actin, to less abundant ones, such as transcription factors—in different cells, tissues, or organs. The variety of post-translational modifications (PTMs) in proteins, which may be stable or transient, is responsible for the vast number of proteoforms, which is an obstacle in most proteomics experiments. This, corroborated with the multiple limitations of proteomics methods, makes the quest to identify the proteome of any given organism a difficult one [6].

The proteomics field consists of a wide range of methodology, which has been largely driven by the modern development of involved technology. The concept of global protein analysis as a complete atlas of human proteins was proposed over 50 years ago [7]; however, proteomics research did not start until the mid-1990s. The beginning of proteomics research was sparked due to parallel developments in four areas: (i) two-dimensional gel electrophoresis (2D-PAGE) evolving into a robust method to rapidly separate proteins contained in biological complex mixtures [8]; (ii) the continuous development of mass spectrometry methods for accurate mass and chemical structure analysis [9,10]; (iii) the constant production of large-scale genome research and enormous number of peptide/protein sequences catalogued in several databases [11]; and (iv) the development of novel bioinformatics tools to simplify the analysis of large volumes of MS data aiming towards identifying intact proteins and their functional or pathological PTMs [12,13].

The field of proteomics comprises a bioanalytical discipline that performs large-scale studies on proteins [14] that may be rooted from a basic science perspective or a clinical one, i.e., proteins that are associated with a broad range of diseases and conditions by means of their altered expression levels and/or PTMs. In addition to fundamental protein research or development of proteomics-related instrumentation, the detailed analysis of the proteome for a specific type of cell (e.g., tumor, blood, or tissue) has the potential to permit the discovery of new protein biomarkers aimed towards diagnostic purposes and novel drug discovery [15,16]. Currently, the knowledge provided by proteomic research adds greatly to the genetic information generated from all genomics studies. The combination of proteomics and genomics research has the potential to play a major role in future biomedical studies and to impact the development of next generation diagnostic and therapeutic approaches.

Mass spectrometry (MS)-based proteomics has led to the possibility of characterizing and quantifying the protein profile of biological specimens, as well as the possibility to discover their complex interactions involved in various specific pathologies. For example, various proteomic approaches combined with genomic analysis have been used in cancer research for obtaining more information about the molecular basis of tumor genesis and the development of more effective anticancer therapies [17–19].

A typical bottom-up proteomics workflow (Figure 1) consists of several major steps: (i) isolation of the protein mixture from the studied biological sample, followed by (ii) quantification of isolated proteins concentrations (e.g., Bradford assay), then (iii) fractionation of proteins by gel electrophoresis or liquid chromatography methods. After fractionation, (iv) the proteins are proteolytically cleaved by enzymes (usually trypsin); followed by (v) a mass spectrometric measurement of the resulting peptides and (vi) a database search for protein identification.

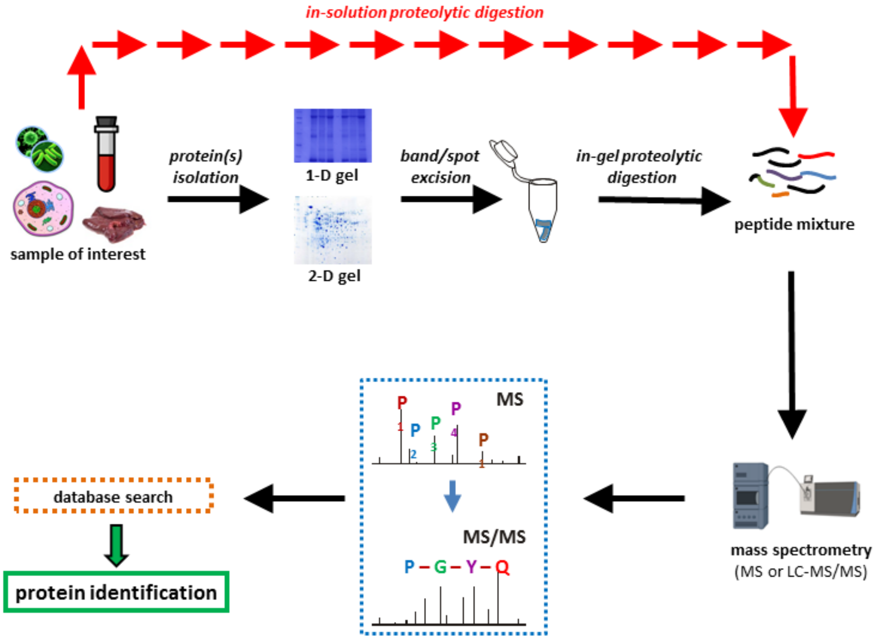

Figure 1. Workflow in a bottom-up proteomics experiment.

The purpose of this review is to briefly introduce the bottom-up proteomic approach from a liquid chromatography–mass spectrometry perspective, to discuss its multidisciplinary development, its strengths and weaknesses, and to identify potential areas of improvement in the current methodology. These potential improvements will be relevant to the current problems in society, including biomedical, clinical, or environmental concerns.

2. Bottom-Up Proteomics

Most proteomic analyses use proteases to digest proteins into peptides with a predictable terminus. Peptides are then analyzed in an MS/MS instrument and their mass to charge ratio and predicted sequence is used to infer information about the proteins in the sample. All experimental set-ups starting from the analysis of peptides from complete protein digests and using a protein database to characterize the open-reading frame this peptide originated from, are grouped under the term "bottom-up" proteomics.

Bottom-up proteomics utilizes the advantages that peptides have over proteins: peptides are more easily separated by reversed-phase liquid chromatography (RPLC), ionize well [20], and fragment in a more predictable manner. This translates into a robust methodology that enables high-throughput analysis, allowing for identification and quantification of thousands of proteins from complex lysates [21]. Today, bottom-up approaches using data-dependent acquisition (DDA) workflows are the core technologies in proteomics. Also termed shot-gun proteomics, these straightforward workflows generate large lists of protein identifications and were used for solving most of the available, complex, full proteomes available today, including the first drafts of the human proteome [22,23]. New data acquisition strategies such as selected reaction monitoring (SRM) further increase the quantification accuracy and reproducibility of bottom-up proteomics studies [24]. Data independent acquisition (DIA) is evolving to become a major driving methodology of the future. It attempts to combine a high number of protein identifications with consistency and accuracy in quantification of protein levels [25].

The hallmark of bottom-up proteomics is the extensive use of protease digestions, which has its drawbacks. Trypsin is the golden standard when it comes to shot-gun approaches and is being used for approximatively 96% of the deposited data sets in the Global Proteome Machine Database [26]. Trypsin is a very efficient protease with high catalytic activity that generates peptides with a basic arginine or lysine at the C-termini, which is ideal for collision-induced dissociation (CID) tandem mass spectrometry analysis [27]. Although efficient, when using trypsin, 56% of all generated peptides are ≤ 6 residues and thus too small to be identified by MS [28]. Moreover, only a fraction of these peptides yields useful fragmentation ladders [29]. The extensive usage of trypsin also leads to analytical tools that are optimized based on tryptic peptide properties. Furthermore, protein identifications are usually inferred by a limited number of typtic peptides. Hence, current bottom-up proteomics offers a rather biased and restricted coverage of the full proteome in a given sample, almost like "tunnel vision" of the proteome [28].

The limited sequence information inferred from small peptides is more often suffice to assign proteins clusters, and not always enough to identify proteoforms. Although possible through the measurement of mass shifts, protein isoform and PTM identification without prior knowledge is extremely limited in bottom-up proteomics.

3. Sample Preparation

As the proteomics field attempts to perform a comprehensive analysis of all the proteins in a given sample [30], sample preparation is the first and most important step in any experimental endeavor. If this step fails to perform, the sensitivity and throughput of the downstream steps are rendered pointless.

Genomics and transcriptomics, the other two major high-throughput -omics fields, benefit from the fact that nucleic acids are relatively uniform in terms of physiochemical properties. As such, "universal" standard methods and buffers are available that can be used to isolate nucleic acids of any size from almost any sample. No such thing exists in proteomics. The functional groups of the amino acid sidechains allow for such a great diversity in terms of protein charge and hydrophobicity, that a universal buffer will most probably never be available [31].

Nucleic acid-based technologies also benefit from the power of the PCR reaction, which allows the production of large quantities of target DNA in a single, reliable, and easy to do step; however, the proteomics field has nothing of the sort. Biological protein mixtures are characterized by a great diversity of proteins in chemistry, shape, and size, and, most importantly, by their range in abundance. Some proteins can be found in quantities as low as one protein per cell, while others up to several million per cell [32]. Despite the great selectivity and sensitivity of mass-spectrometers, no instrument available today can solve such complexity of range in a single measurement. Thereby, various fractionation strategies must be used in order to improve the depth and coverage of proteomic analysis [33].

3.1. The Good

Usually, methodologies for bottom-up proteomic sample preparation include the extraction of proteins from the biological matrix, removal of the nonprotein contaminants such as DNA, sugars, and lipids, removal of the residual salts that may form adducts during ionization, and protein fractionation to reduce sample complexity.

Bottom-up proteomics utilizes unfolded proteins that allow for easy access for proteases since amino acids are more readily available, thus, generating more peptides for MS analysis. In this case, tissue or cell lysis is performed directly in a buffer containing strong denaturants (such as urea or guanidine) and ionic detergents (sodium dodecyl sulfate, SDS, or deoxycholate-SDC). In addition, nonionic zwitterionic detergents (such Triton X-100, NP-40, digitonin, or CHAPS) further help solubilize membrane proteins and can be used when less denaturing conditions are required [34]. Depending on the sample, sometimes protein depletion or enrichment is a crucial step—usually mandatory for serum or plasma samples where albumin represents more than 50% of the proteins. Several methodologies are available for removing high-abundance proteins [35,36] or for enriching cell surface proteins [37], phosphoproteins [38], or various enzyme subclasses [39–41]. The techniques that provide the highest selectivity and sensitivity for the low abundant proteins are affinity chromatography [42] and immunoprecipitation [43], but their applicability depends on the availability of affinity supports or antibodies, as well as the quality of said antibody for the target protein.

For the removal of small molecules that may have originated from the biological sample or were added in the extraction/depletion/enrichment, many methods are available, including dialysis, buffer exchange, size exclusion, protein precipitation, chromatography, or electrophoresis [44,45]. Of all the desalting methods, precipitation with organic solvents (acetone or methanol/chloroform) is the least expensive, simplest, and most scalable option for desalting proteins prior to MS analysis [20].

Two forms of gel electrophoresis, SDS-PAGE and 2D-PAGE, have been extensively used in proteomics analysis as they solve two problems simultaneously—they remove salts, detergent, and other small molecules from the sample and they fractionate the proteins based on apparent molecular mass (SDS-PAGE) or charge and apparent molecular mass (2D-PAGE). SDS-PAGE has the advantage that it is inexpensive, straightforward, and highly reliable. Its separating resolution is higher than size exclusion chromatography, but not as high as 2D-PAGE. 2D-PAGE has a high resolution and, with the introduction of immobilized pH gradient (IPG) strips, its reproducibility has improved considerably, becoming the method of choice for most proteoform studies [46,47]. The development of PTM-specific stains for phosphoproteins (e.g., Pro-Q Diamond) and glycoproteins (e.g., Pro-Q Emerald, Dansylhydrazine) further increased the applicability of 2D-PAGE [48]. 2D gels are very useful for de novo sequencing of proteins from organisms with no genome sequences available and can complement the identification of protein isoforms and modified proteins [49].

3.2. The Bad

When preparing a sample for bottom-up proteomics analysis, many factors can cause the experiment to fail. Some examples include incomplete lysis, incomplete solubilization of the proteins, and contamination during sample handling with compounds that interfere with the downstream analysis steps, including MS. Some of the most troublesome compounds are detergents—both the detergents used for sample solubilization and those used for cleaning laboratory glassware. Detergent-contaminated samples and foul autosampler needles and LC piping reduce column capacity and performance and have poor ionization in both ESI and MALDI [20]. Triton X-100, Tween, or NP-40 contain polyethylene glycol (PEG) chains that elute throughout the LC and overwhelm the MS detector. SDS can cause complete signal suppression at levels as low as 0.01% [50]. Not all detergents are the same—some are considered safer for mass spectrometry, such as N-octyl-beta-glucoside and octylthioglucoside [51]. Moreover, several mass-spectrometry-compatible detergents are available, such as ProteaseMax (Promega), Rapigest (Waters), PPS Silent Surfactant (Expedeon), or Progenta (Protea), which degrade with heat and in the low pH of the acid containing LC-MS buffers [20]. Compounds leaching from poor quality plastics and fittings are also common contaminants. Phthalates, for example, are known to ionize well and overwhelm the mass spectrometer. An extensive list of contaminates in mass spectrometry is provided by Keller, 2008 [52].

In many proteomic experiments, the immunoaffinity methods are used to deplete the most abundant proteins. For example, there are many methods available that deplete albumin from human serum. Although these methods allow for the identification of a larger number of proteins due to the decrease in signal suppression, there may be problems with the nonselective loss of other proteins [53].

Keratins from skin and dust are another common contaminant, which takes up detector time, clutters the raw data, and reduces the number of useful spectra recorded. Keratins are commonly introduced when casting and handling gels and gel pieces [54]. It is thereby imperative to wear gloves and scrupulously wipe down surfaces to minimize this contamination. Other frequent proteinaceous contaminants are lysozyme, DNase, and RNase introduced in the cell lysis step, bovine serum albumin and other weight marker components from electrophoresis, as well as trypsin and other proteases from the digestion step [52]. A significant amount of MS instrument time is spent sequencing peptides from these abundant contaminant proteins and not peptides from the actual sample of interest [55]. Exclusion lists have been put together in an attempt to cleverly remove the burden of these contaminants [55]; however, these lists are useless in certain instances when, for example, keratins are expected to be present in the sample to be analyzed. In this case, it is difficult to distinguish between endogenous and contaminating keratins [56].

SDS-PAGE is an easy, inexpensive method to separate proteins prior to mass spectrometry analysis; however, its separating power is not very high. Although 2D PAGE has better separating power, it is still not uncommon to have multiple proteins in the same spot [57]. Along with this, proteins with a low copy number and hydrophobic proteins are difficult to detect in 2-D electrophoresis [58].

Ultimately, difficulties in sample preparation for bottom-up proteomics analysis are a result of the huge variability of protein properties. There is an overwhelming number of methods and technologies available, each tailored for specific samples and protein groups. With such variety to choose from, one will have a hard time deciding which method to pick and how they should adapt it to fit the specific sample. Moreover, most of the methodologies are long, multistep preparations that are prone to losses, biases, and contaminations, while being extremely time-consuming and labor-intensive. In the end, all of this ultimately contributes to poor reproducibility of the current sample preparation and fractionation methods [59].

3.3. The Future

Ideally, what is needed is a reliable and universal method to extract and fractionate proteins; one that is fast, does not require detergents, requires minimal steps and sample handling, and can be automated. This may be achievable with the development of new approaches that rethink sample preparation in bottom-up proteomics. Sample Preparation by Easy Extraction and Digestion (SPEED) is a new approach that does what its name suggests—reduces sample preparation to three easy steps: (1) acidification with pure trifluoroacetic acid (TFA) to lyse cells and extract proteins, (2) neutralization with TRIS base, and (3) digestion with proteases to generate peptides [59]. All steps and buffers required for sample preparation can be integrated for a straightforward and possibly automated sample preparation. Depending on the protein digestion mechanism, three types of integrated sample preparation methods are emerging: (1) in solution digestion, (2) immobilized-enzyme-reactor, and (3) on bed digestion methods [60]. Integrated in-solution digestion methods include filter-aided sample preparation (FASP) that repurposes centrifugal ultrafiltration concentrators in order to remove detergents, perform protein cleavage and isolate peptide fractions [61,62]. Derived from FASP, an encapsulated in-StageTip (iST) device has an enclosed tip chamber with an inserted membrane where lysis, denaturation, and alkylation take place [63]. The membrane serves both as a filter and as separation support. NanoPOTS are microfluidic chip devices that use nanodroplet-based processing methods. They allow in-solution digestion and processing of single cells and are reproducible [64]. Immobilized enzymatic reactor (IMER) methods make use of special columns that contain immobilized trypsin and can be directly coupled to LC-MS system for automated protein digestion and online LC-MS/MS analysis [65]. On-bead digestion methods perform reduction, alkylation, and digestion steps on functionalized beads by trapping proteins into a very limited void volume. Suspension trapping (STrap) is a method that uses tips packed with a quartz or glass filter and a hydrophobic C18 layer to perform sample clean-up and protein digestion in one step [66]. Single-Pot Solid-Phase-enhanced Sample Preparation (SP3) uses paramagnetic beads for the same purpose [67] and can be easily automated using a liquid handling robot [68]. These methods can be easily automated and integrated.

4. Conventional HPLC/Modern UHPLC Fractionation

Currently, liquid chromatography coupled with mass spectrometry (LC-MS) is a common and indispensable analytical technique for proteomic investigation. This coupling initiated the development of novel ionization methods and led to a broad range of interfaces aimed at separating various biological complex mixtures [69]. Atmospheric pressure ionization (API) was the first method to directly interface a solution stream with a mass analyzer [70]. Several thermospray units were also introduced as a breakthrough for modern LC-MS [71,72].

4.1. The Good

The main applications of HPLC in proteomics are rooted in the concept that peptides can be separated over a time and buffer gradient. This allows for maximum identification by the mass spectrometer, helps with the problem of ion suppression from coeluting peptides, and helps in the identification of the individual protein sequence. Significant research has been aimed towards developing HPLC methods to enable the resolution and identification of all generated peptides from digested proteins in a given proteome. This is a difficult task for many reasons. For example, a serum proteome may contain up to 20,000 proteins with a concentration dynamic range of 10^{11}, which, by proteolytical digestion, may result in more than 600,000 peptides [73] excluding PTMs. Fractionation of such complex peptide mixtures is a critical aspect of their mass spectrometric identification. The general principle of HPLC fractionation of complex peptides mixtures is based on their interaction with a stationary phase (column) and a mobile phase (solvent gradient elution). The three major HPLC modes developed for peptide fractionation or sequential separation utilize differences in peptide size (Size-Exclusion Chromatography (SEC)), net charge (Ion-Exchange Chromatography (IEX)), or hydrophobicity (Reverse-Phase Chromatography (RP-HPLC)) [74,75].

The well-established proteomic methods approaching structural characterization of proteins (top-down and bottom-up approaches), have led to the use of HPLC methods in a particular mode [76]. For example, SEC [77] was used in top-down approaches to study intact proteins as a favored method for size-based separation. SEC has been mostly selected to analyze antibody–drug conjugates to determine their purity [78] and to purify recombinantly expressed proteins [79]. Since SEC is considered a low-resolution chromatographic method that requires diluted samples, the SEC technique was employed in a combination of columns with different pore sizes to achieve sufficiently high resolution in the separation of a complex protein mixture with a broad molecular weight range (10–223 kDa) [80].

Moreover, the combination of SEC with RP-HPLC, in a two dimensional (2D) separating platform exceeded a one-dimensional (1D) RP-HPLC experimental run with 4044 more unique proteoforms identified in a sarcomeric protein mixture [80]. This multidimensional HPLC was introduced as a shotgun approach to analyze very complex protein or peptide mixtures without performing gel electrophoresis, but achieving the same separating resolution as bidimensional gel electrophoresis (2D-PAGE) [81]. In the so-called multidimensional protein identification technology (MudPIT) [82,83], a protein mixture was subject to specific enzymatic digestion, usually using trypsin and endoproteinase LysC, and the resulting peptide mixture was separated by strong cation exchange (SCX) and reversed-phase high performance liquid chromatography (RP-HPLC) [84,85]. A combination of HPLC, liquid phase isoelectric focusing, and capillary electrophoresis was reported as a multimodular approach for obtaining a better separation of complex protein mixtures [86]. Recently, the large dynamic range and complexity of human cancer cells was overcome by employing a 2D separation using a high pH and low pH reversed-phase liquid chromatography technique, and 2778 proteoforms from 628 intact proteins were detected [87].

RP-HPLC is used in most bottom-up proteomic experiments for separating proteolytically generated peptides due to its high peak capacity, reproducibility, and robustness [88,89]. Both bottom-up and top-down proteomics approaches are fully dependent on employed separation technologies to: (i) provide large-scale proteome coverage in a given time; (ii) accomplish higher analytical throughput; and (iii) cover a broad dynamic protein concentration range, including trace amounts of distinct proteins. Many proteomic approaches use a IEX/RPLC [90,91] combination where octadecylsilanes (C18) remains the preferred RP ligand and the choice of ionic ligands experimentally depends on the class of peptides to be enriched and/or fractionated.

The development of an HPLC column with an “ideal and perfect” peptide/protein separation capability is an important, continuous research objective. The main types of analytical columns currently used in proteomics research present different characteristics with regard to the material composition and particle size packed into it, and the length and diameter [92]. Remarkable separation of complex peptide mixtures in proteomic studies was achieved using nano-LC/UPLC and capillary columns containing packed alkyl bonded C8 or C18 [93] and silica-based monolithic capillaries [94,95]. In proteomic analysis, the ability to handle very small amounts of biological material is crucial. Miniaturized HPLC separation systems have been developed by using fused silica capillary columns [96,97] or chip-based devices [98–100]. Monolithic capillary columns (and other columns, such as C8 or C18 columns) can be produced in the laboratory without the need of expensive media, packing solvents, and high-pressure packing instrumentation with the proper training [101]. Even if reproducibility issues of published protocols to produce such columns were reported, they remain cost-efficient and depend on the researcher’s capability to pack it perfectly.

Separation time is an important factor in nanoflow LC/tandem MS technologies, which has been shown by Mann et al. [102] where the results on the yeast proteome were compared using different mobile-phase gradient running times, identifying 5806 peptides in 140 min and 13,682 peptides in 480 min. Furthermore, improved fractionation in chromatography was shown as a straightforward approach to separate coeluting peptides and the pH of the mobile phase was shown to have an important impact on the retention, selectivity, and sensitivity of the separation [103].

In 1997, MacNair introduced a hybrid stationary phase for ultrahigh-pressure reversed-phase liquid chromatography (UHPLC) for rapid separation [104]. The difference between UHPLC and traditional HPLC is that UHPLC uses smaller diameter packing material in the column and higher pump pressures than HPLC. In UHPLC separation, the size of the particles is usually approximately

1.7 µm in diameter, versus the 3–5 µm diameter of HPLC column particles and requires a pressure of up to 1000 bar, versus the 50–600 bar range that is typical in HPLC instruments [105]. UHPLC has raised the level of performance of the separation with significant gain in resolution, speed, and sensitivity. The use of UHPLC doubled the peak capacity and increased the separation speed over nine-fold and the sensitivity by three- to five-fold, as compared with an HPLC run [106]. The problematic thermal effect occurring in the column by using high pressure in UHPLC system was overcome by their narrower diameters. However, as column diameters decrease, technical problems such as (i) sealing the LC system to resist leakage, (ii) very high backpressures at high flow rates, and (iii) contamination were reported. In terms of primordial advantages of UHPLC over HPLC, there is the faster speed and less solvent consumption, but its evident disadvantage is the price [107].

4.2. The Bad

A problem that remains in part unresolved by liquid chromatography is membrane protein solubility, which requires the use of special solubilization agents such as acidic solvents, detergents, and/or chaotropes that need to be diluted or even removed prior to proteolytical digestion or mass spectrometric analysis [108].

To optimize fractionation protocols using multidimensional separation, several factors need to be considered, including the impact of ionic strength, buffer capacity, pH response on the retention time behavior, and peak shape of proteins or peptides [109]. Moreover, when using multiple SCX /HPLC separation, the elution of the peptides from the SCX column is not precise and the same peptide will appear in several of the subsequent HPLC runs, reducing the amount of that peptide and thus decreasing the sensitivity. In samples that are too complex/concentrated, the first few buffer injections will contain the majority of singly charged peptides and di/triprotonated peptides are usually not trapped [110]. The major obstacle for liquid chromatography in using multiple parallel columns is adjusting the hardware and/or software to work in a synchronic manner. The price of the columns and their stability is not to be neglected. Future contributions in column technology require important developments so that LC is not restricted to the conventional single-column fractionation methods.

Broadly, pharmaceutical and clinical laboratories seem to be willing to sacrifice resolution to gain analysis speed. This sacrifice may not be acceptable in fundamental research when analyzing complex peptide/protein mixtures to provide data for developing novel drugs or when identifying new biomarkers with diagnostic purposes. HPLC/UHPLC will continue to be a pivotal analytical technique that, in combination with high resolution mass spectrometry, will further raise the level of performance with significant increases in resolution, speed, and sensitivity required for elucidation of complex biological of proteomes [111].

4.3. The Future

HPLC/UHPLC-MS will remain very useful in "omics" sciences and reference methods will be developed, but the transition to various clinical applications is more desired. In recent years, immunoassays have been progressively replaced by HPLC-MS due to higher sensitivity, significantly less false positives, and the reduced costs of used reagents because multiple analytes can be measured simultaneously [112]. On the other hand, some drawbacks of HPLC-MS analysis include the high cost of instrumentation and the necessity for the initial method development and rigorous validation. As of today, universal chromatographic methods are not available, and each laboratory has to develop their own methods critically depending on their available infrastructure. The improvement of proteome coverage by first reducing sample complexity via chromatography will be a great challenge to researchers. One important requirement that gets increasingly supported by the literature is the fact that the current column hardware is no longer adequate to maintain the very high efficiencies and small peak volumes produced by the high-quality particles (continually smaller) and high-quality packing procedures. The proteomics field has also been utilizing very slow microflow rates in order to increase sensitivity, which cuts down on buffer costs, but also increases analysis time [113].

As several method parameters, such as pH, temperature, buffer concentration, and gradient time are varied simultaneously, a revolutionary software that predicts chromatograms based on HPLC method development is desired to determine the behavior of the separation. This is necessary in order to optimize complex sample separation and to economize resources spent developing and running wet lab activities. With the help of bioinformatics developments, in silico chromatography will be of great help in the near future to predict LC experimental flows for avoiding time, solvent, and sample consumption normally used in optimizing protocols for complex biological peptide/protein mixtures. Recently, a predictive algorithm of peptide and protein retention times in reversed-phase chromatography [114] was reported as a complementary experimental tool for proteomics.

5. MS Analysis: Instrumentation

MS-based proteomics is one of the chosen methods for complex protein sample analysis. It has established itself as a superior technology for complete characterization of proteins [29]. It is widely used for sequence analysis, protein–protein interactions, and identifying PTMs [29]. In general, a mass spectrometer consists of an ion source to ionize the analytes, a mass analyzer to measure the mass to charge ratio (m/z) of the analytes, and a detector that detects the number of ions at each m/z value. Electrospray ionization (ESI) is a commonly used technique to ionize peptides or proteins for MS analysis [115]. It ionizes a liquid solution of sample and hence can be coupled to liquid chromatography for separation [29,115]. There are four different type of mass analyzers used in proteomics: ion trap, quadrupole, time of flight (TOF), and Fourier transform ion cyclotron (FT-MS). These analyzers are the key to maintain sensitivity, mass accuracy, resolution, and to generate information rich ion mass spectra (MS/MS spectra) from peptide fragments [116]. They can be used individually in instruments or combined with each other to take advantage of the strengths of each [29,116].

It is also necessary for peptides be broken up further before being analyzed by the mass spectrometer, by a process known as dissociation. Different dissociation techniques exist for mass spectrometers, including (1) collision-induced dissociation (CID), (2) electron-capture dissociation (ECD), (3) electron-transfer dissociation (ETD), and (4) higher-energy collisional dissociation (HCD).

5.1. The Good

There is a diverse range of MS instruments that cover many possible applications in proteomics. Recent developments in the instrument design have led to new ion activation techniques as well as allowing for lower limits of detection. It has also increased the ability of tandem mass spectrometry for peptide and protein structure elucidation by improving the understanding of gas-phase ion chemistry [117]. In addition, the dynamic range of instruments allows one to optimize LC-MS and LC-MS/MS methods such as Data Dependent Analysis (DDA), Data Independent Analysis (DIA), Selected Reaction Monitoring (SRM), and Parallel Reaction Monitoring (PRM).

DDA is a common data acquisition strategy that selects the most abundant precursor ions for MS/MS analysis [118]. It takes the selection of peptide signals forward for fragmentation and matches them to a predefined database. The method allows for minimal selection of redundant peptide precursors [119]. The semirandom peptide sampling phenomenon in a DDA method has shown to increase the rates at which new peptides are identified, especially during replicate analysis. However, the reproducibility of the low abundant peptides between the runs, remain a challenge in the random sampling situation [118,120].

DIA is a method in which all the peptides within a defined m/z frame are subjected to fragmentation [121]. This method allows for accurate peptide quantification without being restricted to profiling only the predefined peptides of interest. In addition, DIA has the potential to overcome the random sampling problem and to reproduce and quantify low abundant peptides [121]. DIA offers several advantages over DDA for characterizing complex proteins. Unlike DDA, which sequentially detects, selects, and analyzes the individual ions, DIA systematically parallelizes the fragmentation of detectable ions within an m/z window, regardless of their intensity, thus providing a broader dynamic range of detected signals, improved reproducibility for identification, better accuracy and sensitivity for quantification, and enhanced protein coverage [122].

SRM is an MS-based technique for quantitative analyses. It is a nonscanning technique where selectivity is increased through fragmentation [123]. This method has the benefit of being able to control error rates in discovery proteomic experiments and can efficiently generate specific and quantitative assays for a large number of proteins and their PTMs by being both cost and time efficient [124]. It serves as a benchmark to all time segment methods [125]. PRM is another targeted method of quantitation performed using high-resolution mass spectrometers such as quadrupole-Orbitrap (q-OT). The development and application of higher-energy collisional dissociation (HCD) fragmentation enables MS/MS spectra to be acquired in the Orbitrap analyzer with high mass accuracy and high resolution. HCD is a beam-type collisional dissociation similar to the dissociation achieved in QQQ as well as QTOF mass spectrometers. An advantage of using q-OT is that both the discovery and targeted experiments can be performed on the same instrument, and it is convenient to transfer instrumental parameters such as collision energy, retention time, quadrupole isolation window, etc. [126]. This approach enables the acquisition of full MS/MS spectra of a targeted peptide with high mass accuracy and resolution, thus quantifying highly specific proteins. Similar to SRM, PRM can also validate the abundance of proteins and their PTMs [126]. While SRM and PRM are comparable, PRM is the most suitable for an attomole-level detection and quantification of multiple proteins in a complex sample [127]. Its simple and straight forward acquisition method in addition to its high selectivity and specificity of data acquired from high resolution and high mass accuracy makes it a powerful quantitation method [128]. Thus, one can choose from one of the previously mentioned methods based on their specific needs and optimize these methods in the instrument in order to acquire data in a short period of time with accurate peptide quantitation.

Recently, the combination of the linear ion trap with the Orbitrap analyzer has benefitted the advances in high resolution MS greatly [129]. It can achieve high mass resolutions within a fraction of a second, which is important in both qualitative and quantitative applications. Moreover, it allows for the identification and quantification of a compound even in the presence of a background ion that has a nominally identical mass [129]. Complementing the ion trap and Orbitrap combination, the quadrupole mass filter was coupled to an Orbitrap analyzer known as the “Q Exactive” instrument, which features fast collision-induced dissociation peptide fragmentation with high energy because of parallel filling and detection modes, and features high ion currents because of an S-lens. This combination aids multiplexed operation at the MS and tandem MS levels, enabling joint analysis of HCD fragment ions in the Orbitrap analyzer and fragmentation of different precursor masses by the quadrupole analyzer, overall making Q Exactive an exciting instrument for proteomics [129]. SWATH-MS is a DIA–LC-MS technique that has become prominent for quantitative proteomics mainly due to its high quantitation accuracy, increased peptide coverage, generation of a digital map, and excellent reproducibility. In addition to that, it also permits qualitative analyses and allows for small molecule applications such as forensic analysis and the identification of metabolites and metabolomics [130]. MS instruments are usually compared to Nuclear Magnetic Resonance (NMR) instruments because they are expensive and require regular maintenance and troubleshooting. However, the robustness, the relative simplicity of sample preparation steps, high-throughput analysis, and sensitivity of MS remains unmatched [131,132].

Ion mobility mass spectrometry (IM-MS) uses electric fields to drag analytes through a buffer gas, separating the analytes whilst providing structural information, and has had many improvements in recent years [133]. The extra dimension that ion mobility provides increases the peak capacity in LC-MS workflows [134]. It is most common to couple a TOF mass analyzer with ion mobility; however, other analyzers can be used.

5.2. The Bad

Mass spectrometers are very expensive, delicate, and require a significant amount troubleshooting and maintenance. Although they are known to produce extensive information on proteins, the abundance of data can give rise to negatives, false positives, and unassigned spectra. There is a dynamic range of instruments to cover all possible applications; however, there is not one specific instrument that can perform all kinds of experiments [132]. MS detection oscillates with the concentration range of 10^4 and 10^5 and sometimes even 10^7. Overall, the complexity, dynamic range of biological samples, and low abundance of disease-specific biomarkers remains a major challenge for proteomic biomarker discovery, and there is no MS instrument that can simultaneously address these challenges efficiently [135]. In addition, complete characterization of the proteome both quantitatively and qualitatively remains a challenge [136]. These problems can be remedied by reducing the sample complexity before introduction into the MS. Another major challenge is the analysis of a large number of samples in clinical studies or discovery proteomics where there is often experimental variability among the clinical samples. Additionally, the impact of single nucleotide polymorphisms (SNPs) on proteome analysis has still not been fully investigated [136]. Proteomics experiments for hundreds of samples are expensive and time consuming [132]. Very few proteomics techniques allow for high throughput analysis while simultaneously maintaining the sensitivity and robustness. DIA/SWATH, as mentioned above, is an alternative for proteomic analysis of clinical samples on a large scale [132,137].

5.3. The Future

Although mass spectrometers have the upper hand concerning their high throughput analysis and robustness, there is a need for MS with better sensitivity, resolution, and accuracy for analyzing samples such as proteins in the blood plasma. This will further allow the instrument to detect low-abundance proteins and their PTMs or interacting partners [132]. In addition, conventional MS has trouble resolving multiple charge states for species such as protein complexes of several hundred kilodaltons, especially for heterogenous samples such as heavily glycosylated proteins. Instruments such as charge detection mass spectrometry (CDMS), a variant of MS can help overcome such complexities by allowing the detection of m/z and charge states, making it easy to determine the mass of ions in a sample [138]. Advances in MS methods will aid in improving the chances of biomarker discovery for MS-based proteomics. Bottom-up proteomics is the workhorse for proteomic analysis. The middle and top-down approaches must advance in order to completely characterize the protein isoforms as well as the PTMs [136]. Overall, continued improvements are needed in both MS methods and technology in order to overcome the aforementioned challenges.

6. Analysis of Mass Spectrometry Data

Analysis of MS data proves to be a limiting factor in many proteomics experiments. Although direct analysis of the raw data can be very useful, this practice can be tedious and time-consuming. There are many different software pipelines available for a more high-throughput and time-efficient approach; however, these programs have a great number of limitations. Although there are still many obstacles to overcome, in recent years, the limitations in proteomic data analysis have lessened, and high-throughput data analysis continues to improve tremendously.

It is important to understand the basics of peptide fragmentation before one can grasp a firm understanding of the current problems in proteomic data analysis, since most problems arise from lacking complete sequence information for many proteins. Peptides produced by enzymatic (i.e., trypsin) digestion will have between 5 and 20 amino acids. During collision-induced dissociation (CID) MS/MS fragmentation, the peptide is fragmented into many smaller fragment ions. In CID, these fragments are called y ions and are numbered from y1 (the C-terminal amino acid) to y(n), where (n) is the maximum number of amino acids in the peptide. A second type of ion in CID is the b ion. These ions start from the N-terminus, and end at the C terminus of the peptide. Both y and b ions are produced through fragmentation at the peptidic bond—the weakest bond in peptides (Figure 2). Additional ions are also produced in CID and observed in MS/MS: a and c (and b) ions from the N-terminus of the peptide and x and z (and y) ions from the C-terminus of the peptide (Figure 2). Predictability of peptide fragmentation (and production of the y, b, and a ions) allows for identification of peptide sequences using the de novo sequencing and in almost any database search algorithm. Different dissociation techniques produce different types of fragment ions. For example, ECD is complementary to CID as it provides more extensive sequence coverage. Disulfide bonds are preferentially cleaved in this method, and PTMs tend to stay intact; however, this method is specific to FTICR MS instruments [139]. Similarly, ETD fragments in the same type of way, creating longer c and z-type ions and preserving PTMs, but instead uses an RF quadrupole ion trapping device [140]. If used together, these dissociation techniques can give complementary information.

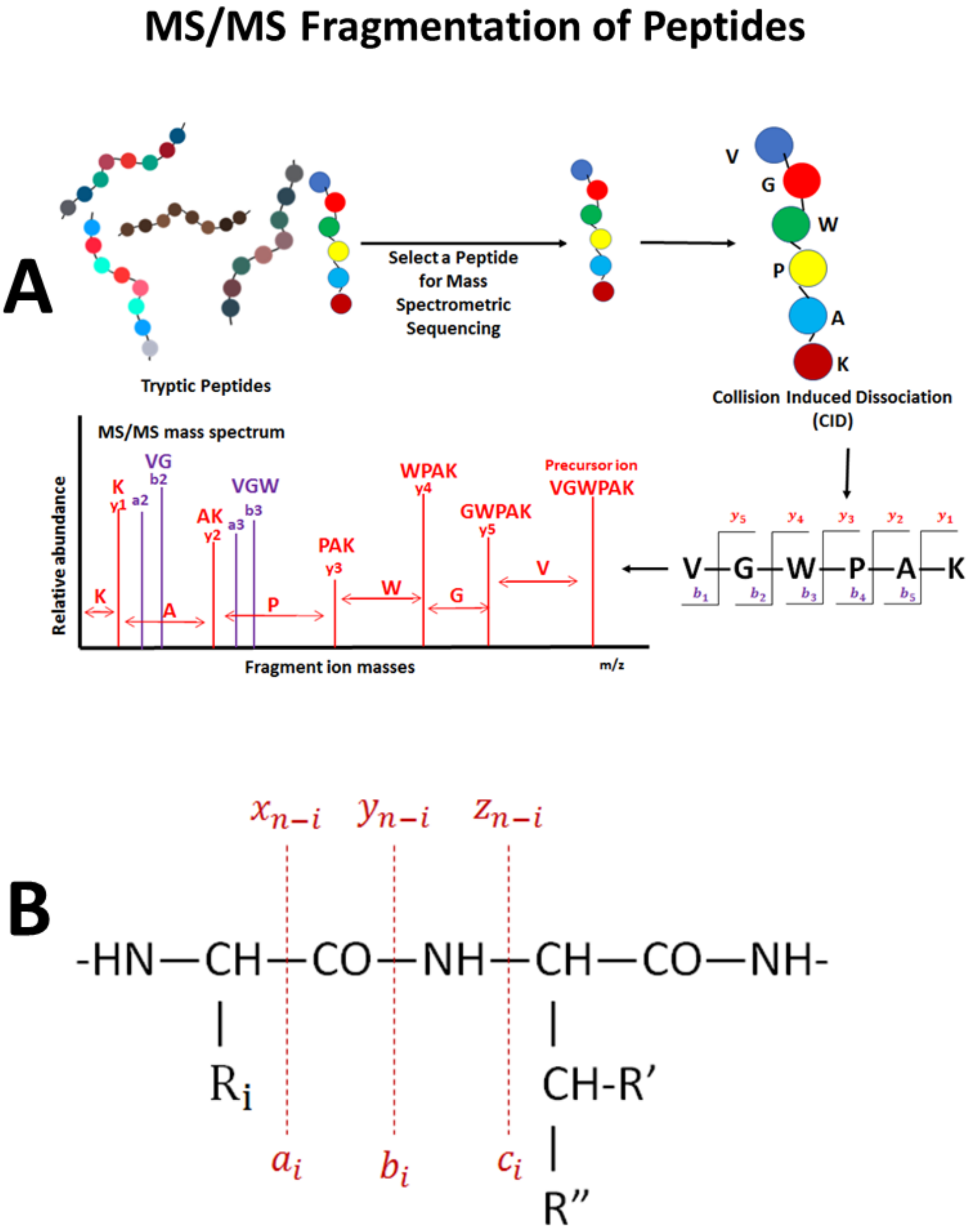

Figure 2. Correspondence between the peptide’s amino acid sequence and the fragment ion peaks that are produced in MS/MS (**A**) and the major theoretical fragment ions that are produced from fragmentation of a peptide (**B**).

A particular advantage in proteomics-based analysis is that one protein upon enzymatic digestion can produce many peptides that can be useful in both identification and characterization of that protein. For example, if a protein can theoretically produce 50 peptides that can be analyzed by MS, only a few of those peptides are needed to identify that protein. These peptides can be native peptides and modified peptides (i.e., with Methionine oxidized), or can be native peptides with a different charge state. In addition, identification of these additional charge states for one peptide or unmodified and modified peptides can be particularly useful in characterization of that protein. For example, most therapeutic proteins have at least one or two methionines within their sequences. Knowing whether they are oxidized or not may be useful for long-term storage of those proteins since a methionine does not oxidize if it is buried inside the protein because it is not exposed to solvent. Having two or three charges may be useful in de novo sequencing, when the number of charges indicate the number of amino acids that can be protonated (i.e., only Arg, Lys, and His are/can be protonated).

6.1. The Good

In a typical proteomics experiment, the raw data file is usually dependent on the MS instrument used and it should be first converted into a universal, readable file. More often, this file is a peak list and will concisely show the mass to charge ratio of precursor ions with their relative intensity and charge, as well as the mass to charge ratio of its fragment ions and their relative intensities. OmicX is an online a resource that lists different -omics software. Once the readable file is obtained, it can be submitted for a search against a target database containing theoretical peptide sequences. These sequences are generated by computationally performing theoretical digestions and MS/MS analysis of all the possible proteins/resulting peptides from a given genomic dataset. The search output is a list of peptide-spectrum matches (PSMs) that are used to identify individual peptides. The PSMs are then further used to infer the identity of proteins present in the sample [141]. Such searches can be performed through PLGS or through other programs such as Mascot (Matrix Science, London, UK), which has a subset of programs like the free Mascot Server, Mascot Daemon, Mascot Distiller, and Mascot Parser. This program can be used for protein identification and scoring. In this software, it is also possible to set the parameters to search for certain PTMs based on the mass shift. It is also possible to customize PTM searches to match specific experiments. In 1993, the Yates lab proposed a method that could be used to correlate mass spectral data with a predicted amino acid sequences in a protein database [142]. Following, a program known as Sequest came into the works [143]. PEAKS (Bioinformatics Solutions Inc, Waterloo, ON) is another protein identification software that also has a de novo sequencing function. Comet [144], is another search program which became publicly available in 2012. For experiments requiring relative quantification and statistical analysis, programs such as Scaffold (Proteome Software Inc., Portland, OR, USA) can be utilized. MaxQuant is an important quantitative software that is freely available [145].

Although typical database and spectral library searching are the main tools used for proteomic data analysis, in some instances, de novo sequencing of the raw MS/MS spectrum data is used to identify peptides in instances of novel proteins, mutations, and PTMs. This approach involves acquiring the data and determining mass and composition of peptides directly from the MS/MS spectrum and predicted fragmentation [146]. Current de novo sequencing algorithms include PEAKS [147], Lutefisk [148], PepNovo [149]. Mascot Distiller [150], Protein Prospector [151], Novor [152], UniNovo [153], and PeptideProphet [154].

Proteogenomics, first seen in 2004, in an approach that uses genomic and transcriptomic sequence information as a reference for MS/MS spectra to identify novel peptides [155]. Some examples of proteogenomic approaches include six-frame translation, ab initio gene prediction, and expressed sequence tags. The ability to pair genomic, transcriptomic, and proteomic methods has been the key factor in many discoveries [156–159].

The alignment of amino acids in proteins from different species can be a beneficial characteristic in proteomic analysis. If a well-developed protein database does not exist for a certain species, protein databases of similar species can be used. The protein hits that are obtained from these other databases can be used as a reference to determine what proteins are present in the species of interest; therefore, databases can be utilized that may not be specifically for the species of interest [160,161]. PEAKS software has a SPIDER algorithm that allows for cross-species homology search and detection of peptide mutations [147].

Many different approaches exist for quantitative proteomics experiments. In a simple, label-free method, a known concentration of an internal standard peptide can be spiked into samples prior to mass spectrometry analysis [148]. Likewise, an external standard peptide sample can be run in between samples on the mass spectrometer, and if the spectral count intensities of the standard runs are the same, the spectral counts/intensities of peptides in the actual samples can be compared as a relative quantitation [162–165]. In this case, the standard acts as proof that running conditions are identical enough that relative quantitation using spectral intensities can be used. There are also many available software for label-free quantitative proteomics, including MapQuant, MZmine, MsInspect, OpenMS, MSight, SuperHirn, and MaxLFQ [166–172]. The advantage of these types of analysis is that they are affordable and time-efficient. This type of quantitation works well enough for pilot and preliminary studies, but findings should be confirmed later using a more precise, reliable method. Labeling methods for quantification are also abundant, including methods such as isotope-coded affinity tag (ICAT), stable isotope labeling by amino acids in cell culture (SILAC), ^{15}N/^{14}N metabolic labeling, ^{18}O/^{16}O enzymatic labeling, isotope coded protein labeling (ICPL), tandem mass tags (TMT), and isobaric tags for relative and absolute quantification (iTRAQ) [173]. The absolute quantification (AQUA) method allows for the precise determination of protein expression and even post-translational modification levels by mimicking the exact peptide of interest, with the exception of stable isotope enrichment [174].

6.2. The Bad

The aforementioned type of proteomic data analysis is based off the premise that peptides are identified by matching their m/z information to a library of known proteins and their theoretical fragmentation. Similarly, peptides can be identified by matching MS/MS spectra against a library of theoretical spectra for that cell, tissue, organism, etc. which is a slowly growing practice. Therefore, it is assumed that all protein-coding gene sequences are known and well-annotated, and that all proteins encoded by these sequences exist in a database. Such databases include NCBI RefSeq or UniProtKB [175]. Unfortunately, this assumption is far from reality. The PTMs are not easily identifiable from genetic data; therefore, most modified peptides for the proteome of most organisms do not exist in a protein database, which makes proteomic data analysis a tricky task to tackle. As previously stated, de novo sequencing is a method that can be used in this instance. Although de novo sequencing is a very thorough type of analysis, problems with this technique include incomplete peptide fragmentation and low mass accuracy. These problems have been improved slightly in the most current software, but not totally. The downfall of proteogenomic approaches like six-frame translation, ab initio gene prediction, and expressed sequence tags is that they make very large reference databases that can sometimes be hard to work with [175].

Although the alignment of amino acids between different species can be beneficial, it can also be a complication in data analysis because most peptides are not unique to one specific protein or proteome. The NCBI Basic Local Alignment Search Tool (BLAST) shows regions of similarity between protein sequences [176]. Proteins of two different species can have more than 99% alignment of amino acids if they are closely related and can still have much more than 80% alignment if they are distantly related. The quality of alignment and conservation of sequences between proteins is interesting because it can hint towards species homology; however, it can be a complication when analyzing proteomics data.

When trying to perform quantitative proteomics experiments, the tools needed for a reliable analysis can be expensive. The cheapest way to quantify peptides is a label-free approach; however, variable running conditions for the mass spectrometer can lead to errors [177]. The amount of instrument time required for technical replicates is also a downfall when it comes to label-free approaches. In addition, it is not possible to accurately quantify the relative abundance of different peptides since these molecules have different physicochemical properties and thus will behave differently under mass spectrometry conditions [178]. The different labeling methods for quantitative proteomics also have their flaws. For example, ^{15}N labeling relies on the number of nitrogen atoms present in the proteins, which is variable and can complicate analysis. A disadvantage of the SILAC method is that its cell culture origin limits its capabilities [178]. A universal problem with most labeling approaches is that there may be incomplete labeling of the peptides, which would cause errors in the quantitative analysis.

6.3. The Future

One of the major problems that needs to be addressed in proteomic data analysis is the lack of well-annotated protein databases. It is necessary for researchers to help develop these databases so that they can be used in future proteomics studies. When using a reference database for proteomic analyses, strict, relevant search parameters should be followed to decrease the likelihood of false positives. Whenever possible, smaller protein databases should be used instead of larger ones to minimize the false positives, peptide thresholds should be set to a reasonable tolerance, and relevant PTMs should be taken into consideration. In addition, if peptide identifications are particularly important, it is wise to verify them using the raw data spectrums.

7. False Positives, False Negatives, and Unassigned Spectra

The final and most biologically relevant result of a proteomics experiment is the identification of the proteins present in the sample. This is a two-step process that includes finding PSMs and inferring protein identity based on sequence alignments. Each step is dependent on a database, and, in each step, as the database grows larger, the probability to find peptides and hence proteins due to chance alone increases. Hits that are not actually present in the sample are called false positive identifications and used to be the most challenging problem in proteomics.

7.1. The Good

In order to assess the quality of the PSMs and hence peptide identity, the false discovery rate (FDR) was introduced as the ratio between false PSMs and total number of PSMs. The lower the FDR, the more meaningful the results. Two major strategies for calculating FDR are available: target-decoy search strategy (TDS) [179] and mixture model-based methods [180].

In TDS, fictious decoy peptides are fabricated by reversing or shuffling the protein sequence from the target database. This decoy database is also used for searches, and the FDR level is controlled based on the amount of decoy peptides found in the search output [181]. This approach is the most widely used, although it has several disadvantages. First, as the target database grows larger, so does the decoy database, which increases search time and complexity. Some strategies for reducing [182] or eliminating the decoy database altogether [183] are available and have yet to find their way into the mainstream. Secondly, decoy database searches are more often biased by different peptide features or score functions [181], but some work has been done to overcome this issue [182].

When it comes to protein inference, many effective protein inference algorithms have been developed, such as ProteinProphet, ComByne, and MSBayesPro. For quality assessment of protein identification, two methodologies are available: p-value-based approaches that calculate a probability parameter for each identified protein, and FDR approaches that apply a single threshold to all proteins identified [141]. Previously, one of the major sources of false positive protein identification was identification of peptides based on a reduced number of fragment ions. However, this is no longer a problem, since identification of a protein based on one peptide alone is no longer acceptable within the scientific community. This strategy clearly eliminates true identification of some proteins, but it also eliminates any doubts in the proteomics dataset.

The most commonly used computational strategy is to use database search algorithms based on searching acquired MS/MS spectra against a protein sequence database [184]. However, a large fraction of spectra remains unexplained despite a substantial improvement in the quality of MS/MS data acquired on modern mass spectrometers [184]. Novel sequences and the vast diversity of PTMSs remain unidentified in traditional database searches [185]. To develop a fast computational strategy that could be broadly applicable for searches using wide precursor mass tolerance of hundreds of Daltons (open database search), a novel fragment-ion indexing method was designed that provides orders of magnitude improvement in speed over existing tools [186]. This method implements a new database search tool called MS Fragger, which can perform open searches with variable modifications, thus identifying peptides with unknown alterations. It further allows open searches for data sets containing millions of MS/MS spectra and is applicable to data obtained from labeling-based qualitative proteomics experiments [186,187]. Open database searching offers a potential solution to the problem of inaccurate FDR estimates due to unaccounted peptide modifications in traditional narrow mass-window searches [186]. Using an independent platform tool like MS Fragger that is not limited to data from a particular MS instrument can be easily incorporated into most of the existing data analysis pipelines [186].

7.2. The Bad

A problem that will always persist in the proteomics community is false-negative identification of peptides, i.e., not detecting a peptide and a protein that contains it, for a variety of reasons. One such reason is physiological/pathological. For example, if a peptide should be identified in a sample, but it is physiologically/pathologically post-translationally modified (PTMs, i.e., phosphorylation, acetylation etc.), then the peptide is not found in any database search. While most researchers do search for common PTMs in peptides such as phosphorylation, they do not necessarily search for rare PTMs such as farnesylation, and are unable to search for unknown PTMs.

While this problem could be solved (i.e., by expanding the database search using even more powerful software), the price is relatively high and each search will require a considerable amount of time. Another factor for false-negative identification of peptides is experimental. For example, if a peptide that should be identified is modified during the sample preparation process by artificial modifications/PTMs such as methionine oxidation, tryptophan oxidation, cysteine oxidation (to cysteic acid), iodoacetamide-based modification of cysteine (to carbamydomethyl cysteine) or cysteineless peptides, or acrylamide-based modification of cysteine-containing peptides (to propionamide), then these peptides will no longer be identified in a database search, unless all these artificial PTMs are selected in a database search.

Physicochemical properties of proteins and peptides are also important and are not always considered. Hydrophobic proteins and natural PTMs in proteins are two examples of this. For example, hydrophobic peptides do not ionize well and, therefore, are not even found in the MS data. PTMs such as glycosylation and disulfide bridges can prevent trypsin cleavage and thus lead to false negative identification of peptides. An example of a peptide with a series of disulfide bridges, in which analysis of the precursor ions works well, but the disulfide bridges prevent a good fragmentation of these precursor ions, and thus leads to poor quality MS/MS spectra, was recently published from our lab [188].

Incomplete enzymatic digestions (usually trypsin) can lead to erroneous proteomics results and false negative identifications of some peptides. One such case is when one or more trypsin cleavage sites are missed. While 1–3 missed cleavages can easily be fixed (i.e., by database search with missed cleavages), peptides larger than 4–5 kDa that are produced by incomplete trypsin digestion are usually not identified, unless a specialized database is used. Furthermore, fragmentation of such large precursor ions does not produce good MS/MS spectra and therefore does not lead to the positive identification of any proteins [132].

Perhaps one of the biggest problems in a proteomics experiment is the unassigned MS/MS spectra (unpublished observations and [132]). Indeed, in a proteomics experiment, about 60–70% of the spectra are not assigned with confidence to any peptide (unpublished observations). This is true for experiments analyzing just one protein as well as analysis of a large set of proteins (unpublished observations). Again, multiple database searches that are sometimes customized can reduce the number of unassigned peptides to perhaps 50% or as low as 40% in the best-case scenarios, but there is still a lot of work that needs to be done in the area of bioinformatics and database searches.

7.3. The Future

False positive results, false negative results, and unassigned spectra can be solved only by advances in database search and superior bioinformatics analysis of proteomes, i.e., additional software that allows identification of additional peptides that correspond to previously unassigned MS/MS spectra.

8. Conclusions

MS and proteomic methods have greatly advanced the field of protein analysis for both fundamental and clinical research. There are well-established methods for sample preparation and great instrumentation (both UPLC and MS) for MS analysis. There is also a diverse variety of database search engines. However, under the assumption that a proteomics experiment is done properly, the greatest challenge that we still face is a full, comprehensive database search of a proteomics dataset. It is clear that what proteomics needs most is a better, more powerful bioinformatics support for database searching.

Funding: This research received no external funding.

Acknowledgments: We thank the past and current members of the Biochemistry & Proteomics Group for the pleasant working environment.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Schwenk, J.M.; Omenn, G.S.; Sun, Z.; Campbell, D.S.; Baker, M.S.; Overall, C.M.; Aebersold, R.; Moritz, R.L.; Deutsch, E.W. The Human Plasma Proteome Draft of 2017: Building on the Human Plasma PeptideAtlas from Mass Spectrometry and Complementary Assays. *J. Proteome Res.* **2017**, *16*, 4299–4310. [CrossRef] [PubMed]
2. Chen, D.; Li, X.Y.; Zhao, X.; Qin, Y.S.; Zhang, X.X.; Li, J.; Wang, J.M.; Wang, C.F. Proteomics and microstructure profiling of goat milk protein after homogenization. *J. Dairy Sci.* **2019**, *102*, 3839–3850. [CrossRef] [PubMed]
3. Mohanta, T.K.; Khan, A.; Hashem, A.; Abd Allah, E.F.; Al-Harrasi, A. The molecular mass and isoelectric point of plant proteomes. *BMC Genom.* **2019**, *20*, 631. [CrossRef] [PubMed]
4. Aslam, B.; Basit, M.; Nisar, M.A.; Khurshid, M.; Rasool, M.H. Proteomics: Technologies and Their Applications. *J. Chromatogr. Sci.* **2017**, *55*, 182–196. [CrossRef]
5. Swinbanks, D. Australia Backs Innovation, Shuns Telescope. *Nature* **1995**, *378*, 653. [CrossRef]
6. Corthals, G.L.; Wasinger, V.C.; Hochstrasser, D.F.; Sanchez, J.C. The dynamic range of protein expression: A challenge for proteomic research. *Electrophoresis* **2000**, *21*, 1104–1115. [CrossRef]
7. Anderson, N.G.; Matheson, A.; Anderson, N.L. Back to the future: The human protein index (HPI) and the agenda for post-proteomic biology. *Proteomics* **2001**, *1*, 3–12. [CrossRef]
8. Sperling, K. From proteomics to genomics. *Electrophoresis* **2001**, *22*, 2835–2837. [CrossRef]
9. Loo, J.A.; DeJohn, D.E.; Du, P.; Stevenson, T.I.; Loo, R.R.O. Application of mass spectrometry for target identification and characterization. *Med. Res. Rev.* **1999**, *19*, 307–319. [CrossRef]
10. Aicher, L.; Wahl, D.; Arce, A.; Grenet, O.; Steiner, S. New insights into cyclosporine A nephrotoxicity by proteome analysis. *Electrophoresis* **1998**, *19*, 1998–2003. [CrossRef]
11. Zhang, N.; Aebersold, R.; Schwilkowski, B. ProbID: A probabilistic algorithm to identify peptides through sequence database searching using tandem mass spectral data. *Proteomics* **2002**, *2*, 1406–1412. [CrossRef]

12. Tsai, P.L.; Chen, S.F. A Brief Review of Bioinformatics Tools for Glycosylation Analysis by Mass Spectrometry. *Mass Spectrom* **2017**, *6*, S0064. [CrossRef] [PubMed]
13. Nefedov, A.V.; Gilski, M.J.; Sadygov, R.G. Bioinformatics Tools for Mass Spectrometry-Based High-Throughput Quantitative Proteomics Platforms. *Curr. Proteom.* **2011**, *8*, 125–137. [CrossRef]
14. Woods, A.G.; Sokolowska, I.; Ngounou Wetie, A.G.; Channaveerappa, D.; Dupree, E.J.; Jayathirtha, M.; Aslebagh, R.; Wormwood, K.L.; Darie, C.C. Mass Spectrometry for Proteomics-Based Investigation. *Adv. Exp. Med. Biol.* **2019**, *1140*, 1–26. [PubMed]
15. Schirle, M.; Bantscheff, M.; Kuster, B. Mass spectrometry-based proteomics in preclinical drug discovery. *Chem. Biol.* **2012**, *19*, 72–84. [CrossRef]
16. Crutchfield, C.A.; Thomas, S.N.; Sokoll, L.J.; Chan, D.W. Advances in mass spectrometry-based clinical biomarker discovery. *Clin. Proteom.* **2016**, *13*, 1. [CrossRef]
17. Tong, M.; Liu, H.; Hao, J.; Fan, D. Comparative pharmacoproteomics reveals potential targets for berberine, a promising therapy for colorectal cancer. *Biochem. Biophys. Res. Commun.* **2020**, *525*, 244–250. [CrossRef]
18. Parsons, J.; Francavilla, C. 'Omics Approaches to Explore the Breast Cancer Landscape. *Front Cell Dev. Biol.* **2019**, *7*, 395. [CrossRef]
19. Saleem, S.; Tariq, S.; Aleem, I.; Sadr-Ul, S.; Tahseen, M.; Atiq, A.; Hassan, S.; Abu Bakar, M.; Khattak, S.; Syed, A.A.; et al. Proteomics analysis of colon cancer progression. *Clin. Proteom.* **2019**, *16*, 44. [CrossRef]
20. Rogers, J.C.; Bomgarden, R.D. Sample preparation for mass spectrometry-based proteomics; from proteomes to peptides. In *Advances in Experimental Medicine and Biology*; Springer New York LLC: New York, NY, USA, 2016; Volume 919, pp. 43–62.
21. Cristobal, A.; Marino, F.; Post, H.; Van Den Toorn, H.W.P.; Mohammed, S.; Heck, A.J.R. Toward an Optimized Workflow for Middle-Down Proteomics. *Anal. Chem.* **2017**, *89*, 3318–3325. [CrossRef]
22. Kim, M.S.; Pinto, S.M.; Getnet, D.; Nirujogi, R.S.; Manda, S.S.; Chaerkady, R.; Madugundu, A.K.; Kelkar, D.S.; Isserlin, R.; Jain, S.; et al. A draft map of the human proteome. *Nature* **2014**, *509*, 575–581. [CrossRef] [PubMed]
23. Wilhelm, M.; Schlegl, J.; Hahne, H.; Gholami, A.M.; Lieberenz, M.; Savitski, M.M.; Ziegler, E.; Butzmann, L.; Gessulat, S.; Marx, H.; et al. Mass-spectrometry-based draft of the human proteome. *Nature* **2014**, *509*, 582–587. [CrossRef] [PubMed]
24. Addona, T.A.; Abbatiello, S.E.; Schilling, B.; Skates, S.J.; Mani, D.R.; Bunk, D.M.; Spiegelman, C.H.; Zimmerman, L.J.; Ham, A.J.L.; Keshishian, H.; et al. Multi-site assessment of the precision and reproducibility of multiple reaction monitoring-based measurements of proteins in plasma. *Nat. Biotechnol.* **2009**, *27*, 633–641. [CrossRef] [PubMed]
25. Gillet, L.C.; Leitner, A.; Aebersold, R. Mass Spectrometry Applied to Bottom-Up Proteomics: Entering the High-Throughput Era for Hypothesis Testing. *Annu. Rev. Anal. Chem.* **2016**, *9*, 449–472. [CrossRef]
26. Fenyö, D.; Beavis, R.C. The GPMDB REST interface. *Bioinformatics* **2015**, *31*, 2056–2058. [CrossRef]
27. Kiser, J.Z.; Post, M.; Wang, B.; Miyagi, M. Streptomyces erythraeus trypsin for proteomics applications. *J. Proteome Res.* **2009**, *8*, 1810–1817. [CrossRef]
28. Tsiatsiani, L.; Heck, A.J.R. Proteomics beyond trypsin. *FEBS J.* **2015**, *282*, 2612–2626. [CrossRef]
29. Aebersold, R.; Mann, M. Mass spectrometry-based proteomics. *Nature* **2003**, *422*, 198–207. [CrossRef]
30. Graves, P.R.; Haystead, T.A.J. Molecular Biologist's Guide to Proteomics. *Microbiol. Mol. Biol. Rev.* **2002**, *66*, 39–63. [CrossRef]
31. Lovrić, J. *Introducing Proteomics: From Concepts to Sample Separation, Mass Spectrometry and Data Analysis*; Wiley-Blackwell: Hoboken, NJ, USA, 2011; p. 283.
32. Futcher, B.; Latter, G.I.; Monardo, P.; McLaughlin, C.S.; Garrels, J.I. A Sampling of the Yeast Proteome. *Mol. Cell. Biol.* **1999**, *19*, 7357–7368. [CrossRef]
33. Kim, H.; Dan, K.; Shin, H.; Lee, J.; Wang, J.I.; Han, D. An efficient method for high-pH peptide fractionation based on C18 StageTips for in-depth proteome profiling. *Anal. Methods* **2019**, *11*, 4693–4698. [CrossRef]
34. Feist, P.; Hummon, A.B. Proteomic challenges: Sample preparation techniques for Microgram-Quantity protein analysis from biological samples. *Int. J. Mol. Sci.* **2015**, *16*, 3537–3563. [CrossRef]
35. Bandow, J.E. Comparison of protein enrichment strategies for proteome analysis of plasma. *Proteomics* **2010**, *10*, 1416–1425. [CrossRef] [PubMed]
36. Polaskova, V.; Kapur, A.; Khan, A.; Molloy, M.P.; Baker, M.S. High-abundance protein depletion: Comparison of methods for human plasma biomarker discovery. *Electrophoresis* **2010**, *31*, 471–482. [CrossRef] [PubMed]

37. Deeb, S.J.; Cox, J.; Schmidt-Supprian, M.; Mann, M. N-linked glycosylation enrichment for in-depth cell surface proteomics of diffuse large B-cell lymphoma subtypes. *Mol. Cell. Proteom.* **2014**, *13*, 240–251. [CrossRef] [PubMed]
38. Nilsson, C.L.; Dillon, R.; Devakumar, A.; Shi, S.D.H.; Greig, M.; Rogers, J.C.; Krastins, B.; Rosenblatt, M.; Kilmer, G.; Major, M.; et al. Quantitative phosphoproteomic analysis of the STAT3/IL-6/HIF1α signaling network: An initial study in GSC11 glioblastoma stem cells. *J. Proteome Res.* **2010**, *9*, 430–443. [CrossRef] [PubMed]
39. Adachi, J.; Kishida, M.; Watanabe, S.; Hashimoto, Y.; Fukamizu, K.; Tomonaga, T. Proteome-wide discovery of unknown ATP-binding proteins and kinase inhibitor target proteins using an ATP probe. *J. Proteome Res.* **2014**, *13*, 5461–5470. [CrossRef]
40. Lemeer, S.; Zörgiebel, C.; Ruprecht, B.; Kohl, K.; Kuster, B. Comparing immobilized kinase inhibitors and covalent ATP probes for proteomic profiling of kinase expression and drug selectivity. *J. Proteome Res.* **2013**, *12*, 1723–1731. [CrossRef]
41. Patricelli, M.P.; Szardenings, A.K.; Liyanage, M.; Nomanbhoy, T.K.; Wu, M.; Weissig, H.; Aban, A.; Chun, D.; Tanner, S.; Kozarich, J.W. Functional interrogation of the kinome using nucleotide acyl phosphates. *Biochemistry* **2007**, *46*, 350–358. [CrossRef]
42. Kim, B.; Araujo, R.; Howard, M.; Magni, R.; Liotta, L.A.; Luchini, A. Affinity enrichment for mass spectrometry: Improving the yield of low abundance biomarkers. *Expert Rev. Proteom.* **2018**, *15*, 353–366. [CrossRef]
43. Ten Have, S.; Boulon, S.; Ahmad, Y.; Lamond, A.I. Mass spectrometry-based immuno-precipitation proteomics—The user's guide. *Proteomics* **2011**, *11*, 1153–1159. [CrossRef] [PubMed]
44. Gundry, R.L.; White, M.Y.; Murray, C.I.; Kane, L.A.; Fu, Q.; Stanley, B.A.; Van Eyk, J.E. Preparation of proteins and peptides for mass spectrometry analysis in a bottom-up proteomics workflow. *Curr. Protoc. Mol. Biol.* **2009**, *90*, 10–25.
45. Tubaon, R.M.; Haddad, P.R.; Quirino, J.P. Sample Clean-up Strategies for ESI Mass Spectrometry Applications in Bottom-up Proteomics: Trends from 2012 to 2016. *Proteomics* **2017**, *17*, 1700011. [CrossRef] [PubMed]
46. Curreem, S.O.T.; Watt, R.M.; Lau, S.K.P.; Woo, P.C.Y. Two-dimensional gel electrophoresis in bacterial proteomics. *Protein Cell* **2012**, *3*, 346–363. [CrossRef] [PubMed]
47. Kendrick, N.; Darie, C.C.; Hoelter, M.; Powers, G.; Johansen, J. 2D SDS PAGE in Combination with Western Blotting and Mass Spectrometry Is a Robust Method for Protein Analysis with Many Applications. *Adv. Exp. Med. Biol.* **2019**, *1140*, 563–574.
48. Anguraj Vadivel, A.K. Gel-based proteomics in plants: Time to move on from the tradition. *Front. Plant Sci.* **2015**, *6*, 1–4. [CrossRef]
49. Rogowska-Wrzesinska, A.; Le Bihan, M.C.; Thaysen-Andersen, M.; Roepstorff, P. 2D gels still have a niche in proteomics. *J. Proteom.* **2013**, *88*, 4–13. [CrossRef]
50. Loo, R.R.O.; Dales, N.; Andrews, P.C. Surfactant effects on protein structure examined by electrospray ionization mass spectrometry. *Protein Sci.* **1994**, *3*, 1975–1983. [CrossRef]
51. Loo, R.R.; Dales, N.; Andrews, P.C. The effect of detergents on proteins analyzed by electrospray ionization. *Methods Mol. Biol. (Cliftonn. J.)* **1996**, *61*, 141–160.
52. Keller, B.O.; Sui, J.; Young, A.B.; Whittal, R.M. Interferences and contaminants encountered in modern mass spectrometry. *Anal. Chim. Acta* **2008**, *627*, 71–81. [CrossRef]
53. Yocum, A.K.; Yu, K.; Oe, T.; Blair, I.A. Effect of immunoaffinity depletion of human serum during proteomic investigations. *J. Proteome Res.* **2005**, *4*, 1722–1731. [CrossRef] [PubMed]
54. Xu, B.; Zhang, Y.; Zhao, Z.; Yoshida, Y.; Magdeldin, S.; Fujinaka, H.; Ismail, T.A.; Yaoita, E.; Yamamoto, T. Usage of electrostatic eliminator reduces human keratin contamination significantly in gel-based proteomics analysis. *J. Proteom.* **2011**, *74*, 1022–1029. [CrossRef] [PubMed]
55. Hodge, K.; Have, S.T.; Hutton, L.; Lamond, A.I. Cleaning up the masses: Exclusion lists to reduce contamination with HPLC-MS/MS. *J. Proteom.* **2013**, *88*, 92–103. [CrossRef] [PubMed]
56. Lyngholm, M.; Vorum, H.; Nielsen, K.; Ehlers, N.; Honoré, B. Attempting to distinguish between endogenous and contaminating cytokeratins in a corneal proteomic study. *BMC Ophthalmol.* **2011**, *11*, 3. [CrossRef]
57. Rabilloud, T. Two-dimensional gel electrophoresis in proteomics: Old, old fashioned, but it still climbs up the mountains. *Proteomics* **2002**, *2*, 3–10. [CrossRef]

58. Wilkins, M.R.; Gasteiger, E.; Sanchez, J.C.; Bairoch, A.; Hochstrasser, D.F. Two-dimensional gel electrophoresis for proteome projects: The effects of protein hydrophobicity and copy number. *Electrophoresis* **1998**, *19*, 1501–1505. [CrossRef] [PubMed]
59. Doellinger, J.; Schneider, A.; Hoeller, M.; Lasch, P. Sample Preparation by Easy Extraction and Digestion (SPEED)—A Universal, Rapid, and Detergent-free Protocol for Proteomics Based on Acid Extraction. *Mol. Cell. Proteom.* **2020**, *19*, 209–222. [CrossRef] [PubMed]
60. Ye, X.T.; Tang, J.; Mao, Y.H.; Lu, X.; Yang, Y.; Chen, W.D.; Zhang, X.Y.; Xu, R.L.; Tian, R.J. Integrated proteomics sample preparation and fractionation: Method development and applications. *Trac-Trends Anal. Chem.* **2019**, *120*, 115667. [CrossRef]
61. Wiśniewski, J.R. Filter-Aided Sample Preparation: The Versatile and Efficient Method for Proteomic Analysis. In *Methods in Enzymology*; Academic Press Inc.: Cambridge, MA, USA, 2017; pp. 15–27.
62. Wiśniewski, J.R. Filter-aided sample preparation for proteome analysis. In *Methods in Molecular Biology*; Humana Press Inc.: Totowa, NJ, USA, 2018; pp. 3–10.
63. Kulak, N.A.; Pichler, G.; Paron, I.; Nagaraj, N.; Mann, M. Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. *Nat. Methods* **2014**, *11*, 319. [CrossRef]
64. Zhu, Y.; Piehowski, P.D.; Zhao, R.; Chen, J.; Shen, Y.F.; Moore, R.J.; Shukla, A.K.; Petyuk, V.A.; Campbell-Thompson, M.; Mathews, C.E.; et al. Nanodroplet processing platform for deep and quantitative proteome profiling of 10–100 mammalian cells. *Nat. Commun.* **2018**, *9*, 1–10. [CrossRef]
65. Yuan, H.M.; Zhang, S.; Zhao, B.F.; Weng, Y.J.; Zhu, X.D.; Li, S.W.; Zhang, L.H.; Zhang, Y.K. Enzymatic Reactor with Trypsin Immobilized on Graphene Oxide Modified Polymer Microspheres To Achieve Automated Proteome Quantification. *Anal. Chem.* **2017**, *89*, 6324–6329. [CrossRef] [PubMed]
66. Zougman, A.; Selby, P.J.; Banks, R.E. Suspension trapping (STrap) sample preparation method for bottom-up proteomics analysis. *Proteomics* **2014**, *14*, 1006–1010. [CrossRef]
67. Hughes, C.S.; Foehr, S.; Garfield, D.A.; Furlong, E.E.; Steinmetz, L.M.; Krijgsveld, J. Ultrasensitive proteome analysis using paramagnetic bead technology. *Mol. Syst. Biol.* **2014**, *10*, 757. [CrossRef]
68. Muller, T.; Kalxdorf, M.; Longuespee, R.; Kazdal, D.N.; Stenzinger, A.; Krijgsveld, J. Automated sample preparation with SP3 for low-input clinical proteomics. *Mol. Syst. Biol.* **2020**, *16*, e9111. [CrossRef] [PubMed]
69. Wolters, D.A.; Washburn, M.P.; Yates, J.R., 3rd. An automated multidimensional protein identification technology for shotgun proteomics. *Anal. Chem.* **2001**, *73*, 5683–5690. [CrossRef] [PubMed]
70. Horning, E.C.; Carroll, D.I.; Dzidic, I.; Haegele, K.D.; Horning, M.G.; Stillwell, R.N. Atmospheric pressure ionization (API) mass spectrometry. Solvent-mediated ionization of samples introduced in solution and in a liquid chromatograph effluent stream. *J. Chromatogr. Sci.* **1974**, *12*, 725–729. [CrossRef]
71. Blakley, C.R.; Carmody, J.J.; Vestal, M.L. A New Soft Ionization Technique for Mass-Spectrometry of Complex-Molecules. *J. Am. Chem. Soc.* **1980**, *102*, 5931–5933. [CrossRef]
72. Blakley, C.R.; Carmody, J.C.; Vestal, M.L. Combined Liquid Chromatograph-Mass Spectrometer for Involatile Biological Samples. *Clin. Chem.* **1980**, *26*, 1467–1473. [CrossRef]
73. Anderson, N.L.; Anderson, N.G. The human plasma proteome—History, character, and diagnostic prospects. *Mol. Cell. Proteom.* **2002**, *1*, 845–867. [CrossRef]
74. Mant, C.T.; Chen, Y.; Yan, Z.; Popa, T.V.; Kovacs, J.M.; Mills, J.B.; Tripet, B.P.; Hodges, R.S. HPLC analysis and purification of peptides. *Methods Mol. Biol.* **2007**, *386*, 3–55.
75. Neverova, I.; Van Eyk, J.E. Application of reversed phase high performance liquid chromatography for subproteomic analysis of cardiac muscle. *Proteomics* **2002**, *2*, 22–31. [CrossRef]
76. Chait, B.T. Chemistry. Mass spectrometry: Bottom-up or top-down? *Science* **2006**, *314*, 65–66. [CrossRef] [PubMed]
77. Duong-Ly, K.C.; Gabelli, S.B. Gel filtration chromatography (size exclusion chromatography) of proteins. *Methods Enzym.* **2014**, *541*, 105–114.
78. Li, Y.; Gu, C.; Gruenhagen, J.; Zhang, K.; Yehl, P.; Chetwyn, N.P.; Medley, C.D. A size exclusion-reversed phase two dimensional-liquid chromatography methodology for stability and small molecule related species in antibody drug conjugates. *J. Chromatogr. A* **2015**, *1393*, 81–88. [CrossRef]
79. Duong-Ly, K.C.; Gabelli, S.B. Using ion exchange chromatography to purify a recombinantly expressed protein. *Methods Enzym.* **2014**, *541*, 95–103.

80. Cai, W.; Tucholski, T.; Chen, B.; Alpert, A.J.; McIlwain, S.; Kohmoto, T.; Jin, S.; Ge, Y. Top-Down Proteomics of Large Proteins up to 223 kDa Enabled by Serial Size Exclusion Chromatography Strategy. *Anal. Chem.* **2017**, *89*, 5467–5475. [CrossRef]
81. Freeman, W.M.; Lull, M.E.; Guilford, M.T.; Vrana, K.E. Depletion of abundant proteins from non-human primate serum for biomarker studies. *Proteomics* **2006**, *6*, 3109–3113. [CrossRef]
82. Kline, K.G.; Wu, C.C. MudPIT analysis: Application to human heart tissue. *Methods Mol. Biol.* **2009**, *528*, 281–293.
83. Motoyama, A.; Venable, J.D.; Ruse, C.I.; Yates, J.R., 3rd. Automated ultra-high-pressure multidimensional protein identification technology (UHP-MudPIT) for improved peptide identification of proteomic samples. *Anal. Chem.* **2006**, *78*, 5109–5118. [CrossRef]
84. Issaq, H.J.; Chan, K.C.; Janini, G.M.; Conrads, T.P.; Veenstra, T.D. Multidimensional separation of peptides for effective proteomic analysis. *.J Chromatogr. B Anal. Technol. Biomed. Life Sci.* **2005**, *817*, 35–47. [CrossRef]
85. Washburn, M.P.; Wolters, D.; Yates, J.R., 3rd. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat. Biotechnol.* **2001**, *19*, 242–247. [CrossRef] [PubMed]
86. Lambert, J.P.; Ethier, M.; Smith, J.C.; Figeys, D. Proteomics: From gel based to gel free. *Anal. Chem.* **2005**, *77*, 3771–3787. [CrossRef] [PubMed]
87. Yu, D.; Wang, Z.; Cupp-Sutton, K.A.; Liu, X.; Wu, S. Deep Intact Proteoform Characterization in Human Cell Lysate Using High-pH and Low-pH Reversed-Phase Liquid Chromatography. *J. Am. Soc. Mass Spectrom.* **2019**, *30*, 2502–2513. [CrossRef] [PubMed]
88. Aguilar, M.I. Reversed-phase high-performance liquid chromatography. *Methods Mol. Biol.* **2004**, *251*, 9–22. [PubMed]
89. Solovyeva, E.M.; Lobas, A.A.; Kopylov, A.T.; Ilina, I.Y.; Levitsky, L.I.; Moshkovskii, S.A.; Gorshkov, M.V. FractionOptimizer: A method for optimal peptide fractionation in bottom-up proteomics. *Anal. Bioanal. Chem.* **2018**, *410*, 3827–3833. [CrossRef]
90. Nogueira, R.; Lammerhofer, M.; Lindner, W. Alternative high-performance liquid chromatographic peptide separation and purification concept using a new mixed-mode reversed-phase/weak anion-exchange type stationary phase. *J. Chromatogr. A* **2005**, *1089*, 158–169. [CrossRef]
91. Phillips, H.L.; Williamson, J.C.; van Elburg, K.A.; Snijders, A.P.; Wright, P.C.; Dickman, M.J. Shotgun proteome analysis utilising mixed mode (reversed phase-anion exchange chromatography) in conjunction with reversed phase liquid chromatography mass spectrometry analysis. *Proteomics* **2010**, *10*, 2950–2960. [CrossRef]
92. Vasconcelos Soares Maciel, E.; de Toffoli, A.L.; Sobieski, E.; Domingues Nazario, C.E.; Lancas, F.M. Miniaturized liquid chromatography focusing on analytical columns and mass spectrometry: A review. *Anal. Chim. Acta* **2020**, *1103*, 11–31. [CrossRef]
93. Zuvela, P.; Skoczylas, M.; Jay Liu, J.; Ba Czek, T.; Kaliszan, R.; Wong, M.W.; Buszewski, B. Column Characterization and Selection Systems in Reversed-Phase High-Performance Liquid Chromatography. *Chem. Rev.* **2019**, *119*, 3674–3729. [CrossRef]
94. Maya, F.; Paull, B. Recent strategies to enhance the performance of polymer monoliths for analytical separations. *J. Sep. Sci.* **2019**, *42*, 1564–1576. [CrossRef]
95. Luo, Q.; Shen, Y.; Hixson, K.K.; Zhao, R.; Yang, F.; Moore, R.J.; Mottaz, H.M.; Smith, R.D. Preparation of 20-microm-i.d. silica-based monolithic columns and their performance for proteomics analyses. *Anal. Chem.* **2005**, *77*, 5028–5035. [CrossRef]
96. Novotny, M.V. Capillary biomolecular separations. *J. Chromatogr. B Biomed. Sci. Appl.* **1997**, *689*, 55–70. [CrossRef]
97. Premstaller, A.; Oberacher, H.; Walcher, W.; Timperio, A.M.; Zolla, L.; Chervet, J.P.; Cavusoglu, N.; van Dorsselaer, A.; Huber, C.G. High-performance liquid chromatography-electrospray ionization mass spectrometry using monolithic capillary columns for proteomic studies. *Anal. Chem.* **2001**, *73*, 2390–2396. [CrossRef] [PubMed]
98. He, B.; Tait, N.; Regnier, F. Fabrication of nanocolumns for liquid chromatography. *Anal. Chem.* **1998**, *70*, 3790–3797. [CrossRef]
99. Ion, L.; Petre, B.A. Immuno-Affinity Mass Spectrometry: A Novel Approaches with Biomedical Relevance. *Adv. Exp. Med. Biol.* **2019**, *1140*, 377–388. [PubMed]

100. Yuan, X.; Oleschuk, R.D. Advances in Microchip Liquid Chromatography. *Anal Chem* **2018**, *90*, 283–301. [CrossRef] [PubMed]
101. Andjelkovic, U.; Tufegdzic, S.; Popovic, M. Use of monolithic supports for high-throughput protein and peptide separation in proteomics. *Electrophoresis* **2017**, *38*, 2851–2869. [CrossRef]
102. Thakur, S.S.; Geiger, T.; Chatterjee, B.; Bandilla, P.; Frohlich, F.; Cox, J.; Mann, M. Deep and highly sensitive proteome coverage by LC-MS/MS without prefractionation. *Mol. Cell. Proteom.* **2011**, *10*, M110.003699. [CrossRef]
103. Van Eeckhaut, A.; Lanckmans, K.; Sarre, S.; Smolders, I.; Michotte, Y. Validation of bioanalytical LC-MS/MS assays: Evaluation of matrix effects. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* **2009**, *877*, 2198–2207. [CrossRef]
104. MacNair, J.E.; Lewis, K.C.; Jorgenson, J.W. Ultrahigh-pressure reversed-phase liquid chromatography in packed capillary columns. *Anal. Chem.* **1997**, *69*, 983–989. [CrossRef]
105. Denoroy, L.; Zimmer, L.; Renaud, B.; Parrot, S. Ultra high performance liquid chromatography as a tool for the discovery and the analysis of biomarkers of diseases: A review. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* **2013**, *927*, 37–53. [CrossRef] [PubMed]
106. Wilson, I.D.; Nicholson, J.K.; Castro-Perez, J.; Granger, J.H.; Johnson, K.A.; Smith, B.W.; Plumb, R.S. High resolution "ultra performance" liquid chromatography coupled to oa-TOF mass spectrometry as a tool for differential metabolic pathway profiling in functional genomic studies. *J. Proteome Res.* **2005**, *4*, 591–598. [CrossRef] [PubMed]
107. Neue, U.D.; Kele, M.; Bunner, B.; Kromidas, A.; Dourdeville, T.; Mazzeo, J.R.; Grumbach, E.S.; Serpa, S.; Wheat, T.E.; Hong, P.; et al. Ultra-performance liquid chromatography technology and applications. *Adv. Chromatogr.* **2010**, *48*, 99–143.
108. Vuckovic, D.; Dagley, L.F.; Purcell, A.W.; Emili, A. Membrane proteomics by high performance liquid chromatography–tandem mass spectrometry: Analytical approaches and challenges. *Proteomics* **2013**, *13*, 404–423. [CrossRef] [PubMed]
109. Fairchild, J.N.; Horvath, K.; Guiochon, G. Theoretical advantages and drawbacks of on-line, multidimensional liquid chromatography using multiple columns operated in parallel. *J. Chromatogr. A* **2009**, *1216*, 6210–6217. [CrossRef] [PubMed]
110. Ball, C.H.; Roulhac, P.L. Multidimensional Techniques in Protein Separations for Neuroproteomics. In *Neuroproteomics*; Alzate, O., Ed.; CRC Press: Boca Raton, FL, USA, 2010.
111. Kota, U.; Stolowitz, M.L. Improving Proteome Coverage by Reducing Sample Complexity via Chromatography. *Adv. Exp. Med. Biol.* **2016**, *919*, 83–143.
112. Kocova Vlckova, H.; Pilarova, V.; Svobodova, P.; Plisek, J.; Svec, F.; Novakova, L. Current state of bioanalytical chromatography in clinical analysis. *Analyst* **2018**, *143*, 1305–1325. [CrossRef]
113. Ucles Moreno, A.; Herrera Lopez, S.; Reichert, B.; Lozano Fernandez, A.; Hernando Guil, M.D.; Fernandez-Alba, A.R. Microflow liquid chromatography coupled to mass spectrometry–an approach to significantly increase sensitivity, decrease matrix effects, and reduce organic solvent usage in pesticide residue analysis. *Anal. Chem.* **2015**, *87*, 1018–1025. [CrossRef]
114. Spicer, V.; Krokhin, O.V. Peptide retention time prediction in hydrophilic interaction liquid chromatography. Comparison of separation selectivity between bare silica and bonded stationary phases. *J. Chromatogr. A* **2018**, *1534*, 75–84. [CrossRef]
115. 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. [CrossRef]
116. Strathmann, F.G.; Hoofnagle, A.N. Current and Future Applications of Mass Spectrometry to the Clinical Laboratory. *Am. J. Clin. Pathol.* **2011**, *136*, 609–616. [CrossRef] [PubMed]
117. Chalmers, M.J.; Gaskell, S.J. Advances in mass spectrometry for proteome analysis. *Curr. Opin. Biotechnol.* **2000**, *11*, 384–390. [CrossRef]
118. Bateman, N.W.; Goulding, S.P.; Shulman, N.J.; Gadok, A.K.; Szumlinski, K.K.; MacCoss, M.J.; Wu, C.C. Maximizing peptide identification events in proteomic workflows using data-dependent acquisition (DDA). *Mol. Cell. Proteom.* **2014**, *13*, 329–338. [CrossRef] [PubMed]
119. Neilson, K.A.; Ali, N.A.; Muralidharan, S.; Mirzaei, M.; Mariani, M.; Assadourian, G.; Lee, A.; Van Sluyter, S.C.; Haynes, P.A. Less label, more free: Approaches in label-free quantitative mass spectrometry. *Proteomics* **2011**, *11*, 535–553. [CrossRef]

120. Tabb, D.L.; Vega-Montoto, L.; Rudnick, P.A.; Variyath, A.M.; Ham, A.-J.L.; Bunk, D.M.; Kilpatrick, L.E.; Billheimer, D.D.; Blackman, R.K.; Cardasis, H.L. Repeatability and reproducibility in proteomic identifications by liquid chromatography-tandem mass spectrometry. *J. Proteome Res.* **2010**, *9*, 761–776. [CrossRef] [PubMed]
121. Doerr, A. DIA mass spectrometry. *Nat. Methods* **2014**, *12*, 35. [CrossRef]
122. Bilbao, A.; Varesio, E.; Luban, J.; Strambio-De-Castillia, C.; Hopfgartner, G.; Müller, M.; Lisacek, F. Processing strategies and software solutions for data-independent acquisition in mass spectrometry. *Proteomics* **2015**, *15*, 964–980. [CrossRef]
123. Lange, V.; Picotti, P.; Domon, B.; Aebersold, R. Selected reaction monitoring for quantitative proteomics: A tutorial. *Mol. Syst. Biol.* **2008**, *4*, 14. [CrossRef]
124. Maiolica, A.; Jünger, M.A.; Ezkurdia, I.; Aebersold, R. Targeted proteome investigation via selected reaction monitoring mass spectrometry. *J. Proteom.* **2012**, *75*, 3495–3513. [CrossRef]
125. Stone, P.; Glauner, T.; Kuhlmann, F.; Schlabach, T.; Miller, K. *New Dynamic MRM Mode Improves Data Quality and Triple Quad Quantification in Complex Analyses*; Agilent Publication: Santa Clara, CA, USA, 2009.
126. Rauniyar, N. Parallel reaction monitoring: A targeted experiment performed using high resolution and high mass accuracy mass spectrometry. *Int. J. Mol. Sci.* **2015**, *16*, 28566–28581. [CrossRef]
127. Ronsein, G.E.; Pamir, N.; von Haller, P.D.; Kim, D.S.; Oda, M.N.; Jarvik, G.P.; Vaisar, T.; Heinecke, J.W. Parallel reaction monitoring (PRM) and selected reaction monitoring (SRM) exhibit comparable linearity, dynamic range and precision for targeted quantitative HDL proteomics. *J. Proteom.* **2015**, *113*, 388–399. [CrossRef]
128. Gallien, S.; Duriez, E.; Crone, C.; Kellmann, M.; Moehring, T.; Domon, B. Targeted proteomic quantification on quadrupole-orbitrap mass spectrometer. *Mol. Cell. Proteom.* **2012**, *11*, 1709–1723. [CrossRef] [PubMed]
129. Michalski, A.; Damoc, E.; Hauschild, J.-P.; Lange, O.; Wieghaus, A.; Makarov, A.; Nagaraj, N.; Cox, J.; Mann, M.; Horning, S. Mass Spectrometry-based Proteomics Using Q Exactive, a High-performance Benchtop Quadrupole Orbitrap Mass Spectrometer. *Mol. Cell. Proteom.* **2011**, *10*, M111.011015. [CrossRef] [PubMed]
130. Bonner, R.; Hopfgartner, G. SWATH data independent acquisition mass spectrometry for metabolomics. *Trac. Trends Anal. Chem.* **2019**, *120*, 115278. [CrossRef]
131. Lenz, E.; Bright, J.; Knight, R.; Westwood, F.; Davies, D.; Major, H.; Wilson, I. Metabonomics with 1H-NMR spectroscopy and liquid chromatography-mass spectrometry applied to the investigation of metabolic changes caused by gentamicin-induced nephrotoxicity in the rat. *Biomark* **2005**, *10*, 173–187. [CrossRef] [PubMed]
132. Channaveerappa, D.; Ngounou Wetie, A.G.; Darie, C.C. Bottlenecks in Proteomics: An Update. *Adv. Exp. Med. Biol.* **2019**, *1140*, 753–769.
133. Morris, C.B.; Poland, J.C.; May, J.C.; McLean, J.A. Fundamentals of Ion Mobility-Mass Spectrometry for the Analysis of Biomolecules. *Methods Mol. Biol.* **2020**, *2084*, 1–31.
134. Distler, U.; Kuharev, J.; Navarro, P.; Tenzer, S. Label-free quantification in ion mobility-enhanced data-independent acquisition proteomics. *Nat. Protoc.* **2016**, *11*, 795–812. [CrossRef]
135. Rifai, N.; Gillette, M.A.; Carr, S.A. Protein biomarker discovery and validation: The long and uncertain path to clinical utility. *Nat. Biotechnol.* **2006**, *24*, 971–983. [CrossRef]
136. Han, X.; Aslanian, A.; Yates, J.R., 3rd. Mass spectrometry for proteomics. *Curr. Opin. Chem. Biol.* **2008**, *12*, 483–490. [CrossRef]
137. Liu, Y.; Buil, A.; Collins, B.C.; Gillet, L.C.J.; Blum, L.C.; Cheng, L.-Y.; Vitek, O.; Mouritsen, J.; Lachance, G.; Spector, T.D.; et al. Quantitative variability of 342 plasma proteins in a human twin population. *Mol. Syst. Biol.* **2015**, *11*, 786. [CrossRef] [PubMed]
138. Contino, N.C.; Pierson, E.E.; Keifer, D.Z.; Jarrold, M.F. Charge Detection Mass Spectrometry with Resolved Charge States. *J. Am. Soc. Mass Spectrom.* **2013**, *24*, 101–108. [CrossRef] [PubMed]
139. Zubarev, R.A. Electron-capture dissociation tandem mass spectrometry. *Curr. Opin. Biotechnol.* **2004**, *15*, 12–16. [CrossRef]
140. Mikesh, L.M.; Ueberheide, B.; Chi, A.; Coon, J.J.; Syka, J.E.; Shabanowitz, J.; Hunt, D.F. The utility of ETD mass spectrometry in proteomic analysis. *Biochim. Biophys. Acta.* **2006**, *1764*, 1811–1822. [CrossRef] [PubMed]
141. Wu, G.Y.; Wan, X.; Xu, B.H. A new estimation of protein-level false discovery rate. *BMC Genom.* **2018**, *19*, 567. [CrossRef]
142. Eng, J.K.; McCormack, A.L.; Yates, J.R. 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. [CrossRef]

143. Sadygov, R.G. Using SEQUEST with Theoretically Complete Sequence Databases. *J. Am. Soc. Mass Spectrom.* **2015**, *26*, 1858–1864. [CrossRef]
144. Eng, J.K.; Hoopmann, M.R.; Jahan, T.A.; Egertson, J.D.; Noble, W.S.; MacCoss, M.J. A Deeper Look into Comet-Implementation and Features. *J. Am. Soc. Mass Spectrom.* **2015**, *26*, 1865–1874. [CrossRef]
145. Tyanova, S.; Temu, T.; Cox, J. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat. Protoc.* **2016**, *11*, 2301–2319. [CrossRef]
146. Webb-Robertson, B.J.M.; Cannon, W.R. Current trends in computational inference from mass spectrometry-based proteomics. *Brief. Bioinform.* **2007**, *8*, 304–317. [CrossRef]
147. Ma, B.; Zhang, K.Z.; Hendrie, C.; Liang, C.Z.; Li, M.; Doherty-Kirby, A.; Lajoie, G. PEAKS: Powerful software for peptide de novo sequencing by tandem mass spectrometry. *Rapid Commun. Mass Spectrom.* **2003**, *17*, 2337–2342. [CrossRef] [PubMed]
148. Taylor, J.A.; Johnson, R.S. Sequence database searches via de novo peptide sequencing by tandem mass spectrometry. *Rapid Commun. Mass Spectrom.* **1997**, *11*, 1067–1075. [CrossRef]
149. Frank, A.; Pevzner, P. PepNovo: De novo peptide sequencing via probabilistic network modeling. *Anal. Chem.* **2005**, *77*, 964–973. [CrossRef] [PubMed]
150. Kalb, S.R.; Baudys, J.; Rees, J.C.; Smith, T.J.; Smith, L.A.; Helma, C.H.; Hill, K.; Kull, S.; Kirchner, S.; Dorner, M.B.; et al. De novo subtype and strain identification of botulinum neurotoxin type B through toxin proteomics. *Anal. Bioanal. Chem.* **2012**, *403*, 215–226. [CrossRef]
151. Medzihradszky, K.F.; Chalkley, R.J. Lessons in De Novo Peptide Sequencing by Tandem Mass Spectrometry. *Mass Spectrom. Rev.* **2015**, *34*, 43–63. [CrossRef]
152. Ma, B. Novor: Real-Time Peptide de Novo Sequencing Software. *J. Am. Soc. Mass Spectrom.* **2015**, *26*, 1885–1894. [CrossRef] [PubMed]
153. Jeong, K.; Kim, S.; Pevzner, P.A. UniNovo: A universal tool for de novo peptide sequencing. *Bioinformatics* **2013**, *29*, 1953–1962. [CrossRef]
154. Keller, A.; Nesvizhskii, A.I.; Kolker, E.; Aebersold, R. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal. Chem.* **2002**, *74*, 5383–5392. [CrossRef] [PubMed]
155. Jaffe, J.D.; Berg, H.C.; Church, G.M. Proteogenomic mapping as a complementary method to perform genome annotation. *Proteomics* **2004**, *4*, 59–77. [CrossRef] [PubMed]
156. Fertig, E.J.; Slebos, R.; Chung, C.H. Application of genomic and proteomic technologies in biomarker discovery. *Am. Soc. Clin. Oncol. Educ. Book* **2012**, *1*, 377–382. [CrossRef]
157. Sheynkman, G.M.; Shortreed, M.R.; Cesnik, A.J.; Smith, L.M. Proteogenomics: Integrating Next-Generation Sequencing and Mass Spectrometry to Characterize Human Proteomic Variation. *Annu. Rev. Anal. Chem. (Palo Alto. Calif.)* **2016**, *9*, 521–545. [CrossRef] [PubMed]
158. Lai, Y.J.; Yu, D.; Zhang, J.H.; Chen, G.J. Cooperation of Genomic and Rapid Nongenomic Actions of Estrogens in Synaptic Plasticity. *Mol. Neurobiol.* **2017**, *54*, 4113–4126. [CrossRef] [PubMed]
159. Whiteaker, J.R.; Halusa, G.N.; Hoofnagle, A.N.; Sharma, V.; MacLean, B.; Yan, P.; Wrobel, J.A.; Kennedy, J.; Mani, D.R.; Zimmerman, L.J.; et al. CPTAC Assay Portal: A repository of targeted proteomic assays. *Nat. Methods* **2014**, *11*, 703–704. [CrossRef] [PubMed]
160. Dupree, E.J.; Crimmins, B.S.; Holsen, T.M.; Darie, C.C. Developing Well-Annotated Species-Specific Protein Databases Using Comparative Proteogenomics. *Adv. Exp. Med. Biol.* **2019**, *1140*, 389–400. [PubMed]
161. Dupree, E.J.; Crimmins, B.S.; Holsen, T.M.; Darie, C.C. Proteomic Analysis of the Lake Trout (Salvelinus namaycush) Liver Identifies Proteins from Evolutionarily Close and -Distant Fish Relatives. *Proteomics* **2019**, *19*, e1800429. [CrossRef]
162. Mihasan, M.; Babii, C.; Aslebagh, R.; Channaveerappa, D.; Dupree, E.; Darie, C.C. Proteomics based analysis of the nicotine catabolism in Paenarthrobacter nicotinovorans pAO1. *Sci. Rep.* **2018**, *8*, 16239. [CrossRef]
163. Aslebagh, R.; Channaveerappa, D.; Arcaro, K.F.; Darie, C.C. Proteomics analysis of human breast milk to assess breast cancer risk. *Electrophoresis* **2018**, *39*, 653–665. [CrossRef]
164. Channaveerappa, D.; Lux, J.C.; Wormwood, K.L.; Heintz, T.A.; McLerie, M.; Treat, J.A.; King, H.; Alnasser, D.; Goodrow, R.J.; Ballard, G.; et al. Atrial electrophysiological and molecular remodelling induced by obstructive sleep apnoea. *J. Cell. Mol. Med.* **2017**, *21*, 2223–2235. [CrossRef]
165. Ngounou Wetie, A.G.; Wormwood, K.L.; Russell, S.; Ryan, J.P.; Darie, C.C.; Woods, A.G. A Pilot Proteomic Analysis of Salivary Biomarkers in Autism Spectrum Disorder. *Autism. Res.* **2015**, *8*, 338–350. [CrossRef]

166. Leptos, K.C.; Sarracino, D.A.; Jaffe, J.D.; Krastins, B.; Church, G.M. MapQuant: Open-source software for large-scale protein quantification. *Proteomics* **2006**, *6*, 1770–1782. [CrossRef]
167. Pluskal, T.; Castillo, S.; Villar-Briones, A.; Oresic, M. MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinform.* **2010**, *11*, 395. [CrossRef] [PubMed]
168. Bellew, M.; Coram, M.; Fitzgibbon, M.; Igra, M.; Randolph, T.; Wang, P.; May, D.; Eng, J.; Fang, R.H.; Lin, C.W.; et al. A suite of algorithms for the comprehensive analysis of complex protein mixtures using high-resolution LC-MS. *Bioinformatics* **2006**, *22*, 1902–1909. [CrossRef] [PubMed]
169. Sturm, M.; Bertsch, A.; Gropl, C.; Hildebrandt, A.; Hussong, R.; Lange, E.; Pfeifer, N.; Schulz-Trieglaff, O.; Zerck, A.; Reinert, K.; et al. OpenMS-An open-source software framework for mass spectrometry. *BMC Bioinform.* **2008**, *9*, 1–11. [CrossRef] [PubMed]
170. Palagi, P.M.; Walther, D.; Quadroni, M.; Catherinet, S.; Burgess, J.; Zimmermann-Ivol, C.G.; Sanchez, J.C.; Binz, P.A.; Hochstrasser, D.F.; Appel, R.D. MSight: An image analysis software for liquid chromatography-mass spectrometry. *Proteomics* **2005**, *5*, 2381–2384. [CrossRef] [PubMed]
171. Mueller, L.N.; Rinner, O.; Schmidt, A.; Letarte, S.; Bodenmiller, B.; Brusniak, M.Y.; Vitek, O.; Aebersold, R.; Muller, M. SuperHirn—A novel tool for high resolution LC-MS-based peptide/protein profiling. *Proteomics* **2007**, *7*, 3470–3480. [CrossRef] [PubMed]
172. 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. Proteom.* **2014**, *13*, 2513–2526. [CrossRef] [PubMed]
173. Zhu, W.H.; Smith, J.W.; Huang, C.M. Mass Spectrometry-Based Label-Free Quantitative Proteomics. *J. Biomed. Biotechnol.* **2010**, *2010*, 840518. [CrossRef]
174. Kirkpatrick, D.S.; Gerber, S.A.; Gygi, S.P. The absolute quantification strategy: A general procedure for the quantification of proteins and post-translational modifications. *Methods* **2005**, *35*, 265–273. [CrossRef]
175. Nesvizhskii, A.I. Proteogenomics: Concepts, applications and computational strategies. *Nat. Methods* **2014**, *11*, 1114–1125. [CrossRef]
176. Matsuda, F.; Tsugawa, H.; Fukusaki, E. Method for Assessing the Statistical Significance of Mass Spectral Similarities Using Basic Local Alignment Search Tool Statistics. *Anal. Chem.* **2013**, *85*, 8291–8297. [CrossRef]
177. Old, W.M.; Meyer-Arendt, K.; Aveline-Wolf, L.; Pierce, K.G.; Mendoza, A.; Sevinsky, J.R.; Resing, K.A.; Ahn, N.G. Comparison of label-free methods for quantifying human proteins by shotgun proteomics. *Mol. Cell. Proteom.* **2005**, *4*, 1487–1502. [CrossRef] [PubMed]
178. Chen, X.L.; Wei, S.S.; Ji, Y.L.; Guo, X.J.; Yang, F.Q. Quantitative proteomics using SILAC: Principles, applications, and developments. *Proteomics* **2015**, *15*, 3175–3192. [CrossRef] [PubMed]
179. 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. [CrossRef] [PubMed]
180. Choi, H.; Nesvizhskii, A.I. Semisupervised model-based validation of peptide identifications in mass spectrometry-based proteomics. *J. Proteome Res.* **2008**, *7*, 254–265. [CrossRef]
181. Danilova, Y.; Voronkova, A.; Sulimov, P.; Kertesz-Farkas, A. Bias in False Discovery Rate Estimation in Mass-Spectrometry-Based Peptide Identification. *J. Proteome Res.* **2019**, *18*, 2354–2358. [CrossRef]
182. Kim, H.; Lee, S.; Park, H. Target-small decoy search strategy for false discovery rate estimation. *BMC Bioinform.* **2019**, *20*, 438. [CrossRef]
183. Gonnelli, G.; Stock, M.; Verwaeren, J.; Maddelein, D.; De Baets, B.; Martens, L.; Degroeve, S. A Decoy-Free Approach to the Identification of Peptides. *J. Proteome Res.* **2015**, *14*, 1792–1798. [CrossRef]
184. 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*. [CrossRef]
185. Skinner, O.S.; Kelleher, N.L. Illuminating the dark matter of shotgun proteomics. *Nat. Biotechnol.* **2015**, *33*, 717. [CrossRef]
186. Kong, A.T.; Leprevost, F.V.; Avtonomov, D.M.; Mellacheruvu, D.; Nesvizhskii, A.I. MSFragger: Ultrafast and comprehensive peptide identification in mass spectrometry–based proteomics. *Nat. Methods* **2017**, *14*, 513. [CrossRef]

187. Meyer, J.G. Fast Proteome Identification and Quantification from Data-Dependent Acquisition-Tandem Mass Spectrometry using Free Software Tools. *Methods Protoc.* **2019**, *2*, 8. [CrossRef] [PubMed]
188. Wormwood, K.L.; Ngounou Wetie, A.G.; Gomez, M.V.; Ju, Y.; Kowalski, P.; Mihasan, M.; Darie, C.C. Structural Characterization and Disulfide Assignment of Spider Peptide Phalpha1beta by Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2018**, *29*, 827. [CrossRef] [PubMed]