

Review

Proteomes Are of Proteoforms: Embracing the Complexity

Katrina Carbonara, Martin Andonovski and Jens R. Coorsen *

Faculties of Applied Health Sciences and Mathematics & Science, Departments of Health Sciences and Biological Sciences, Brock University, 1812 Sir Isaac Brock Way, St. Catharines, ON L2S 3A1, Canada; kcarbonara@brocku.ca (K.C.); mandonovski@brocku.ca (M.A.)

* Correspondence: jcoorsen@brocku.ca

Abstract: Proteomes are complex—much more so than genomes or transcriptomes. Thus, simplifying their analysis does not simplify the issue. Proteomes are of proteoforms, not canonical proteins. While having a catalogue of amino acid sequences provides invaluable information, this is the Proteome-lite. To dissect biological mechanisms and identify critical biomarkers/drug targets, we must assess the myriad of proteoforms that arise at any point before, after, and between translation and transcription (e.g., isoforms, splice variants, and post-translational modifications [PTM]), as well as newly defined species. There are numerous analytical methods currently used to address proteome depth and here we critically evaluate these in terms of the current ‘state-of-the-field’. We thus discuss both pros and cons of available approaches and where improvements or refinements are needed to quantitatively characterize proteomes. To enable a next-generation approach, we suggest that advances lie in transdisciplinarity via integration of current proteomic methods to yield a unified discipline that capitalizes on the strongest qualities of each. Such a necessary (if not revolutionary) shift cannot be accomplished by a continued primary focus on proteo-genomics/-transcriptomics. We must embrace the complexity. Yes, these are the hard questions, and this will not be easy...but where is the fun in easy?

Keywords: proteomics; top-down; bottom-up; immunoassay; mass spectrometry; two-dimensional gel electrophoresis; Western blotting

Citation: Carbonara, K.; Andonovski, M.; Coorsen, J.R. Proteomes Are of Proteoforms: Embracing the Complexity. *Proteomes* **2021**, *9*, 38. <https://doi.org/10.3390/proteomes9030038>

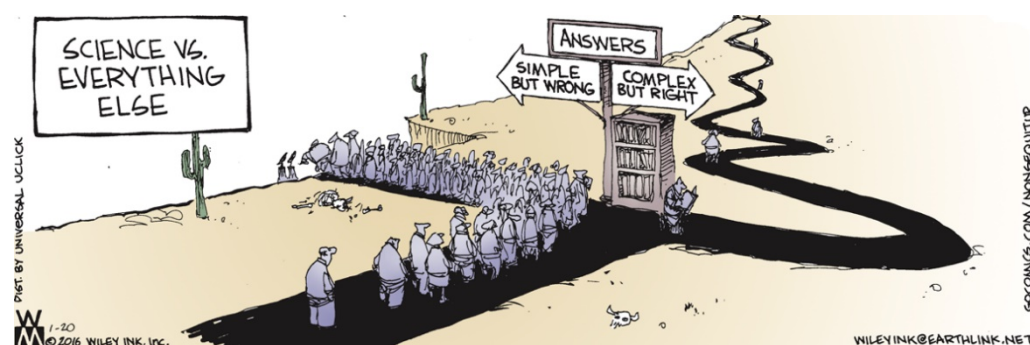
Academic Editor: Julian Whitelegge

Received: 1 August 2021
Accepted: 29 August 2021
Published: 31 August 2021

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).



(NON SEQUITUR © Wiley Ink, Inc. Dist. By ANDREWS MCMEEL SYNDICATION. Reprinted with permission. All rights reserved.)

1. Introduction

Throughout a lifespan, the genome remains essentially unaltered in every somatic cell of an organism, the exception being epigenetic influences on gene expression, and hence proteins. In contrast, the proteome is constantly changing and responding to stimuli, both internal and external [1]. The proteome was first defined as, “the PROTEin complement expressed by a genOME” [2] but here we will emphasize how and why it is much more complex than this original definition implied. It is this complexity that must be respected and addressed to ensure the best possible analyses of proteomes.

For the field of proteomics to effectively move forward, there must first be a discussion and broader consensus on what a ‘protein’ is. To what exactly does the word ‘protein’ refer? It is well known that proteins serve as antibodies, enzymes, messengers, and so much more [3]. However, the central dogma that we are so familiar with is as follows: DNA is transcribed into mRNA and then translated into an amino acid sequence, which is the primary structure of a protein [4–6]. Amino acids are thus the backbone of proteins, but they do not fully or effectively define its function, localization, or interactions. DNA/RNA coding sequence, mutations, and variability of translational start site are key elements that reveal information about primary protein structure and chemistry. These elements—along with the immense number of possible modifications that can occur at any given point before translation, after transcription, or in-between (e.g., RNA splicing, alternative splicing, alternate open reading frames (AltORF) [7–9], single nucleotide polymorphisms (SNP) [10,11], and mRNA editing [12–17]), post-translational modifications (PTM), and adducts)—each yield multiple different protein species of a given amino acid sequence, and each such species, or proteoform, differs in its biological function (Figure 1) [5,18,19]. Being the critical players in the majority of biological activities, the presence/absence or alterations in the abundance of given proteoforms can disrupt physiological functions, thus leading to disease phenotypes [20–22]. This makes proteoforms the most selective and useful indicators (e.g., biomarkers) of disease, remission, response to therapy, as well as the most specific targets for the rationale design of therapeutics.

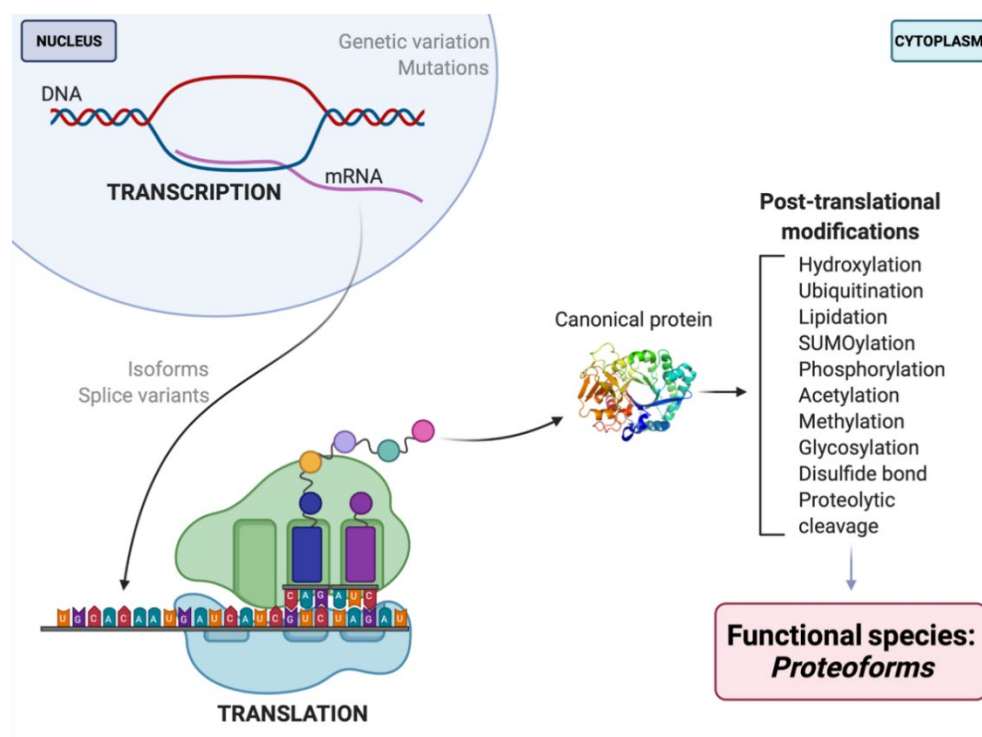


Figure 1. Schematic illustration of proteoform synthesis. Depicted are a handful (but not all) factors contributing to final proteoform configuration that are not seen in, nor predicted by, the central dogma. The PTM noted are but examples of the 100s of currently identified native modifications [23,24]. Each modification that occurs throughout the development and lifespan of a given amino acid backbone will yield multiple different proteoforms, each differing in their biological localization and/or function.

This then raises the question of what exactly is a ‘protein’? In trying to address the complexity of the proteome, ‘protein’ is essentially a nonspecific term that generically refers to a canonical amino acid sequence (which is itself a proteoform). Proteoforms define function (or dysfunction) within an organism and drive cellular functions at the molecular level; therefore, one gene = one protein = one function is far from any longer being an

acceptable logic [3,5,6,20,21,23,25–29]. The generic ‘protein’ term is thus useful for its breadth in referring to a class of molecules. As the genuine complexity of proteomes becomes increasingly more obvious, use of more specific terms—‘protein species’ or ‘proteoforms’—more accurately conveys the complexity being addressed, while ‘canonical protein’ more accurately describes only a common amino acid backbone.

While only the terms proteoform or protein species will be used here, there are other terms that have sought to deal with the limitations of the central dogma but only hint at the real complexity of proteomes. ‘Moonlighting proteins’ are defined as a class of multifunctional canonical proteins derived from a single gene that exhibit multiple biochemical and/or biophysical functions said to be due to changes in cellular localization, cell type, oligomeric state, PTM, or the cellular concentration of a ligand, substrate, cofactor, or product [5,30–33]. This does not include gene fusions, multiple RNA splice variants, or pleiotropic effects, although these can also modify the function(s) of a given canonical protein. Another term, ‘morpheins,’ is used to describe proteins that change shape and assemble in alternate configurations with different functions [34,35]. Noncanonical open reading frames (ORF) have been the subject of intense investigation since their discovery, and their functional products are proving quite important [36]. An ‘alternative protein’ is defined as the product of an AltORF that does not match the sequence of the protein from a reference ORF of the same gene [7–9]. Additionally, ‘microproteins’ challenge the common definition of a gene, being functional entities arising from small ORF and consisting of ~10–100 amino acids [37–40]; these are missed in the standard definition of an ORF and thus of a canonical protein. Finally, the recent revelation of a reverse transcriptase activity in human cells (via Polymerase θ) may also indicate additional complications for genome and proteome analyses [41]. These are only a few of many observations that indicate the sheer complexity of proteomes and what actually constitutes a ‘protein’; it is crucial to incorporate all into our discussions and assessments if we are to genuinely address proteome complexity.

Currently, there are a number of methods that are used for proteomic analysis, and in an ideal world these would identify and quantify all proteoforms in a cell or tissue at a particular time and under a given set of environmental conditions [1,4]. However, as the complexity of the proteome is greater (and more dynamic) than that of the genome or transcriptome, deep analysis—routinely identifying and characterizing proteoforms—is much more difficult, particularly because amino acid sequences cannot be amplified like DNA/RNA.

Furthermore, proteomics does not assess protein expression, but rather the abundance of species or inferred canonical proteins depending on the methods used. Protein expression is difficult to quantify as there are conflicting data concerning the relationship between mRNA levels and that of the corresponding canonical proteins [42,43]. This is due to the fact that there are multiple processes beyond transcript concentration that contribute to establishing the amount of a given canonical protein present at any given time [44]. The presence of an mRNA tends to indicate the presence of the corresponding canonical protein however, it does not provide information as to what happens following transcription (e.g., proteolytic processing or other PTM, complexation) nor how quickly it may be degraded. Thus, whether a direct correlation exists between the levels of mRNA and the canonical amino acid sequences coded for, it seems unlikely—if not impossible—that any such correlation exists between mRNA and specific proteoforms. That is the critical point missed in all attempts to draw such correlations; conventional proteogenomics/transcriptomics based on the central dogma does not address the actual biological complexity.

Simply, proteomes are complicated and multifaceted, and thus proteomic analyses must be as well: this must be recognized, appreciated, and respected. Simplifying the analysis and/or increasing its throughput does not simplify the issue. Thus, it must also be recognized and accepted that each analytical method currently applied in the study of proteomes has strengths and weaknesses, and that in some cases previously identified

weaknesses have been effectively addressed while in others, weaknesses are only now being identified and must not be dogmatically perpetuated.

Our goal here is to critically evaluate the current state-of-the-art in proteomics: what are the pros and cons of available approaches, and thus what works and to what extent, what perhaps works less well or has pushed as far as it can regarding information, and where are improvements, refinements, or recalibrations needed if we are to fully develop a field that has the capacity to contribute as robustly as possible to systems biology [27–29,45]. We will also propose that the only way for proteomics to accomplish this is to be transdisciplinary. This includes understanding that proteomics is fundamentally based in analytical chemistry, operates at multiple disciplinary interfaces, and the need to integrate the numerous proteomic methods that currently exist to yield a unified discipline. There is also the need to better integrate proteomics with the other ‘omics’ to understand changes underlying disease and to identify rational biomarkers and treatment targets—but that is a discussion for another time [46,47].

While aspects of what we will discuss have been raised and reviewed over the years [1,3,27–29,48–53], here we seek to focus on the integrated assessment of proteomics as a whole—indeed, its need for integration—beginning with questions concerning terminology used. Our goal is to broadly consider the sheer complexity of proteomics and how this is addressed by current operating philosophies and associated analytical approaches. Yes, these are the ‘hard’ questions, but they must be addressed as directly, transparently, and realistically as possible to ensure that the field can move most effectively forward with routine quantitative assays that address the depth and breadth of proteomes as well as the need for effectively targeted analyses.

2. What Is a Proteome?

The definition of a proteome differs depending on who you ask. For some, the sum of all protein species during the entire lifespan of an individual is considered the ‘proteome’ [4], while others believe it is the sum of proteoforms expressed in a select biological material at a given time (e.g., specific tissue, cell type, organelle, or fluid) [50]. While both are reasonable, a consensus on a definition would be useful as concise and consistent language is key to the clearest, most unambiguous communication. The literature is replete with reports of the ‘total’ proteome analysis of a particular sample (e.g., blood, brain, and saliva) when only a sub-fraction of the proteome was investigated (i.e., specific ranges of molecular weight (MW)/isoelectric point (pI) [i.e., 3–20 kilodalton (kDa), pI 6–10], soluble fraction, and/or only canonical proteins). There must be clarity in defining what is being analyzed. If we are to adopt the definition, “*all protein species during a lifetime*,” [4], then there is currently no lab nor methodology that has or can effectively identify and quantify the human proteome, let alone that of any other species. Perhaps, then, the definition, “*the sum of proteoforms expressed in a selected type of biological material at a certain time*,” [50], is more appropriate from a purely practical standpoint, if perhaps not yet fully technologically feasible. Sub-proteome investigations would thus constitute any analyses focusing on only a select portion of the total sample (e.g., total membrane, total soluble, or <40 kDa).

Since completion of initial human genome sequencing in 2001, a number of studies have demonstrated numerous errors and confirmed that a large number of sequences (~8–10%) are missing from the reference genome (to some extent possibly due to inherent bias) [54–59]. This also does not take into consideration the ongoing identification of new AltORFs and small ORFs. As there is still no firm consensus on the size of the human genome, and the number of protein-coding genes is currently thought to be ~19,000–20,000, it is difficult, at best, to reasonably estimate the size of the human proteome [23,56,59–61]. Notably, this number also does not include regions coding for alternate RNA species (e.g., long non-coding and micro-RNA) that in many cases also affect canonical protein levels. If the one gene = one protein hypothesis were correct, then there should be a fixed number of identifiable, unmodified (canonical) human proteins. However, this simple definition is inconsistent with the presence of variants of those canonical proteins.

It is therefore not surprising that attempts to estimate the number of proteoforms within the human proteome range anywhere from 98,000 to 6 million, and even >1 billion potential species [5,23,50,61]. Nonetheless, much research continues to only infer the presence of canonical proteins. Despite the uncertainty of how many proteoforms exist within the human proteome, there are methods to determine the variety and quantity of proteoforms present in different samples, although not yet with the sensitivity to fully address the potential dynamic range of the species present (i.e., one copy—a few copies—millions of copies). Rather than despairing over this or bemoaning the fact that granting agencies simply cannot afford—for both financial and logical reasons—to put the latest new mass spectrometer into every lab, every year, we must capitalize on the techniques, instrumentation, and analyses that are proven, and open to even further improvement. We must embrace the complexity of proteomes by continuously striving for better, deeper analyses.

3. Proteomics

Proteomics is thus defined as the systematic analysis of proteoforms that constitute a given proteome, including their diverse properties. However, this is by far not a consensus definition considering that the bulk of current published research seeks to assess or infer the presence of only canonical amino acid sequences as opposed to proteoforms. Nevertheless, with the working definition presented, the aims of proteomics as a discipline are to provide accurate, reliable, and detailed descriptions of the proteoforms present and their functions, thereby providing rational insight into the molecular mechanisms underlying biological systems in both health and disease [3,4,27–29,51,62,63].

At a practical, day-to-day research level, the main goal of proteomics is to identify and quantify all the proteoforms present in, for example, a cell or tissue—a goal that has yet to be achieved for perhaps any species. For this to be accomplished, sequence, quantity, nature and state of modification, interactions with other proteoforms/molecules, activity, subcellular distribution, and structure of every proteoform would have to be known [1,5,62]. While this is not currently possible, many different approaches have been designed to assess proteomes, at least at the compositional level, although some have only a limited focus on proteoforms, or on only a single type of PTM (e.g., phosphorylation).

Currently, there are two general approaches to address proteome composition: ‘discovery’ or a ‘targeted’ (Figure 2). Discovery proteomics is exploratory, untargeted—hopefully unbiased—and aims at mapping a given proteome or sub-proteome as deeply as possible without any preconceived notions as to what exactly will be found beyond differences between two or more conditions. Using the identifications made in a discovery approach, a targeted approach is critical for validating identified protein species of interest on a larger scale (i.e., validation of potential biomarkers). Thus, currently, targeted proteomic workflows generally involve the selective detection and assessment of particular canonical proteins of interest, hopefully with high sensitivity, quantitative accuracy, and reproducibility [64–67].

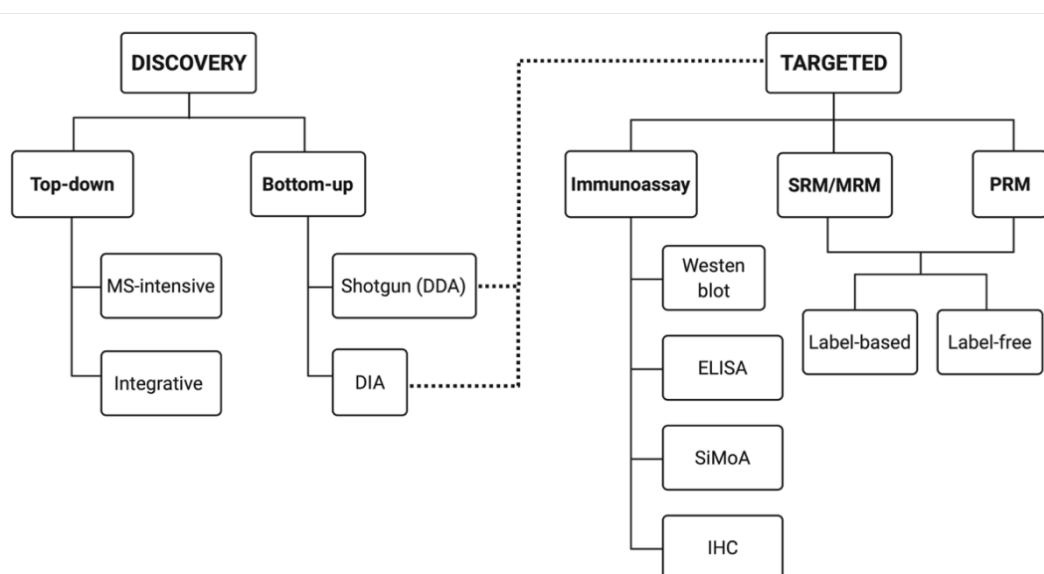


Figure 2. Proteomics: discovery and targeted approaches. Discovery proteomics is defined by two main approaches: top-down (resolution of intact protein species) and bottom-up (peptide mass spectrometry (MS) of proteolytic digests). Targeted proteomics involves either antibody- or MS-dependent approaches. Data dependent acquisition (DDA) and data independent acquisition (DIA) were initially developed for discovery but can be modified to also serve in a targeted approach.

However, if the discovery approach has misidentified a particular proteoform or change in abundance, then the targeted approach will produce contrasting results, or vice versa. For example, a study identifies phosphorylated protein X to be significantly linked in causation of a disease. An antibody-based approach (Section 5.2) is then used to validate this result but finds that there is no statistically significant link between total detected canonical protein X and the disease condition. This occurs because a discovery proteomic method was used to identify a specific proteoform but the ‘standard’ validation attempt in a targeted approach usually only identifies total canonical amino acid sequences. This highlights the need for discovery *and* targeted proteomics to be as sensitive, accurate, and integrated as possible. Unless both approaches work in tandem, with the same definitions and objectives, they can simply confound one another.

4. Discovery Proteomics

Within discovery proteomics, there are two general analytical approaches: top-down and bottom-up (Figure 3). There is, however, a notable discrepancy concerning the definition of top-down. Some use purely instrumentation-centric definitions based solely on what enters the mass spectrometer, ignoring the importance and indeed use of upfront separations/prefractionations in determining that [68]. Noting again that proteomics is analytical chemistry [28], we emphasize that, following proteome extraction, top-down approaches resolve intact proteoforms prior to their identification (e.g., by mass spectrometry [MS] or immunoblotting) whereas bottom-up methods use peptides generated from a gross proteolytic digestion of the extracted proteome as surrogates of the original intact proteoforms to only *infer* the presence of *potentially intact* canonical proteins [3,69]. The bottom-up method thus provides a rapid scan to identify likely canonical proteins but lacks the capability of routinely providing critical information concerning the myriad of proteoforms that define proteomes.

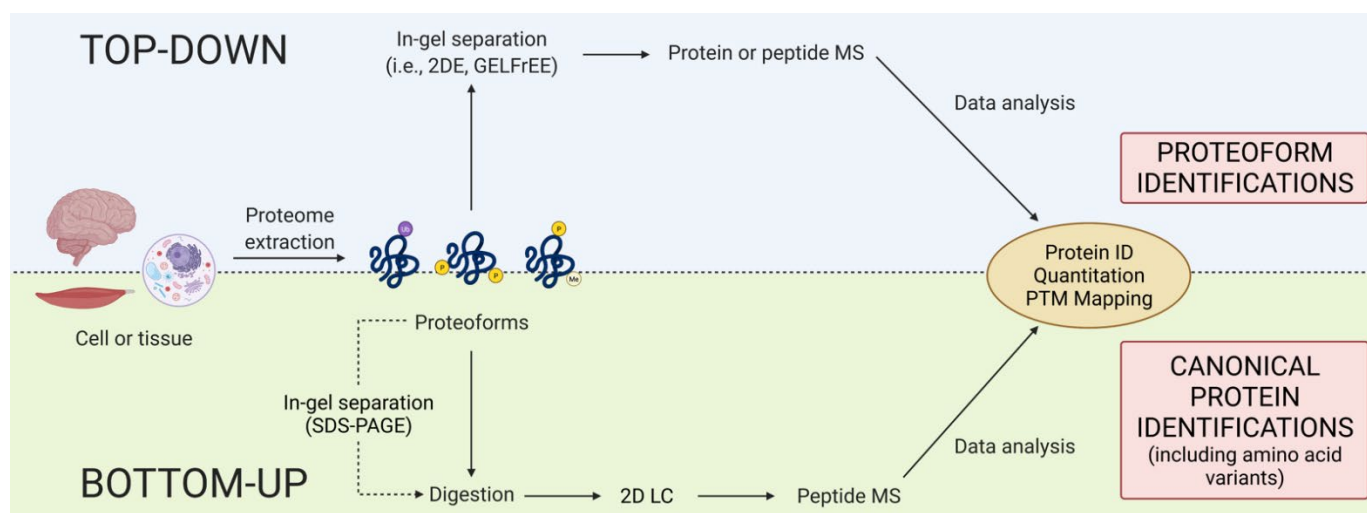


Figure 3. Top-down versus bottom-up proteomics. This schematic depicts a general description of the workflows for these two discovery approaches. While both rely on final MS analysis for identifications (not to oversimplify the analysis of intact proteoforms), the main differences lie in the up-front analytical approaches. Top-down resolves intact proteoforms prior to MS while bottom-up generally bypasses any initial separation technique. Thus, top-down provides proteoform information while bottom-up can only provide (limited) amino acid sequence information. Nonetheless, perhaps the most important point to immediately emphasize is the critical importance of high quality/high resolution MS to proteomics as an integrative discipline, now and into the future.

Nonetheless, this bottom-up approach has enabled the Human Proteome Project (HPP) of the Human Proteome Organization (HUPO) to recently announce an almost complete (~90%) cataloging of canonical proteins corresponding to known/accepted gene coding regions [70,71]. Conservatively, an estimated 1899 canonical proteins remain to be identified; what that corresponds to in terms of proteoforms is anyone's guess. This critical update thus qualifies that biologically functional proteoforms have not been yet identified due to the difficulty in their measurement, and future projects will thus also be needed to incorporate heterogeneous RNA splicing, PTM, and single amino acid variants [70]; we suggest that international efforts should also include alternate proteins [7–9] and very small coding regions [8,72]. Additionally, researchers will need to be cognizant of ongoing developments in genome analysis and incorporate them into any future analyses. It is also now clear that some previously identified genes do not code for amino acid sequences, and thus these do not in fact yield proteins. Thus, while this new catalogue of canonical protein sequences is indeed a critical milestone, it is the minimally essential step forward in terms of a fundamental tool to now undertake the necessary critical deep analytical dive(s) into proteomes.

4.1. Bottom-up

In recent years, bottom-up has come to dominate the field of proteomics [27,52,73,74]. The main reasons for the preferred use of bottom-up over top-down are *claims* of high throughput, ability to produce large volumes of raw data, and improved sensitivity, as well as the refrain that it is 'faster' and 'easier,' although, even if true, those are perhaps not ideal reasons to select an analytical approach. This approach—colloquially known as shotgun proteomics—combines liquid chromatography (LC) and tandem MS (MS/MS) to analyze peptide mixtures obtained from total proteolytic digests of proteome extracts (Figure 4). Data dependent acquisition (DDA) methods are predominantly used to identify canonical proteins, automatically selecting precursor ions from those detected in a survey scan. The approach is designed to select precursor ions in order of decreasing relative abundance while requiring a minimum threshold of abundance [21,25,66,75]. The identified peptides are then matched to a corresponding canonical protein (i.e., amino acid

sequence) by searching protein sequence databases [76]. While peptide MS is an extremely powerful technology, there are several concerns with its application.

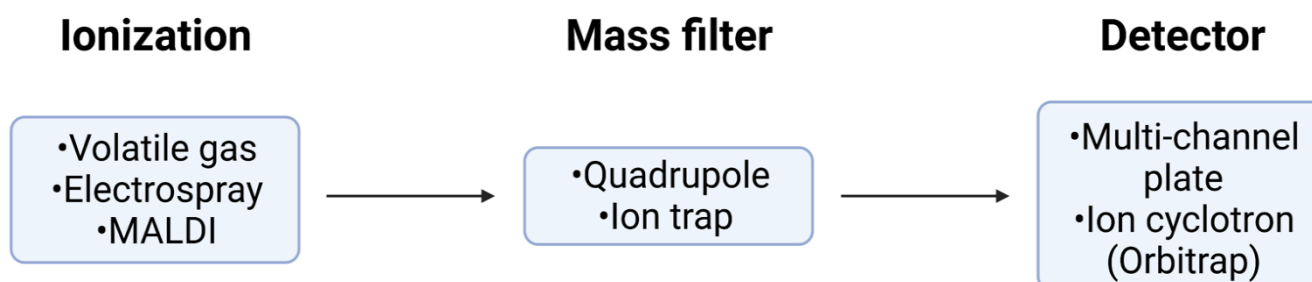


Figure 4. Schematic of MS/MS. A basic overview of the four main systems of MS/MS and the different methods for each. Peptides undergo separation via LC prior to ionization. Peptides are then transformed into ions before entering the mass filter where precursor ions are then selected prior to collision-induced dissociation. The resulting fragment ions are then separated and transmitted to the detector. The mass filter measures the mass of the ions and the detector counts the ions. This information can then be combined to determine the mass-to-charge ratio (m/z), leading to identification of a peptide.

The most noted is that hydrophobic peptides tend to stick to LC columns and also produce poor MS signals when routine soft ionization techniques (e.g., electrospray ionization [ESI]) are used. Thus, improving the coverage of hydrophobic proteins has been one critical area of focus. DDA methods also tend to be biased toward the most abundant proteins, making them vulnerable to high sample complexity and/or dynamic range [21,25]. Overall, the bottom-up approach is purely assumptive as it identifies proteolytic peptides and from those, infers the primary structure of a presumably intact canonical protein; however, it therefore does not actually sequence proteoforms [6,63,77]. Thus, the real question for bottom-up analyses is ‘how many corresponding proteoforms does the apparent change in total abundance of a putative intact canonical protein actually represent?’ This is further complicated as there is no general consensus on the number of peptides needed for a positive identification of a canonical protein [6,28,76,78,79]. Criteria seemingly vary from lab-to-lab (and journal-to-journal). Thus, substantial sequence coverage—some argue, with solid rationale, that only 100% coverage is optimal [5] or at least several peptides that roughly cover the full span of the sequence (i.e., minimally near the C and D termini and the middle)—is required to confidently assume a canonical protein identification [28,80]. We suggest using a minimum of three peptides, that span the range of the apparent canonical protein sequence, as criteria for a positive identification: this is further strengthened if proteoforms have first been resolved (i.e., top-down analysis) [27,28,81]. Indeed, there are many claims in the literature of canonical protein identifications that are based on very limited data (i.e., one or two peptides), making the inherent assumptions of even greater concern [27,82]. Notably, manufacturers (who heavily influence directions in proteomics research) promote new MS instruments based on the weakest possible sequence criteria and thus, each year claim a certain increase in the number of peptides and thus supposedly canonical protein identifications. However, these claims omit proteoform identifications.

Thus, one of the main complications that arises is a lack of standardized methods and thus poor reproducibility between different laboratories [80,83]. Regrettably, this has likely been a problem since the first SDS-PAGE gel was resolved in the *second* lab to ever try the method. In the case of identifying canonical proteins from available databases, a key example of this issue is an independent and unbiased study involving 20 highly purified recombinant proteins (expressed in *E. coli*), and that contained at least one unique tryptic peptide of 1250 ± 5 Dalton (Da) [76]. This sample was sent to 27 independent MS laboratories with the task of identifying all 20 canonical proteins and unique peptides us-

ing their routine procedures and instrumentation. Only seven labs identified all 20 proteins and only one lab reported all tryptic peptides of 1250 Da. Of the other 20 labs, the list of identified peptides differed and there was a general inability to identify even highly purified proteins in such a low complexity sample; the issues were largely attributed to differing protocols and methods across labs, as well as stochastic and irreproducible ion selection [76,84]. When not carried out using stringent protocols and criteria, in particular routine technical replicates, this approach has shown a lack of reproducibility, yielding inconsistent results even when analyzing the same sample, as well as high rates of false positives and false negatives [5,76,79,85,86]. Additionally, it is difficult to identify variants/modifications with routine peptide MS as these peptides are most often of lower abundance and more difficult to identify from their fragmentation spectra compared to nonmodified peptides (Figure 5) [82,87]. It is important to note that while it is entirely feasible to identify PTM on peptides, it is far from a quick and easy task (Section 4.3), and certainly confounds the promotion of shotgun analyses as ‘high-throughput’ (not to mention requiring appropriately sensitive instrumentation) [27].

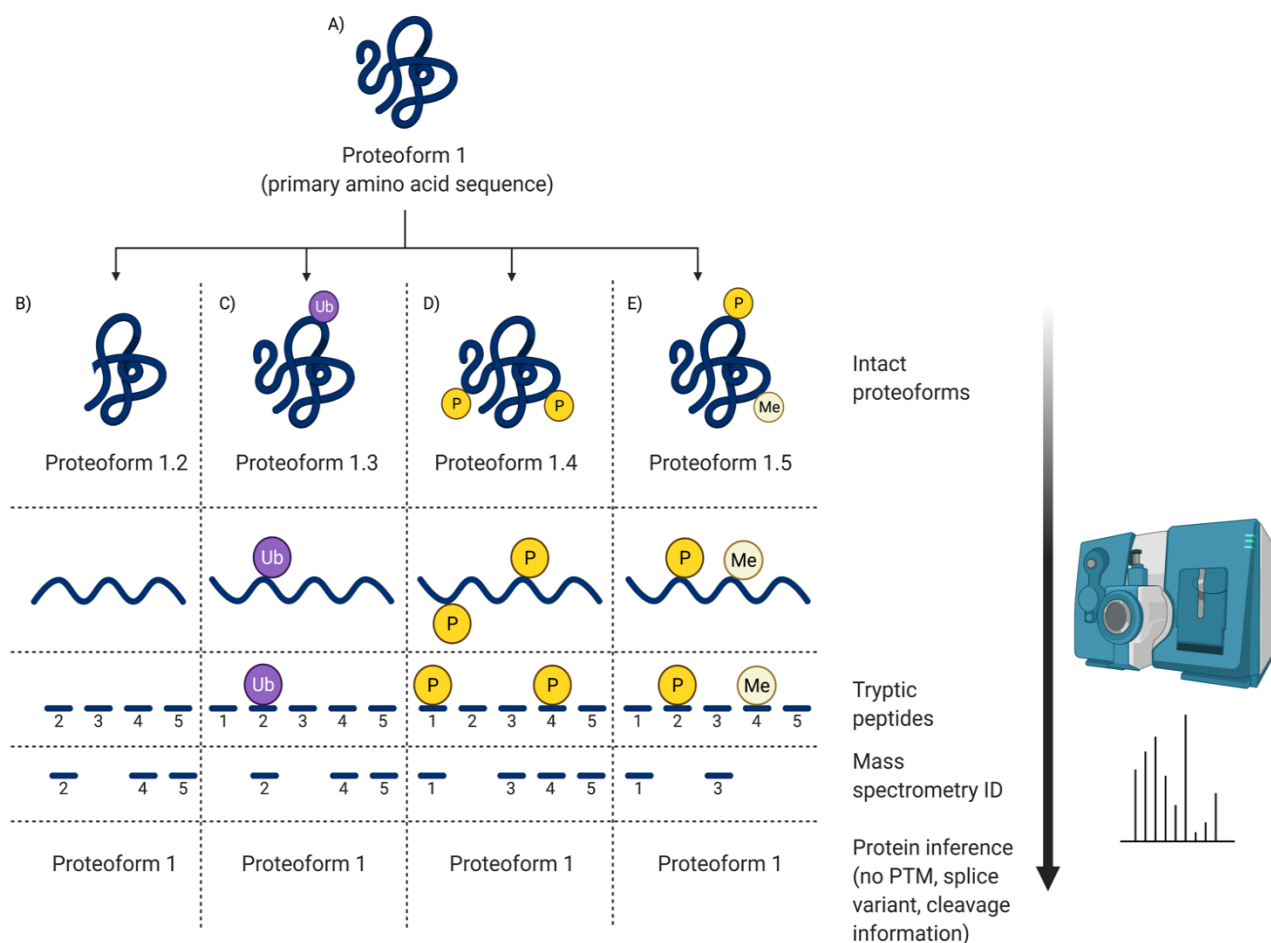


Figure 5. Peptide MS. This illustrates the information obtained via routine peptide MS. (A) Canonical protein (primary amino acid sequence); (B) PTM = Proteolytic cleavage; (C) PTM = Ubiquitination; (D) PTM = Two phosphorylations; (E) PTM = Phosphorylation and methylation. As only peptides are being sequenced, the ‘canonical protein’ identifications are based on inference; thus, as shown in (B), even though there has been a native proteolytic cleavage to generate another proteoform (i.e., likely to modify the biological activity of the canonical protein—Proteoform 1), it will not be detected by inference identification. Notably, other than potentially identifying SNP, no proteoform information is obtained via peptide MS without specific additional processing and assays.

Furthermore, any attempt to quantify proteoforms in such bottom-up analyses would be impossible considering that all information essential to doing so is lost upon the

initial gross proteolytic digestion of the sample. Thus, as widely noted, the absence of a canonical protein from the list of those identified does not indicate the absence of the protein from the sample (nor, more importantly, select, low abundance, proteoforms) [66]. In all proteomic analyses, the concern is always what is missed—might it be the key player in the mechanism or condition being studied? This would thus also appear to be the current issue with newer approaches to protein/peptide sequencing (e.g., nanopores and DNA-PAINT), which focus on amino acid sequences but not PTM [6,88].

Seeking deeper analyses, data independent acquisition (DIA) methods have been used as an alternative to DDA. These methods either acquire fragmentation spectra of the entire mass range simultaneously or set predefined windows to cover the whole mass-to-charge ratio (m/z) ranges of proteolytic peptides, thereby eliminating the ‘one peak at a time’ selection process used in DDA [21,25,89]. All the peptide mass ranges within this window are acquired without pre-selection, leading to an unbiased fragmentation spectrum record of the complete set of peptide precursors of a given sample [21,84]. A popular variant of this method is Sequential Window Acquisition of all Theoretical Mass Spectra (SWATH-MS). SWATH-MS fragments ionized peptides systematically using large precursor isolation windows and records all fragment ions simultaneously, yielding high specificity identification of canonical proteins [84,90,91]. This method has numerous advantages including decreased fragmentation spectra complexity, improved precursor ion selectivity, and increased proteome coverage. While it is low cost and requires simple sample preparation (in comparison to label-based methods, Section 5.3.1), there is a strong need for standardization of equipment and protocols between laboratories [21]—an issue so true of all techniques used to analyze proteomes. Thus, while it is possible to use SWATH-MS to create fragment ion maps of all MS-measurable peptides in a proteome that can be used universally to analyze and compare samples *in silico* [90], lack of standardized operating procedures severely limits the likelihood of consistent and comparable data between different lab groups. Additionally, there are some claims that SWATH-MS is slightly limited in terms of sensitivity and dynamic range, indicating the need for further refinements before it can be used most effectively for biomarker discovery and validation [21,84].

Thus, while bottom-up may serve as a rapid, low-resolution scan of a proteome, it enables deep, high-resolution proteome analyses when paired with two-dimensional gel electrophoresis (2DE). The 2DE protocol was developed in 1975, essentially initiating the discipline of proteomics [92,93]. Since inception, this method has also received its share of criticism, mainly around original issues concerning the resolution of low abundance proteoforms with extreme *pI* and MW, and hydrophobic protein species [73,94,95]. Most notable, however, being a mature technology, these issues have been addressed and the approach has been substantially refined over the last two or more decades to address purported shortcomings (Section 4.2.2). Unfortunately, many of the negative claims that still appear in review articles have simply become dogma, often perpetuated by those who have no experience with the technique, nor certainly reviewed the relevant primary literature from at least the last twenty years [1,29,49,81,96–105]. Simply, refined 2DE and its modifications can effectively resolve many (hundreds of) thousands of proteoforms across a broad range of classes and physico-chemical characteristics (i.e., soluble, membrane, acidic, basic, large, and small) including those of low abundance, and do so in parallel technical replicates [27,29,81]. Like bottom-up approaches, the biggest issue with the method is likely the lack of consistent protocols between labs, making unified gel databases somewhat untenable unless direct and routine calibration of both *pI* and MW are implemented.

4.2. Top-Down

4.2.1. Integrative

There is currently only one high resolution and high sensitivity method that can provide a genuinely deep assessment of proteomes at the critical level of proteoforms: 2DE coupled with LC/MS/MS [1,27–29,52]. The first dimension of 2DE is isoelectric focusing (IEF) which separates protein species according to their pI —the specific pH at which the net charge of the proteoform is zero [106–108]. Following IEF, species are then resolved by size (i.e., nominally and MW) via sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) [92,109]. The importance of the two separate dimensions lies in the fact that canonical proteins speciate into proteoforms that differ in MW and/or pI , and thus resolve into different spots on the gel, despite having identical (i.e., canonical) amino acid backbones (Figure 6). Following staining and quantitative image analysis, select protein spots are excised from the gel and proteolytically digested prior to LC/MS/MS [81,98,102,104,105,110–112]. Selective staining, deep imaging, and third-dimension separations (3DE) may also be employed to further improve resolution (Section 4.2.2).

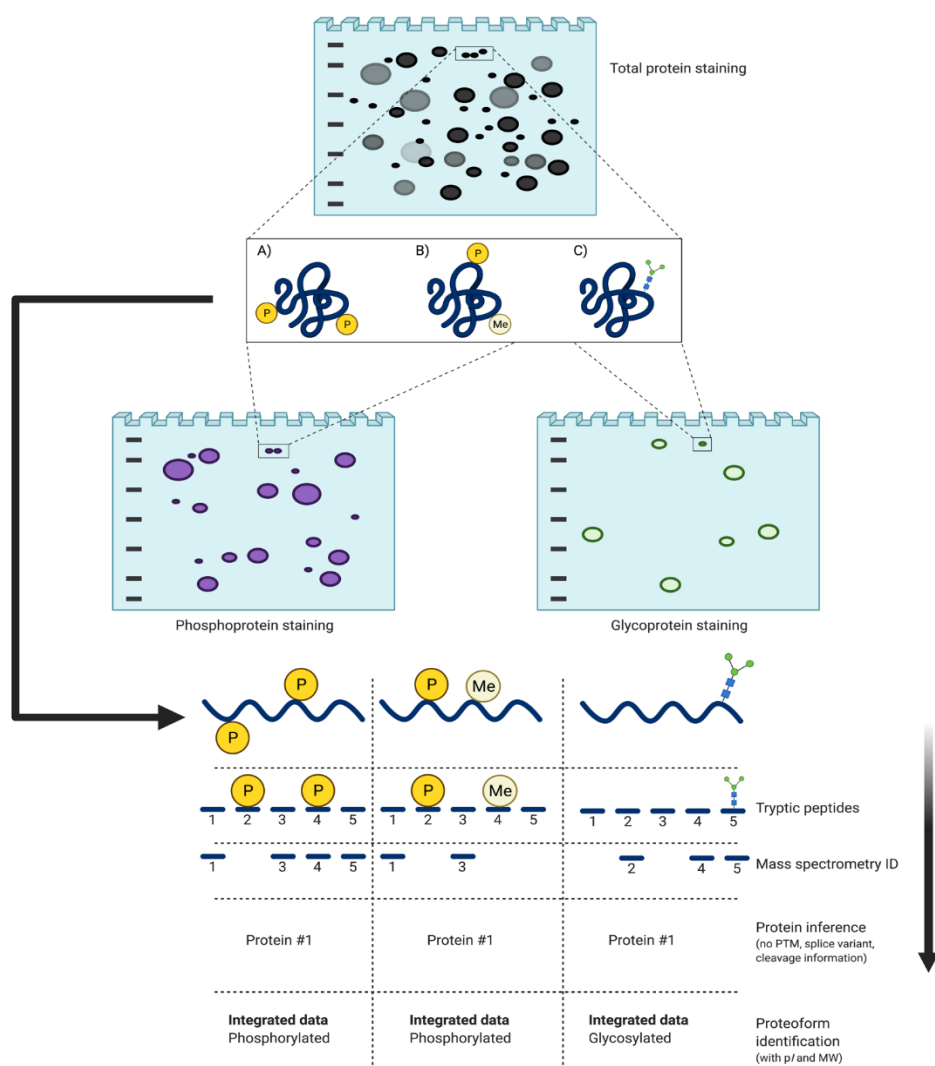


Figure 6. Integrative top-down proteomics via 2DE and PTM post-staining. (A) PTM = two phosphorylation sites; (B) Phosphorylation and methylation; (C) Glycosylation. Different PTM can change the pI and MW of a protein species thus, altering its final resolution in a 2D gel, which can be seen using a total staining method. Additional selective staining (e.g., phospho- and glyco-protein staining) can be used to identify these proteoforms prior to digestion and MS. Phosphorylation yields more acidic species and sugar groups increase MW [113]. Typically, a chain of protein species as seen in the 2D gel is often indicative of an identical canonical protein with varying modifications.

Many choose to bypass this method, supposedly due to the fact that it requires manual dexterity and demands technical precision. Strangely, this hackneyed comment not only describes most if not all scientific techniques but, again, tends to generally appear in reviews written by those with seemingly little to no 2DE experience. While bottom-up is praised for ostensibly being a high-throughput technology, it requires its own fair share of time, manual dexterity, and technical precision in packing, optimizing, and subsequently effectively flushing/cleaning different LC columns, adjusting fittings, fixing pumps, and cleaning clogged electrospray systems, not to mention the need to run *sequential* technical replicates (assuming this is even being done, since parallel replicates are not possible). Simply trying to improve LC resolution has been a major topic for decades, including the need for better columns [114,115]. It would thus behoove researchers using MS core facilities or commercial services to fully understand exactly how the analysis of their samples was conducted and whether the shotgun data they receive are based on technical replicates. This begs the question: does this approach truly require less time and technical input or are many researchers simply unaware of the issues associated with it? Overall, then, we respectfully suggest that the focus should be on the quality of the data that can be delivered, and a willingness to invest the time to acquire the best possible data at each stage of analysis. *It is not the rate or volume of data generated but rather the quality that ultimately matters.*

4.2.2. 2DE: Addressing the Dogma

Extraction of membrane (i.e., hydrophobic) proteoforms has long been considered an issue with 2DE sample buffer. This is in part heavily dependent on effective sample handling before extraction [97]. Thus, while it has been established that hydrophobic proteins are effectively resolved by 2DE, their detection in-gel is often ‘swamped’ by the more abundant soluble proteoforms. It would thus seem that contrary to two decades or more of dogma, the real issue with 2DE is proteoform detection rather than resolution [81,99,100,102,103,116,117]. Again, the maturity of the method has yielded effective solutions.

There are numerous stains that may be used for total proteome detection. While popular, SYPRO Ruby, a sensitive, easy to use, and fully MS compatible fluorescent stain, is quite expensive (as are variants that have appeared since its release ~20 years ago) [99,116]. To address this, a colloidal Coomassie Brilliant Blue (cCBB) staining protocol was tested and refined as a near infrared (NIR) dye rather than a densitometric stain; this proved a broad initial solution to the issue of in-gel detection with sensitivity comparable to SYPRO Ruby while also being easy to use, MS compatible, and very cost effective [99,100,116,118]. Thus, current in-gel detection limits for *intact* proteoforms are in the low femto-to-attomole range, comparable to *peptide* detection in routine shotgun analyses [103].

To dig still deeper into the proteome, it was found that re-imaging 2D gels after excising ~20% of the highest abundance spots of near-saturating signal strength enabled the detection of very low abundance species [102]. These spots, along with areas at the pH extremes and unresolved small species/peptides in the migrating front, can be subjected to a third round of electrophoretic separations (i.e., 3DE), further enhancing the depth of proteome analysis [96,119]. While it was originally thought that one spot on a 2D gel contained perhaps 1–2 protein species, deeper analyses with ever more sensitive mass spectrometers has confirmed that there can be in the range of 200 or more proteoforms in a given spot, depending of course on spot size, density, and thus also quality of resolution [96,98,104,105]. Therefore, optimal in-gel resolution is needed to ensure the best possible assessment and identification of constituent proteoforms by LC/MS/MS [81,96,102].

As a mature technology, there also exist variations of 2DE, the most extensively used being two-dimensional *difference* electrophoresis (2D-DIGE). Utilizing the same fundamental methodology as 2DE, 2D-DIGE enables multiple protein extracts to be resolved on the same 2D gel (i.e., multiplexing) [120]. This method is often used to compare distinct

samples and involves labeling each with one of two related fluorophores (i.e., Cy2 or Cy3) and a pool of both samples is labelled with a third (Cy5) [121]. In theory, the objective of this multiplexing is to reduce inter-gel variability.

Two labeling methods can be employed in 2D-DIGE: minimal (lysine-reactive) or saturation (cysteine-reactive). The lysine-reactive dyes are hydrophobic and label by covalent modification, leading to the removal of multiple charges from the protein species [95,107]. Unfortunately, this causes a decrease in solubility and can lead to sample loss through precipitation. Since lysine residues make up ~6% of all amino acids in human proteins, a saturation labeling approach cannot be used as it can cause large *pI* and MW shifts [122–124]. Thus, with minimal labeling, <5% of each protein species is thought to be labeled (estimated as one lysine residue on one out of twenty protein molecules) [95,107,122,125–127]. It has been seen that a single dye molecule (434–464 Da mass addition) per protein species has negligible impact on *pI* and MW of proteins <30 kDa, however proteins of these smaller sizes have poor mobility compared to their unlabeled counterparts, decreasing the quantitative capacity of the approach [95,122,126]. Additionally, these dyes react with thiol-based reductants (i.e., dithiothreitol) [95] and therefore the samples cannot be reduced. However, sample reduction is extremely important for breaking disulfide bridges prior to electrophoresis which helps reduce streaking of spots within 2D gels and the potential identification of ‘false’ proteoforms [128–130].

Saturation labeling is said to be somewhat less of a concern with cysteine residues as there are, on average, fewer per protein compared to lysine [122,126]. Cysteine-reactive dyes are fluorophores using maleimide chemistry to label all free cysteine residues. Since the reaction is with the thiol group of cysteine, the sample requires reduction prior to labeling to expose the residues [126,127]. Unfortunately, only ~96% of human canonical proteins possess at least one cysteine residue [131], leaving the remaining canonical sequences unlabeled; an additional proportion may not be labeled if the cysteine is not free to react following reduction.

Similar to issues noted with Isobaric Tags for Relative and Absolute Quantification (iTRAQ) (Section 4.3), it is possible that lower abundance proteoforms are less likely to be labelled, especially if lysine or cysteine residues are blocked by PTM or modifications to neighboring residues [117,126,132]. Lysine-reactive dyes are based on *N*-hydroxy-succinimidyl (NHS) esters which undergo nucleophilic substitution with ϵ -amine groups of the lysine residues [126]. Importantly, of the 20 amino acids, lysine is one of the most heavily modified [133]. Lysine is often found at many functional sites (e.g., enzyme active sites and interfaces mediating protein–protein interactions) and is frequently covalently modified by acetyl, hydroxyl, propionyl, butyryl, crotonyl, ubiquitinyl, ubiquitinyl-like (SUMOylation, ISGylation, NEDDylation), formyl, malonyl, succinyl, and methyl groups ([124,133,134] and references within [133]). Lysine methylation occurs when up to three methyl groups are transferred to the ϵ -amine [133]; thus, any proteoforms with these PTM will not be detected with the NHS-CyDye. Notably, the chemical labelling of a proteoform with a CyDye can be considered an adduct and thus, there is also the possibility of the dye disrupting inherent modifications and thus changing proteoforms (although this has yet to be demonstrated).

While cysteine residues are not as prone to PTM as lysine residues, the thiol groups have unique nucleophilic and redox properties that support modifications including oxidation, S-nitrosylation, palmitoylation, prenylation, and Michael addition with oxidized lipid species. Further, while it is rare, cysteine has been seen to be methylated or phosphorylated ([135,136] and references within [135]). Therefore, numerous modifications may block dye binding, further decreasing the quantitative capacities of 2D-DIGE. Again, the reactive dye may also alter existing PTM (e.g., via disruption of thioester bonds). Thus, while an interesting concept, 2D-DIGE does not appear to ensure genuine quantitative assessments of proteomes. Indeed, this appears to be an issue with all covalent labelling-based methods.

Considering the extreme complexity of proteomes, it would thus seem that such targeted methods will not quite work in the manner originally hoped. Noncovalent dyes binding to multiple residues (i.e., SYPRO Ruby and cCBB) thus appear to provide better quantitative detection but it does not appear that a comparable covalent dye has been identified. That is not to say that the binding of noncovalent dyes cannot also be blocked by PTM, but the likelihood of that occurring across the entire surface of a proteoform seems, on average, far less likely.

Thus, while the integrative analytical approach provides high-resolution proteoform separations, there is still room for improvement. In this regard, two principal issues remain with 2DE. First, occasional precipitation in IEF, usually of high abundance species, can interfere with quantitative transfer into the second dimension. This can also confound the identification of genuine protein oligomers as opposed to non-native aggregates formed due to co-precipitation. Although analyzing membrane and soluble proteomes separately [97,102], and supplementing the standard CHAPS detergent with others, in particular a lysolipid [137], seems to help, this issue must be more fully addressed to ensure the best possible proteome analyses.

The second issue concerns the capacity to recover intact protein species embedded in the polyacrylamide matrix; this has been a long-standing challenge in terms of subsequent MS analyses. As the Bis (N,N'-methylene-bis-acrylamide) cross-linked polyacrylamide matrix is insoluble, it is difficult, if not impossible, to quantitatively recover the embedded intact proteoforms; recovery via electroelution is possible but of variable-to-low efficiency, especially for high MW species, leading to a reduced depth of quantitative proteome analysis [138]. Alternatively, the peptides of resolved proteoforms are more routinely released via digestion, most commonly tryptic. While the original concept of a 'molecular scanner' [139] might in theory address this concern it would, minimally, still require genuine quantitative recovery (i.e., transfer) of all proteoforms resolved in the 2D gel. Furthermore, while dissolvable polyacrylamide matrices have been known for several decades [140,141], the chemicals needed would likely result in nonspecific and non-native alterations to the proteome.

Another important approach to top-down proteomics that has emerged involves the application of MS to analyze intact proteoforms as opposed to identifying proteolytic peptides [142–144]. As this method has not yet been integrated with 2DE/gel-based analyses (aside from Gel-Eluted Liquid Fraction Entrapment Electrophoresis [GELFrEE]—Section 4.2.3), it is considered MS-intensive top-down (Figure 7). The aim is to characterize intact proteoforms (i.e., including PTM) and it thus differs from the integrative approach as it seeks to avoid proteolytic digestion, and thus full sequence analysis occurs in the mass spectrometer using alternate approaches to disrupt the amino acid backbone [142]; these include collision-induced dissociation, higher-energy C-trap dissociation, electron-transfer dissociation, or ultraviolet photodissociation. However, this approach is also not without technical issues.

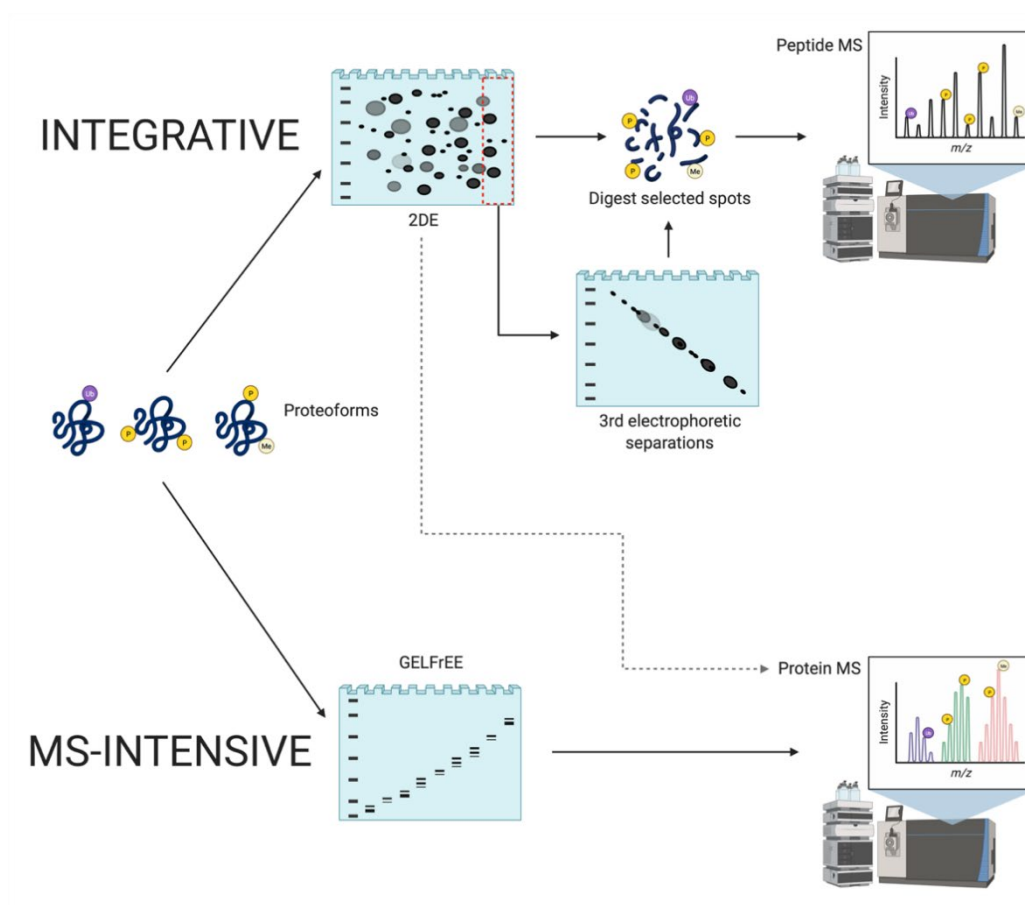


Figure 7. Integrative and MS-intensive proteome analysis. This schematic depicts the workflows of these two top-down approaches. Integrative MS involves the separation of intact protein species via 2DE prior to peptide MS. Additionally, spots of high abundance or areas at the pH extremes and unresolved small peptides in the migrating front can be further subjected to 3rd electrophoretic separations. MS-intensive involves separation of intact protein species, currently mainly via GELFrEE, prior to intact protein MS. Dashed line represents the potential combination of integrative and MS-intensive approaches, which has not yet been pursued.

4.2.3. MS-Intensive

MS-intensive top-down proteomic analyses require the transition of intact proteoforms (i.e., in solution) to positively charged molecular ions (gas-phase). The exact m/z of the intact proteoforms of interest are then measured [142–144]. While this method removes any inference associated with bottom-up peptide MS and provides full sequence coverage of proteoforms, including number, position, and type of PTM on a single polypeptide chain, there remain many technical challenges to be addressed [87,145,146]. First, intact protein MS cannot match the high throughput obtained by peptide MS and is currently unable to handle large-scale analyses [143,147]. Specifically, the inherent difficulty is in producing extensive gas-phase fragmentation of intact proteoforms. Currently, MS technologies are incapable of handling intact proteoforms of larger sizes and mixtures of proteoforms with different physico-chemical properties; this results in loss of certain components or incompatibility of the protein species with MS [78]. Thus, this method can only be consistently applied to species of less than ~30 kDa; there are some reports of select identifications of proteoforms >30 kDa, even up to 104 kDa, although this is completely dependent on the capacity of a few random species to effectively fragment using available technology [18,143–145,147,148]. Thus, in studies identifying larger proteoforms (e.g., >100 kDa), these are currently exceptions—the majority of species that can be effectively analyzed are in the low MW range (i.e., 3–30 kDa) and only a very small handful of larger MW species can currently be effectively identified.

To address these limitations, technological innovations have been developed to improve the assessment of high MW biomacromolecules in various types of samples. However, this process becomes increasingly difficult as sample complexity increases because several components can have exceedingly small differences in their m/z , which makes the analysis of their mass spectrum difficult. Due to the high dynamic range of concentrations in native samples, proteins of higher abundance can suppress the signal of lower abundance proteins (as also seen in 2DE—Section 4.2.2) [96,149]. To address this, a high-resolution mass spectrometer can be used to discern small differences in mass, along with a prefractionation technique to reduce sample complexity and chemical noise. Thus, front-end separation of total protein extracts (i.e., the proteome) is necessary to obtain accurate and reproducible results. GELFrEE has been introduced as a prefractionation step prior to MS-intensive analyses. Quite simply, this is single dimension continuous elution tube gel electrophoresis providing low resolution separation of proteins by MW [109,150]. Such continuous elution gel electrophoresis approaches are well-established, and comparable equipment has been commercially available for more than two decades (i.e., the Prep Cell, Model 491, and BioRad) [151,152] as well as additional designs [153], including simple adaptations to widely used SDS-PAGE gel systems [154]. Such low resolution SDS-PAGE gel-based separations are done prior to the MS-intensive analyses in order to decrease sample complexity. However, there is the issue of possible inconsistency of run times, and this additional up-front gel-based separation step does not correct the current inability of the in-line approach to analyze larger proteoforms but rather fractionates them from the lower MW species that can be analyzed.

In addition to GELFrEE, passive elution of species from SDS-PAGE gels has also been trialed. When comparing the mass spectrums from samples handled with this workflow vs. GELFrEE, they yielded a similar number of proteoform identifications; however, both prefractionation methods still resulted in fewer identifications as the MW of species increased. This can be partially attributed to the decreased recovery rate of high MW proteoforms as the median recovery rate for proteins below 100 kDa was 68%, whereas for those above 100 kDa it was lower (~57%) [155]. Thus, while the MS-intensive approach can indeed analyze some fraction of proteoforms, it is nonetheless also reliant on front-end gel-based separations, raising interesting questions concerning complementarity and integration of approaches to drive the most robust proteome analyses.

Considering the sheer amount of detailed data that the MS-intensive approach can extract per proteoform, one must also understand how the analyses are carried out. The use of a soft ionization technique is paramount as this does not induce dissociation and maintains the integrity of the proteoforms under investigation. Among all available ionization techniques, ESI is the most commonly employed due to its high sensitivity, ease of integration with LC, and its ability to produce multiply charged ions [156,157]. Matrix-assisted laser desorption/ionization (MALDI) is another frequently used soft ionization technique, however, because it predominantly yields singly or doubly charged ions, conventional MALDI is less favorable than ESI for the study of large biomacromolecules [158]. This is a significant disadvantage as multiple charging decreases the m/z of the ions so they may be analyzed within the ranges of most standard mass spectrometers [156,158]. Recently, the development of small emitter tips has allowed for the use of physiological concentrations of non-volatile salts in ESI systems as the use of smaller droplets reduces salt adduction and improves the resolution of charge-state distributions [159]. While analysis of large proteoforms is possible with ESI on older mass spectrometers, it is difficult as the signal-to-noise ratio (S/N) decreases as a function of increasing MW. As the MW of a species increases, so too does the number of charges it can carry and hence, the number of possible charge states, seen as a peak on a mass spectrum, and can vary depending on a number of factors such as pH, protein conformation, or ambient pressure [160,161]. Each peak is surrounded by a cluster of smaller peaks, the number of which depends on the various combinations of naturally occurring isotopes, further contributing to noise. Indeed, the isotopic effects on S/N are more pronounced at a low MW, whereas charge state

effects begin to dominate at a higher MW, however, both need to be considered. Additionally, the effect of chemical noise stemming from various factors such as analyte clustering, multimers, or interfering species, further compounds the arduousness of intact protein detection and analysis. When modelling the decay in S/N against increasing MW with the aforementioned effects considered, there is a pronounced decrease in S/N at a mass of 20–30 kDa. This emphasizes the importance of developing effective separation strategies to remove interfering species, as well as high-resolution mass spectrometers which can distinguish compounds that would otherwise appear as a single peak [161].

Fourier-transform ion cyclotron resonance (FTICR) MS is the gold standard in terms of high-resolution MS and can be coupled with ESI, making it the most valuable instrumentation for intact protein MS-analysis (and by far the most expensive). An FTICR functions by using a Penning trap, which confines ions radially and axially using a magnetic and electric field, respectively. Once excited to their resonant cyclotron frequencies, the ions travel near detection electrodes on which they induce an image current. The signal is then converted to the frequency domain via Fourier transform, from which m/z can be calculated [162,163]. Given that FTICR-MS performance metrics are directly related to the field strength of their magnets, the new 21T FTICR-MS offers the highest resolving power and mass accuracy available ($m/\Delta m_{50\%} > 2,700,000$ at 400 m/z and 80 ppb, respectively) [162,164]. However, the increase in field strength has not yielded an expected proportional increase in mass resolving power when compared to the 18T FTICR-MS. Currently, the potential of the 21T FTICR-MS is restricted by the inability to produce a sufficient vacuum near the ICR cell because the mass resolving power in an FTICR-MS is equal to the frequency resolving power, which depends on the acquisition time, assuming the collision-free motion of an excited ion. A greater acquisition time allows for more data points to be collected, which results in a greater resolution; however, collisions between the analytes and background gas molecules results in signal decay and decreased resolving power. A high vacuum in the ICR cell is therefore required for optimal function [162,165,166]. Recently, a concept was developed for a modified dynamically harmonized cell with a new “zigzag” ion trap configuration, which can improve the vacuum by decreasing the surface area of the cell and incorporating the vacuum tube directly into the working region of the cell however, this has not yet been implemented in an FTICR-MS so the practicality of this new cell remains to be demonstrated [167]. Additional disadvantages of the FTICR include its extremely high upfront and maintenance cost due to the requirement of cryogenic cooling for its magnets [168].

An alternative to FTICR is the Orbitrap, another Fourier transform MS, which provides comparable resolution and accuracy. In contrast to the FTICR, in the Orbitrap, ions are trapped only using an electrostatic field, rather than a magnetic field [169]. The motion of ions differs from the FTICR in that they oscillate along and rotate around the central electrode. Due to the electrostatic field, the kinetic energy of the ions within an orbitrap is typically greater because it is dependent on the force exerted by the electric field. Conversely, in an FTICR the kinetic energy decreases as $(m/z)^{-1}$, allowing for longer acquisition times and hence, greater resolution [170–172]. Consequently, Orbitraps experience a faster signal decay as a function of increasing MW due to greater intermolecular collisions, limiting its routine analysis of intact proteins to those under 30 kDa. However, by modifying the Orbitrap, researchers have attempted to improve its capabilities in niche circumstances [170,171]. For instance, it has been demonstrated that trapping ions in the higher energy collisional dissociation cell and replacing the nitrogen environment with helium can reduce the number of ion collisions, resulting in greater resolution and the mass determination of a 148.7 kDa IgG1 antibody [173]. While interesting, most such analyses of intact high MW molecules have been conducted under ideal conditions, using highly purified, commercially supplied proteins, rather than extracted native proteomes which carry a significant degree of chemical noise.

An interesting recent development is individual ion MS (I²MS), which demonstrates great potential for the characterization of complex proteoform mixtures. In I²MS, each individual ion is analyzed independently, and the image current induced is plotted as a linear function of acquisition time, the slope of which is proportional to its charge. Using this function, each ion is assigned a charge which is used with m/z to produce a true mass spectrum [144,174–176]. In a recent study using I²MS, 550 proteoforms were identified, along with a group of unidentified proteoforms resolved between 20–25 kDa, from HEK293T cell lysate fractionated by GELFrEE [176]. Furthermore, two engineered virus-like particles (VLPs) produced in *E. coli* and carrying DNA and RNA were analyzed using I²MS yielding high-resolution mass distributions with masses of 990 ± 16 kDa and 3190 ± 38 kDa [144,176]. Nonetheless, while quite promising, I²MS is also not without limitations. Given that multiple ions are analyzed in one acquisition, more than one ion may produce the same frequency corresponding to the same m/z , which would lead to false charge assignments. Furthermore, while 550 proteoforms were identified, it was not directly reported how many canonical proteins this represented. Additionally, the variability in the linear function used to assign charges decreases proportionally to (ion survival time)^{1/2} which, as discussed above, is a limitation of the Orbitrap. In an effort to rectify this, voltage has been decreased to lower the kinetic energy of the ions, however, the proteins analyzed were either small enough (<30 kDa) or too large (0.99 and 3.19 MDa) to be significantly impacted by inter-molecular collisions [176].

As the MS-intensive approach can currently analyze primarily a lower MW sub-set of species, it is not yet capable of large-scale investigations of ‘complete’ proteomes [18,143–145,147,148]. In saying that, this method can comprehensively characterize proteoforms, strongly suggesting that MS-intensive top-down will be an extremely powerful tool in the future, when current technical limitations are overcome, enabling it to be more broadly applicable, routine, and far more cost-effective. It would be ideal to pair the current MS-intensive and integrative methods, although this will not be possible without further refining *both* methods. Clearly, this is an extremely critical area for development.

4.2.4. So, What Does Top-Down Really Mean?

The development of proteomics clearly began with a top-down approach, 2DE. Until the advent of the MS-intensive approach, 2DE was essentially the only method to resolve intact proteoforms and ensure the best subsequent analyses to identify species [29,104,177]: it arguably still is. The evolution of this method has resulted in some debate in the field as to which analytical approach is truly ‘top-down’. Since proteomics is analytical chemistry and both methods first resolve intact proteoforms, then, regardless of how proteoforms are eventually identified (i.e., fragmented before or after entering the mass spectrometer, or via Western blotting), they are both top-down methods.

To put this into perspective, when 2DE ‘was’ proteomics, and Western blotting was the primary means to identify specific proteins (along with occasional analyses by Edman degradation), it was the recognized top-down analysis. As analytical instrumentation progressed, as it does, different forms of MS were progressively adopted to complement the resolution achieved with 2DE and to identify canonical protein sequences; doing so with an alternate, low resolution up-front gel electrophoresis method (e.g., SDS-PAGE/GELFrEE) does not change that approach. Should another, ‘better’ instrument/methodology arrive in the future for proteome analysis, would the MS-intensive approach then cease to be top-down?

Therefore, instrument-centric definitions are somewhat personal adoptions whereas general systematic approaches are what should be described by the terminology. Accordingly, the integrative and MS-intensive approaches, from an analytical chemistry perspective, both provide top-down analyses. We will not even consider the term middle-down, which is apparently used in some bottom-up and modified MS-intensive circles to describe a protocol that uses limited proteolysis/larger sizes of peptides to enable analyses. However, it has also been used to describe a number of different approaches in the field,

confusing matters to the point that its use becomes essentially meaningless [28,68,178,179]. Again, here is another area in which consensus would be useful and important, to move the field forward with a unified terminology and understanding of the genuine pros and cons of available analytical approaches [27–29,51].

4.3. Additional Analytical Variations on Peptide MS Analyses

Alternate analytical approaches that seek to enable quantitative proteome assessments include mainly variants of peptide MS such as iTRAQ. This multiplex approach seeks to identify changes in the abundance of canonical protein sequences, simultaneously, in up to eight biological samples, using isobaric tags to label the N termini and lysine side chains of peptides, ostensibly for either relative or absolute quantification [180–185]. This method is claimed to have high sensitivity and reproducibility, although it has recognized biases and underestimation issues [182,183,185]. Specifically, when two or more precursor ions with similar m/z and retention times are selected in the same fragmentation window and are sequenced and quantified together, both peptides contribute to one MS/MS signal. Thus, both are sequenced and quantified at the same time even though they are two separate species [183]. One of the crucial issues with labeling methods such as iTRAQ is that not all peptides have lysine side chains and, therefore, not all peptides will be labeled, thus yielding results that cannot be considered quantitative. Furthermore, any such labelling reactions, with reactive fluorescent groups or isotopes, will not be 100% efficient and hyperabundant species will dominate the reaction. Thus, any quantification must be thoroughly validated by orthogonal methods. Only stable isotope labeling by amino acids in cell culture (SILAC; Section 5.3.1) is likely to yield absolute quantification but is clearly not applicable to most sample types analyzed in proteomic studies (e.g., tissue and fluids); the exception may be *Drosophila* embryos fed labelled *S. cerevisiae* [186,187].

While peptide MS largely focuses on identifying amino acid sequences, it is possible to identify PTM, although it demands substantial additional work and time, and significant complications can arise. Foremost, these approaches generally require enrichment techniques—increasing the concentration of select proteins or peptides to improve their downstream analysis—and these selection techniques differ (including in quality and rigor) for every specific PTM that is to be analyzed [188–190]. Additional issues may arise in the MS analysis. For example, while identification is possible, phosphopeptides are not ionized and fragmented as efficiently as unmodified peptides making MS identification more difficult [190–192]; this is further complicated by the potential presence of ‘non-standard’ phosphorylation of histidine, arginine, and lysine. Another PTM, ubiquitination, results in a mass shift of 114.043 Da. Unfortunately, other events (e.g., a cleavage between a lysine-asparagine motif on other peptides) have been seen to cause an identical mass shift making it difficult to distinguish them from ubiquitination [190,193]. Different enrichment techniques also exist for other PTM (e.g., glycosylation, acetylation, methylation, and cysteine redox modifications), although a separate sample aliquot is needed for each identification and once used to identify one PTM, it cannot be re-used to identify another type of modification: clearly this may be an issue depending on amount of sample available. The question also arises as to which PTM should be analyzed, as each choice thus introduces added analytical bias while still not ensuring definitive identification of proteoforms since the starting material is a proteolytic digest of the proteome. Furthermore, as each PTM requires a separate analysis in addition to the original shotgun process, this essentially turns a high throughput discovery approach into a targeted search for specific alterations based on an assumption of what the important PTM might be. Again, it would seem impossible to do this systematically or exhaustively and thus, impossible to effectively analyze a native proteome (i.e., the full spectrum of constituent proteoforms). However, if there are specific PTM of interest (i.e., known or strongly suspected to be involved in a biological process of interest), then a targeted approach for specific species

can be used—with the usual caveat of the need for attention to weaknesses of the methods in order to capitalize best on the strengths.

5. Targeted Proteomics

A discovery approach can potentially generate 100s–1000s of hits, necessitating a targeted approach to validate the identified protein species if they are to later be assessed on a larger and/or more routine scale. Targeted proteomic workflows thus involve the detection of a canonical protein or proteoform of interest, hopefully with high sensitivity, selectivity, quantitative accuracy, and reproducibility [66,67]. This approach is essentially the bridge that connects discovery proteomics to the validation of biomarkers, potential targets for drug development, and other research efforts. Unfortunately, the quality of this bridge is heavily dependent on a number of factors and criteria [194,195]. There is a bottleneck as the techniques for targeted proteomic studies cannot keep up with the number of reasonably strong hits being made by discovery approaches. This has resulted in a seemingly perpetual gap in the identification of protein targets vs. the testing/validation of potential biomarkers and drug targets. It is most often seen that only a very small number of protein/proteoform candidates (most often those that show the most significant differences in abundance between two or more conditions) are further tested using a targeted approach as current methods can be extremely time consuming and expensive, obviating the capacity to separately test each potential candidate found in discovery studies [195,196]. Furthermore, quantitative rigor varies between methods and how they are applied, again emphasizing the need for consensus to address inter-lab variability or outright irreproducibility of findings. Thus, similar to discovery proteomics, there are numerous approaches used for targeted proteomic analyses although, again, each has its own strengths and weaknesses. Initially we will address those more common methods that utilize antibodies for detection of the target species.

5.1. Antibodies

Immunoassays are likely the most widely used method for targeted proteomics, utilizing antigen-antibody interactions for detection and quantification of the target species. Before discussing the numerous immunoassay methods that exist, it is important to delve into the advantages, as well as the limitations, of the main component used.

There are two main types of antibodies: monoclonal (mAb) and polyclonal (pAb). The former is monospecific, recognizing only a single epitope per antigen; pAb are heterogenous, each antibody component in the mix recognizes a different epitope on the same antigen. mAb are most useful for their strict specificity, and thus in evaluating changes in molecular conformation, protein–protein interactions, and PTM. However, this means that slight changes in the epitope (e.g., genetic polymorphism, untargeted PTM, and denaturation) can affect the binding of the mAb. This can be addressed by pooling multiple mAb of desired specificities, although this can be difficult, expensive, and time-consuming. Aside from epitope specificity, generally the biggest advantages of mAb, compared to pAb, is their high concentration, purity, and reproducibility, which arise from their capacity to be generated from a constant and renewable source. An additional antibody type that is easily sequenced and resynthesized are nanobodies. These monoclonal-like antibodies are devoid of light chains making them small, with high thermostability, superior solubility, and cost effectiveness [197–199]. Nonetheless, in some instances the monospecificity of mAb is considered an issue and pAb are preferred [200–202].

pAb are more stable over a broad range of pH and salt concentrations, often enabling their use under a variety of experimental conditions. Unfortunately, they are non-renewable, and their avidity is at risk of changing as they are harvested over time, and quantity of pAb obtained is limited by the size and lifespan of the host animal [19,200,202]. Additionally, pAb recognize multiple epitopes per antigen thus, if one or more of the clones in

the pAb mix recognizes a highly conserved protein ‘domain’ (e.g., a calcium binding domain)—which is found on numerous unrelated proteins/proteoforms—then it can lead to false identifications.

Overall, antibodies do not exist for all proteins, and vary widely in quality as most that are commercially available often seem to be subjected to limited validation [203–205]. Critically, there are even fewer proteoform-selective antibodies available. If an antibody does not perform as expected, then alternative antibodies with better performance or an antibody-independent approach must be considered which, unfortunately, is costly and time-consuming [19]. Thus, as with all methods, there are clearly pros and cons to be considered in using antibodies in proteomic analyses (although much can be achieved with rigorous controls and optimization; see below). To reduce interference commonly seen with full-size antibodies, Fab fragments can be used; removing the Fc fragment from the antibody provides for smaller binding components which improve binding capacity and can thus improve assay sensitivity [206,207].

Regarding identification of proteoforms, antibodies are generally raised to identify amino acid epitopes on canonical protein sequences; thus, a PTM (e.g., methyl, phosphate, and sugar group) at, or neighboring, the epitope will likely block binding of the antibody, thus preventing detection of the target or at least one or more related proteoforms (Figure 8) [208,209]. In contrast, while there exist antibodies that broadly recognize PTM (e.g., phosphotyrosine), these are generally of poor specificity and selectivity. For the antibody to effectively identify a specific proteoform, it must recognize both a specific PTM and the sequence surrounding it [202,208]. This can be challenging if the same PTM is present on more than one sequence of the amino acid backbone. Using two different antibodies—to the specific sequence and to the PTM—can often help address such detection issues. Notably, 2DE offers significant advantages in alleviating some of these issues since proteoforms have already been resolved prior to detection and thus a single antibody to an unmodified backbone epitope is often sufficient, provided all other control and optimizations steps have also been taken into account (see Western blotting, Section 5.2).

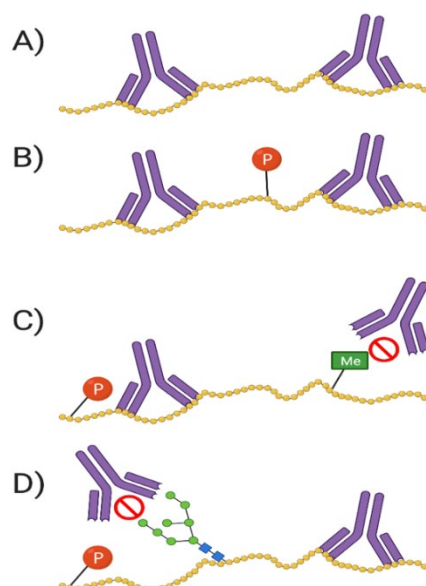


Figure 8. Antibodies and proteoforms. As antibodies are mainly raised to identify amino acid epitopes, it is possible that a PTM at, or near, the epitope will interfere with binding of the antibody. This may prevent the detection of the target. (A) Antibody binding without any interference; (B) Antibody binding without phosphate group interfering; (C) Antibody binding blocked by methyl group; (D) phosphate and sugar group adjacent to epitope affect/block antibody binding.

5.2. Immunoassays

Of all the targeted proteomic methods that exist, immuno- or Western blotting is most commonly used [210,211]. Western blots are effective for small-scale protein analyses and, at least in theory, are relatively simple and cost-effective, and produce data that is easily interpreted [19,212]. While the original goal of the Western blot was to provide a yes or no answer about the presence of a target protein, refinements over the last several decades have enabled sensitive and reproducible quantification of a given target in a native sample extract [64,65,213]; however, this can only be done when the complete sample is represented (i.e., no proteins/proteoforms are (non)specifically removed due to prior use of fractionation techniques) [214]. Unfortunately, Western blotting lacks the throughput to routinely quantify large fractions of a proteome and quantification depends heavily on the quality of the techniques and antibodies used [19,64,84,205]. There are many factors that can affect the reproducibility and quantitative capacity including large protein loads (i.e., signal saturation) and failure to optimize buffers, blocking reagents, or transfer conditions; the presence of lipids or carbohydrates can also interfere with resolution and detection [19,205]. Additionally, at the initial sample preparation stage, improper handling, poor homogenization, as well as inadequate detergents and protease inhibitors can lead to decreased reproducibility and sensitivity [215].

Regarding technique, transfer method/conditions, gel composition, and type of blotting membrane can all substantially affect the quantitative efficiency of protein transfer from the gel to the blotting membrane [19,216]. Two commonly used transfer methods are wet (complete immersion of a gel-membrane sandwich in buffer) or semi-dry (gel-membrane sandwich is placed between absorbent paper soaked in transfer buffer). Wet transfer has high efficiency but takes more time whereas semi-dry is convenient and saves time, although this is often at the expense of transfer quality; specifically, higher MW proteins may not transfer as effectively as they do with a wet transfer [19,216,217]. Notably, a more complete and higher quality transfer is seen when thinner gels are used but there is the added risk of gel cracking/ripping during handling [19].

Transfer conditions (i.e., current, voltage, and buffer) also play key roles in transfer efficiency. Similar to the concept of protein separation with SDS-PAGE, low MW species generally transfer faster than those of higher MW. Therefore, under conditions optimized for transfer of lower MW species, larger proteins have low transfer efficiency and under conditions optimized for high MW proteins, those of lower MW can be driven completely through the blotting membrane [64,216]. This can be avoided with vacuum-assisted solvent flow (transfer of proteins from gel to membrane using suction power); however, as stated in Section 4.2.2, this is difficult with acrylamide gels and if transfer is extended longer than 45 min, the gel is at risk of drying out [216,218].

Furthermore, some PTM (e.g., glycosylation) can markedly affect transfer efficiency, so conditions must always be optimized for species of interest [216,219,220]. The practice of diluting the primary antibody to only detect a single band has been used by some as a proxy for appropriate optimization steps but can prove to be quite misleading. Additional bands may be indicative of notable proteoforms and/or modification (e.g., cleavage) due to suboptimal handling rather than be 'spurious' (e.g., the result of nonspecific antibody binding) [19]. The presence of proteoforms and aggregates/multimers in different bands can be easily confirmed via MS/MS, ensuring the most thorough and quantitative Western blot analyses. Thus, a well validated primary antibody is critical to success, as is a high-resolution separation of species present in the sample; this also emphasizes the risk of interpreting dot blots beyond simply indicating that the antibody being used is binding to something in the sample (which may or may not include your species of interest).

Overall, while it is relatively simple to detect proteins blotted from gels, doing so reproducibly and with quantitative rigor requires more effort than is commonly seen in the literature. Simply, one size does not fit all, and the common assumption that the 'standard' protocol used in a given lab will work effectively for every sample or species is false. Select optimizations and standards for quantitative calibration may well be needed

for each distinct species under investigation. Routine controls should always include (i) ensuring uniform total protein loads per sample (rather than relying on the misleading and outdated use of ‘housekeeping’ proteins for normalization) and (ii) quantitative assessment of transfer efficiency using the highest sensitivity in-gel detection available. This will minimally enable quantification relative to the control samples.

Regarding detection methods, the two most known are chemiluminescence and fluorescence. Colorimetric methods do exist but their performance varies based on purity of substrate and buffer components [221]. Chemiluminescence is used most widely and involves an enzyme-substrate reaction to generate light. The emitted light then decays to ground state and the signal fades quickly, often within a very brief time [221,222]. This degradation is quite disadvantageous as the signal is not consistent, and once it is terminated, the ability to retrieve any additional quantitative data in the future becomes more difficult. An improved method utilizes chemifluorescence, which yields stronger and more stable signals, enabling more sensitive detection via signal integration over time [64,65,221,223]. Additionally, antibodies can be tagged with fluorophores, some detectable in the infrared spectrum. NIR reduces the risk of and thus interference by autofluorescence (natural emission of light) [223]. Since the excitation/emission capacity of these fluorophores does not diminish over time, they enable extended signal integration and thus greater sensitivity of detection. Furthermore, using different fluorophores enables multiplex Western blot assays [224].

Another commonly used method for targeted protein detection is an enzyme-linked immunosorbent assay (ELISA). This method is similar to Western blotting in that it requires an antibody to identify the target. However, ELISA differs as its visual readout is in a 96-well plate as opposed to a blotted SDS-PAGE gel and thus there is no resolution of species (e.g., always leading to the same potential concerns as with a dot blot). The type of catalytic label used can produce different methods of visual detection similar to that of Western blotting (i.e., colorimetric, chemiluminescent, and fluorescent) [225,226]. While ELISAs are often thought of as reasonably fast with a high degree of sensitivity, specificity, and reproducibility, they unfortunately share a critical limitation with Western blots: whether or not the technique is optimized to achieve maximal assay quality, the method is largely dependent on the quality of the antibodies used [194,226]. Unfortunately, these are largely considered proprietary by the commercial firms producing most of the ELISA kits used in current research and, certainly in our experience when purchasing kits, it is impossible to also get samples of the antibodies to verify their target specificity in the samples being analyzed. Like dot-blot, one is assuming that only the target of interest is contributing to the signal, with the understanding that the signal likely represents a host of proteoforms.

Similar to ELISAs are single molecule array assays (SiMoA). This method uses paramagnetic beads that are coupled with biotinylated detection antibodies, streptavidin-labeled enzyme, and the target molecules (proteins) to form immunocomplexes. These beads are then added to a microwell array and a non-fluorescent resorufin- β -D-galactopyranoside (RDG) is added. RDG is converted to a fluorescent product when it reacts with the streptavidin-labeled enzyme. The wells in the plates used for SiMoA allow one type of labeled bead per well (i.e., targeting at a single canonical protein); this ensures that the ratio of active beads to number of beads located in the wells is directly correlated to the concentration of the target protein in the sample being examined [227–230]. The benefit of this assay format is increased sensitivity and ability to detect sub-femtomolar concentrations of the target [230,231] although, it is quite expensive and has similar disadvantages to an ELISA, in that output only represents total abundance of a canonical protein.

Although quite different from the other methods discussed, another antibody-based approach of note is immunohistochemistry (IHC), which is used to detect target in fixed cells and tissue slices [232,233]. This is particularly useful for localizing target species of interest, although quantification is limited to relative comparisons between samples. Nonetheless, much work has been done to improve quantitative assessments. In typical

IHC staining, active immunocomplexes produce a broad colorimetric response however, it makes quantitative observation difficult [234]. An alternative method, quantitative IHC, uses an additional enzymatic amplification which converts the antibody/antigen complexes into defined dots, allowing for counting. Prior to labelling, a pre-determined fraction of secondary antibodies is labeled and thus the ratio between number of labeled sites and number of labeled antibodies can be used as a direct correlation [235]. IHC differs from Western blots, ELISAs, and SiMoAs as it allows visualization of cellular components, can provide morphological information, and thus determines target localization within a cell or tissue. Unfortunately, the IHC protocol is somewhat long and detailed, and thus throughput can be quite limited. Furthermore, as all of these methods use antibodies there is still the risk of non-specific interactions as well as the many limitations discussed earlier. While these methods have been invaluable, they clearly also have limitations. To some extent, recent rapid developments in MS provide some help or alternate approaches in this regard.

5.3. Mass Spectrometry

5.3.1. Label-Based

As an alternative to immunoassay approaches, MS can be used to identify canonical proteins in a targeted approach either with a label-based or label-free method (these are generally discovery-based methods but can be modified for targeted use). The former involves labelling the peptides in a sample digest, prior to shotgun analysis, using different reagents that are chemically identical but differ in their isotopic composition [66,236]. The peptides can undergo chemical, metabolic, or enzymatic labeling—each again with distinct advantages and disadvantages [66,236,237].

The most popular method, metabolic labeling (i.e., SILAC), involves the addition of a stable isotope label to growth media, enabling its incorporation into metabolically active cells [238–240]. As this method does not target functional groups and the isotopic label is introduced prior to protein extractions, each newly synthesized protein is labelled efficiently compared to other labelling methods [239]. While the only truly absolute quantitative method, unfortunately, it is limited to cell cultures and not applicable to the vast bulk of proteomic studies that focus on tissues or biofluids [241]. Furthermore, SILAC introduces an additional shift in the isotopic envelope of the mass spectrum that may result in peak overlap which, coupled with the decreased likelihood of complete labelling as a function of increasing protein size, further complicates data analysis of larger proteins [242].

Chemical labeling (isotope-coded affinity tagging (ICAT), isotope-coded protein labeling (ICPL), tandem mass tags (TMT), and iTRAQ [Section 4.3]) introduce isotopic or isobaric labels at the protein/peptide level following protein extraction and can be used to tag numerous different types of samples [238,240,243–245]. Since the labels are introduced following protein extraction, they must target certain functional groups. ICAT favors cysteine-containing proteins, TMT favors NHS ester-based reactive groups, and ICPL and iTRAQ favor primary amines (i.e., N-terminus and lysine side chains). This is a drawback as not all proteins contain cysteine and/or lysine; thus, unlabeled peptides cannot be used for quantification [239,240]. Furthermore, these methods require additional steps for labelling and protein/peptide recovery in the analytical workflow, and the commercially available labelling reagents are expensive [246].

Enzymatic labeling involves the addition or removal of water using ^{18}O tracers which results in a 4 Da mass shift between the same peptides in two different samples being compared. The main advantage of this method is that it does not target specific amino acids or require added enrichment steps. Additionally, this method is simple, low-cost, and applicable to all types of samples; its main limitations are seen to be incomplete labelling and the lack of capacity for multiplexing [239–241].

5.3.2. Label-Free

In contrast, the label-free method is straightforward, cost-effective, and requires minimal sample manipulation [75,241,247,248], measuring peptides by ion intensity or spectral counting, and is often used to compare two or more conditions (Figure 9) [237,247,249]. Label-free uses the acquired spectra of a given peptide as a proxy for the relative amount of the corresponding canonical protein in a given sample. Again, as a shotgun approach, this assumes all peptides are from intact canonical proteins and may or may not work well with different related proteoforms depending on the nature of the modifications. Thus, this method also tends to lack accuracy, precision, and reproducibility as it does not compare peptides to an internal, chemically identical standard enriched with a stable isotope [75,177,237,250]; that said, this is an area of ongoing refinement [251]. Thus, most simply, label-free provides deeper coverage of the canonical proteome, while label-based methods in theory provide better quantification [177]. Integration with 2DE for front-end proteoform resolution would likely enhance the specificity and quantitative rigor of both approaches [29].

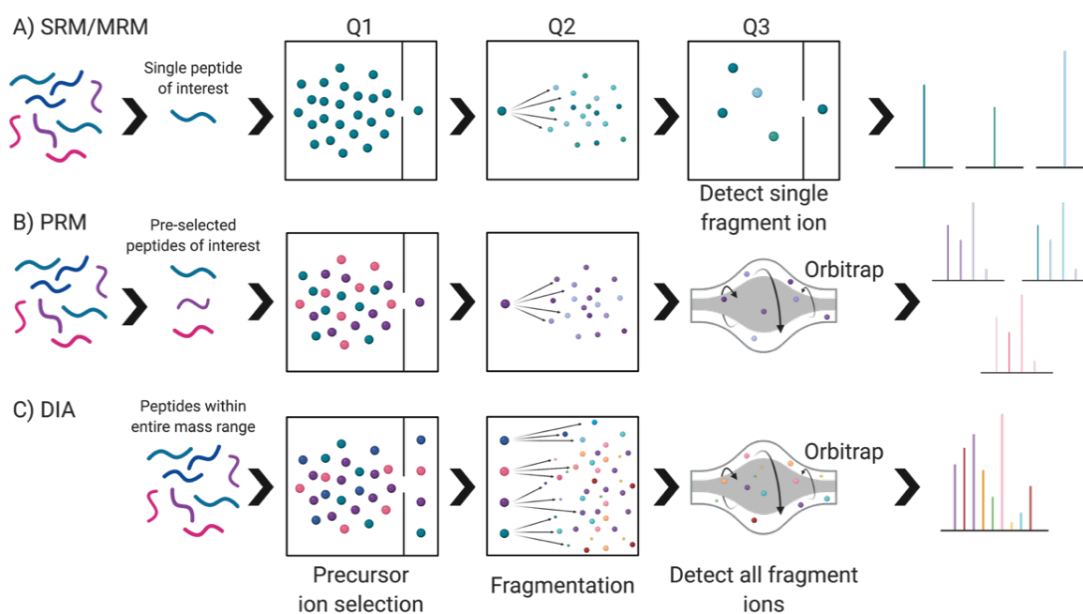


Figure 9. MS-based targeted proteomics. Shown are the different acquisition modes commonly used for targeted detection of protein species with MS. (A) SRM—quantifies specific, predetermined ions from peptide of interest; (B) PRM—simultaneously analyzes all fragment ions of the pre-selected peptides of interest; (C) DIA—analyzes all peptide mass ranges within the window without pre-selection.

Currently, one of the most used methods for quantification of a given canonical protein by MS is Selective Reaction Monitoring (SRM, and the related Multiple Reaction Monitoring [MRM]) [21,79]. SRM is carried out using a triple quadrupole mass spectrometer, sequentially monitoring for fragment ions from the same peptide between two levels of selection and isolating predefined precursor and fragment ions [67,89]. As this method requires that an assay be developed for each target protein and quantifies specific, predetermined ions, it is somewhat similar to Western blotting, although SRM seems to be somewhat superior with regard to data quality and performance characteristics, provided consistent instrumentation and protocols are used [252]. However, although SRM provides high sensitivity and quantitative accuracy (in terms of theoretically intact canonical proteins), it lacks throughput compared to other methods (i.e., DIA and parallel reaction monitoring [PRM]) as it sequentially samples only one fragment ion at a time [67,253].

In contrast, PRM produces full MS/MS spectra for each precursor and simultaneously analyzes all fragment ions of the pre-selected peptides of interest; SRM only monitors the predefined product ions [67,89,254]. This is advantageous as it provides the flexibility to

select fragment ions following data acquisition [67]. Thus, PRM provides higher selectivity, dynamic range, and S/N compared to SRM. Although seemingly not used in any effort to assess specific proteoforms, both SRM and PRM have high sensitivity, specificity, and reproducibility, but they lack multiplexing capabilities. To overcome this limitation, DIA-based targeted quantification may be used.

DIA (Section 4.1) is primarily a discovery-based approach but can also be used as a targeted method. This approach is aimed at utilizing the full capabilities of mass spectrometers to maximize MS acquisition time and to address the need to expand the detectable dynamic range, lower the limit of detection, and improve on the overall confidence of peptide identifications and relative quantification measurements. However, with this multiplexing ability, it is resource demanding and has somewhat decreased sensitivity, specificity, and reproducibility [89].

Like all methods, each of these three label-free methods has its own unique advantages and disadvantages, with some capability sacrificed in each method. Most notable perhaps is their essentially exclusive use in proteogenomic analyses rather than addressing the need to assess proteoforms. Overall, the ‘ideal’ method would provide high throughput, multiplexing capabilities, and high sensitivity, specificity, and reproducibility, and retain these qualities even for the analysis of specific, even closely related proteoforms (i.e., be able to differentiate between them, which is likely impossible at the peptide level). While the instrumentation has seen a steady series of significant improvements over the last two decades, technical limitations may well stand in the way of further major advances. Regardless, any such refinements will also come with a hefty price. For now, we can either wait for this ideal instrument and associated methodologies that will provide thorough MS-intensive analyses of intact proteoforms and thus full proteomes or use the best available technology to address pressing research concerns in medicine, environmental, agricultural, and other areas. This is by no means a suggestion that technological refinement and optimization should not continue, for *all* available approaches to proteome analysis. Rather, it is a comment on better capitalizing on what is available, in parallel with improvements to instrumentation.

6. What Next?

Proteomes are of proteoforms, not canonical proteins. There is no one-size-fits-all method for every type of proteoform or proteomic study—discovery or targeted. We hopefully adapt to capitalize on the strengths and limit the impact of the weaknesses in each case. Currently, no single approach is close to reaching the goal of identifying and characterizing *all* proteoforms in a proteome; for that matter, we cannot even guarantee full proteome extraction from any given sample. Does this mean we cannot move proteomic research forward without a significant disruptive change? While such sudden advancements can revolutionize a field, they are not the only means of carrying out the best possible proteome analyses. Thus, stringent optimization of and consensus on available methods and criteria for data quality, would be the most obvious and straightforward approach to effectively addressing the inherent complexity of proteomes. Indeed, openly accepting the complexity, by consistently addressing proteoforms as the critical species that must be resolved and identified, would seem the first step. Simply, continuing a purely proteo-genomic/-transcriptomic approach to cataloging amino acid sequences will add a few more entries to databases, but will not effectively address proteome complexity or thus enhance our understanding of molecular mechanisms or identify selective therapeutic targets and biomarkers. We need to integrate and capitalize on what works best—*but use it better*—and put effort into consistently and critically improving approaches to proteoform resolution and assessment to reach the goal of routine, full proteome analyses. Obviously, it is unrealistic to expect such routine, deep proteome analyses within even the next decade considering it took concerted international effort coordinated by HUPO over >10 years to reach the current milestone of 90.4% coverage of the conservatively defined genome at the level of amino acid sequences (i.e., Proteome-lite) [70,71]. Since things

are more complex than originally defined, we need to have a firm grasp on the genome if we hope to effectively define a proteome with myriad potential proteoforms that also vary temporally [56]. Importantly, this recent HUPO announcement [70,71] is a new starting point, and a new opportunity for the next generations of researchers to take a bold new perspective on what needs to be done and thus how best to rigorously address the complexity of proteomes.

With these caveats in mind, in terms of a complementary and broadly applicable approach, 2DE/3DE (with a host of well-established modifications and variations to enhance resolution and detection sensitivity/selectivity) coupled with LC/MS/MS (also with established variations to optimize both sample and data analyses) seems to be the only current approach that can effectively resolve, identify, and quantify the largest number of *proteoforms* in a given sample. Essentially this is about escaping technique- or technology-centric biases and integrating existing top-down and bottom-up approaches: capitalizing on existing strengths and minimizing different technical limitations in the process. In saying this, it is also important to note that even though they have undergone numerous enhancements, the core methods can always be further improved. Many critical questions still need to be addressed. Can we further improve protein extraction with alternative detergents and methods? Can we further improve resolution in IEF? Can we further automate any steps within the full analytical process? These are but a few examples deserving of more robust consideration rather than continuing to do things ‘because that is how they have always been done and others do it similarly as well’. If the latter common rationale were actually true—that is, things are *not* in fact done identically between labs, even apparently using the same protocol and instrumentation—there would not exist the current level of inter-lab variation and irreproducibility of findings [255]. The real issue seems to be whether the field seeks to continue its now almost blind commitment to the speed of analyses or whether quality, depth, and quantification in proteoform analysis become recognized as the critical objectives, as will be absolutely necessary with respect to proteome complexity. It is thus time to look forward and fully embrace the genuine complexity of proteomes, and what it will mean to effectively analyze them, particularly in any sort of routine manner. This will require an innovative and transdisciplinary mindset to create an integrative, state-of-the-art *proteomics* that (re)defines the discipline. This is the next generation approach that will future-proof the field and enable nimble integration of advances in both sample and data analysis.

Thus, while MS-intensive analysis now seems to be the most promising long-term venture in top-down proteomics, it currently cannot deliver the depth of proteome coverage provided by the integrative approach. Strangely, it seems a genuine interface of the two approaches (i.e., 2DE and MS-intensive) has never even been tested, although some have reported limited success with proteoform identification following passive elution from SDS-PAGE gels [155], certainly suggesting that this could be quite informative. Furthermore, as we work toward this goal of mapping the proteome [70], all available approaches/methodologies/subdisciplines need to work in unison. As but one example, with the current refined state of 2DE, every spot in the gel essentially becomes a mini bottom-up experiment as 10s–100s of proteoforms (or more) can be found in a single spot [98,104,256]. Thus, the best such analyses depend not only on the highest quality 2D/3D gels, but also the highest resolution MS coupled with the best possible data analysis/database interrogations to ensure solid proteoform identifications and, hopefully, deeper analyses to fully characterize the inherent PTM. What if each such spot could be quantitatively eluted and the intact proteoforms fully assessed using ongoing refinements in MS-intensive analyses? Given the current status of technology in the field, it is actually somewhat surprising that the major MS instrumentation firms—that do much to define and manage the direction of current proteomic analyses—do not offer a refined front-end 2DE suite to complement their latest LC/MS instrumentation packages. That would define a very real

interest in deep proteome analysis rather than a continuing primary focus on proteogenomics, which will simply not effectively address the complexity of proteomes. Analytical rigor in addressing proteoforms is needed, not just new instruments.

Unfortunately, there has come to be an almost wholesale emphasis on the speed of analysis—we must analyze proteomes at the same rate we analyze genomes—yet apparently ignore all the issues that have arisen and are still appearing with those approaches despite the astounding technologies that have been developed to address genome sequencing demands. In this regard, rather than almost exclusively focusing on ‘fast’ analyses of canonical protein sequences, it seems likely that funding agencies are also expecting to see more analytical depth from investments already made for instrumentation. Realistically however, there is currently no panacea to either deep genome or deep proteome analysis. Yet we continue to claim otherwise rather than appreciating the inherent complexities of the systems—which is where the actual answers to our most critical research questions lie. Those who are seemingly convinced that we will sort this all out ‘tomorrow’ undoubtedly also have a new technology they want to sell you. Noting these issues and that the proteome is so much more complex than the genome, we perhaps need to come to yet another firm consensus as a field by asking a critical question: what is more important, speed or the actual quality of the analysis and resulting data? Then, can those data be turned into knowledge?

We need to take the time to refine and optimize these methods, and present trials/papers exactly as such, rather than claiming they are full solutions and ready to provide a ‘breakthrough’. Some of this may be attributed to publication hype by some journals, but rather than moving the field forward, it is muddying the waters. To achieve any type of Systems Biology, we need to stop ignoring the flaws in our respective approaches/methods. We are at a point where each one of these methods is like a separate piece of a car. If we try to drive with just one piece (i.e., engine) we will not get anywhere, even if the part is in great working condition. To get from point A to point B, we need to bring all the pieces together to make an integrated, functional unit.

If nothing else, there is an overwhelming need to recognize complementarity. Thus, we need to move away from the idea of operating in an intradisciplinary manner and push for a sense of transdisciplinarity within the field of proteomics. As we have shown here, all available methods have strengths and weaknesses. However, if we are to work together and take the best qualities of these methods and integrate them, we could well revolutionize the field by fully driving the necessary analyses beyond the current general confines of proteogenomics. Until this is done, we will be unable to advance our methods or thus deep and routine proteomic analyses to the extent that is both necessary and sufficient for unbiased identification of genuine highly selective biomarkers and drug targets. That said, there are indeed critical ongoing efforts to improve both sample and data analyses, and the best of these must be integrated into a continuously developing unified approach to proteoform and thus proteome analysis [29,81,88,257–274]. We are hopeful and confident of a more collaborative and unbiased future for the discipline of proteomics.

For a successful technology, reality must take precedence over public relations, for Nature cannot be fooled.

Richard P. Feynman

Author Contributions: Conceptualization, K.C. and J.R.C.; writing—original draft preparation, K.C.; writing—additional information, review, and editing, K.C., M.A. and J.R.C.; visualization, K.C.; supervision, J.R.C. All authors have read and agreed to the published version of the manuscript.

Funding: JRC acknowledges the support of the Natural Sciences and Engineering Research Council of Canada (NSERC grant number 2019-04324).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Acknowledgments: K.C. acknowledges BioRender.com for creation of Figures 1–9.

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

References

1. Fey, S.J.; Larsen, P.M. 2D or not 2D. *Curr. Opin. Chem. Biol.* **2001**, *5*, 26–33, doi:10.1016/S1367-5931(00)00167-8.
2. Wilkins, M.R.; Sanchez, J.-C.; Gooley, A.A.; Appel, R.D.; Humphrey-Smith, I.; Hochstrasser, D.F.; Williams, K.L. Progress with proteome projects: Why all proteins expressed by a genome should be identified and how to do it. *Biotechnol. Genet. Eng. Rev.* **1996**, *13*, 19–50, doi:10.1080/02648725.1996.10647923.
3. Duncan, M.W.; Yergey, A.L.; Gale, P.J.; Kate, Y. Quantifying proteins by mass spectrometry. *LC-GC N. Am.* **2014**, *32*, 726–735.
4. Jungblut, P.R.; Holzhütter, H.; Apweiler, R.; Schlüter, H. The speciation of the proteome. *Chem. Cent. J.* **2008**, *2*, 16, doi:10.1186/1752-153X-2-16.
5. Jungblut, P.R.; Thiede, B.; Schlüter, H. Towards deciphering proteomes via the proteoform, protein speciation, moonlighting and protein code concepts. *J. Proteom.* **2016**, *134*, 1–4, doi:10.1016/j.jprot.2016.01.012.
6. Timp, W.; Timp, G. Beyond mass spectrometry, the next step in proteomics. *Sci. Adv.* **2020**, *6*, eaax8978, doi:10.1126/sciadv.aax8978.
7. Vanderperre, B.; Lucier, J.-F.; Bissonnette, C.; Motard, J.; Tremblay, G.; Vanderperre, S.; Wisztorski, M.; Salzet, M.; Boisvert, F.-M.; Roucou, X. Direct detection of alternative open reading frames translation products in human significantly expands the proteome. *PLoS ONE* **2013**, *8*, e70698, doi:10.1371/journal.pone.0070698.
8. Delcourt, V.; Brunelle, M.; Roy, A.V.; Jacques, J.-F.; Salzet, M.; Fournier, I.; Roucou, X. The protein coded by a short open reading frame, not by the annotated coding sequence, is the main gene product of the dual-coding gene MIEF1. *Mol. Cell. Proteom.* **2018**, *17*, 2402–2411, doi:10.1074/mcp.RA118.000593.
9. Brunet, M.A.; Brunelle, M.; Lucier, J.-F.; Delcourt, V.; Levesque, M.; Grenier, F.; Samandi, S.; Leblanc, S.; Aguilar, J.-D.; Dufour, P.; et al. OpenProt: A more comprehensive guide to explore eukaryotic coding potential and proteomes. *Nucleic Acids Res.* **2019**, *47*, D403–D410, doi:10.1093/nar/gky936.
10. Randles, L.G.; Dawes, G.J.S.; Wensley, B.G.; Steward, A.; Nickson, A.A.; Clarke, J. Understanding pathogenic single-nucleotide polymorphisms in multidomain proteins—studies of isolated domains are not enough. *FEBS J.* **2013**, *280*, 1018–1027, doi:10.1111/febs.12094.
11. Robert, F.; Pelletier, J. Exploring the impact of single-nucleotide polymorphisms on translation. *Front. Genet.* **2018**, *9*, 507–507, doi:10.3389/fgene.2018.00507.
12. Chen, S.H.; Habib, G.; Yang, C.Y.; Gu, Z.W.; Lee, B.R.; Weng, S.A.; Silberman, S.R.; Cai, S.J.; Deslypere, J.P.; Rosseneu, M. Apolipoprotein B-48 is the product of a messenger RNA with an organ-specific in-frame stop codon. *Science* **1987**, *238*, 363–366, doi:10.1126/science.3659919.
13. Hospattankar, A.V.; Higuchi, K.; Law, S.W.; Meglin, N.; Brewer, H.B. Identification of a novel in-frame translational stop codon in human intestine ApoB mRNA. *Biochem. Biophys. Res. Commun.* **1987**, *148*, 279–285, doi:10.1016/0006-291X(87)91107-7.
14. Powell, L.M.; Wallis, S.C.; Pease, R.J.; Edwards, Y.H.; Knott, T.J.; Scott, J. A novel form of tissue-specific RNA processing produces apolipoprotein-B48 in intestine. *Cell* **1987**, *50*, 831–840, doi:10.1016/0092-8674(87)90510-1.
15. Kim, D.D.Y.; Kim, T.T.Y.; Walsh, T.; Kobayashi, Y.; Matise, T.C.; Buyske, S.; Gabriel, A. Widespread RNA editing of embedded alu elements in the human transcriptome. *Genome Res.* **2004**, *14*, 1719–1725, doi:10.1101/gr.2855504.
16. Nicholas, A.; de Magalhães, J.P.; Kravtsov, Y.; Richfield, E.K.; Levanon, E.Y.; Khrapko, K. Age-related gene-specific changes of A-to-I mRNA editing in the human brain. *Mech. Ageing Dev.* **2010**, *131*, 445–447, doi:10.1016/j.mad.2010.06.001.
17. Knoop, V. When you can't trust the DNA: RNA editing changes transcript sequences. *Cell. Mol. Life Sci.* **2011**, *68*, 567–586, doi:10.1007/s00018-010-0538-9.
18. Schlüter, H.; Apweiler, R.; Holzhütter, H.-G.; Jungblut, P.R. Finding one's way in proteomics: A protein species nomenclature. *Chem. Cent. J.* **2009**, *3*, 11, doi:10.1186/1752-153X-3-11.
19. Gorr, T.A.; Vogel, J. Western blotting revisited: Critical perusal of underappreciated technical issues. *Proteom. Clin. Appl.* **2015**, *9*, 396–405, doi:10.1002/prca.201400118.
20. Issaq, H.J.; Veenstra, T.D. The role of electrophoresis in disease biomarker discovery. *Electrophoresis* **2007**, *28*, 1980–1988, doi:10.1002/elps.200600834.
21. Anjo, S.I.; Santa, C.; Manadas, B. SWATH-MS as a tool for biomarker discovery: From basic research to clinical applications. *Proteomics* **2017**, *17*, 1600278, doi:10.1002/pmic.201600278.
22. Xu, H.; Wang, Y.; Lin, S.; Deng, W.; Peng, D.; Cui, Q.; Xue, Y. PTMD: A database of human disease-associated post-translational modifications. *Genom. Proteom. Bioinform.* **2018**, *16*, 244–251, doi:10.1016/j.gpb.2018.06.004.
23. Aebersold, R.; Agar, J.N.; Amster, I.J.; Baker, M.S.; Bertozzi, C.R.; Boja, E.S.; Costello, C.E.; Cravatt, B.F.; Fenselau, C.; Garcia, B.A.; et al. How many human proteoforms are there? *Nat. Chem. Biol.* **2018**, *14*, 206–214, doi:10.1038/nchembio.2576.
24. UNIMOD: Protein Identifications for Mass Spectrometry. Available online: http://www.unimod.org/modifications_list.php? (accessed on 1 July 2021).

25. Chapman, J.D.; Goodlett, D.R.; Masselon, C.D. Multiplexed and data-independent tandem mass spectrometry for global proteome profiling. *Mass Spectrom. Rev.* **2013**, *33*, 452–470, doi:10.1002/mas.21400.
26. Smith, L.M.; Kelleher, N.L. Proteoform: A single term describing protein complexity. *Nat. Methods* **2013**, *10*, 186–187, doi:10.1038/nmeth.2369.
27. Oliveira, B.M.; Coorsen, J.R.; Martins-de-Souza, D. 2DE: The phoenix of proteomics. *J. Proteom.* **2014**, *104*, 140–150, doi:10.1016/j.jprot.2014.03.035.
28. Coorsen, J.R.; Yergey, A. Proteomics is analytical chemistry: Fitness-for-purpose in the application of top-down and bottom-up analyses. *Proteomes* **2015**, *3*, 440–453, doi:10.3390/proteomes3040440.
29. Zhan, X.; Li, B.; Zhan, X.; Schlüter, H.; Jungblut, P.R.; Coorsen, J.R. Innovating the concept and practice of two-dimensional gel electrophoresis in the analysis of proteomes at the proteoform level. *Proteomes* **2019**, *7*, 36, doi:10.3390/proteomes7040036.
30. Jeffery, C.J. Moonlighting proteins. *Trends Biochem. Sci.* **1999**, *24*, 8–11, doi:10.1016/S0968-0004(98)01335-8.
31. Jeffery, C.J. Why study moonlighting proteins? *Front. Genet.* **2015**, *6*, 211–211, doi:10.3389/fgene.2015.00211.
32. Jeffery, C.J. Protein species and moonlighting proteins: Very small changes in a protein's covalent structure can change its biochemical function. *J. Proteom.* **2016**, *134*, 19–24, doi:10.1016/j.jprot.2015.10.003.
33. Jeffery, C.J. Protein moonlighting: What is it, and why is it important? *Philos. Transactions. Biol. Sci.* **2018**, *373*, 20160523, doi:10.1098/rstb.2016.0523.
34. Jaffe, E.K.; Lawrence, S.H. The morpheein model of allostery: Evaluating proteins as potential morpheeins. *Methods Mol. Biol.* **2012**, *796*, 217–231, doi:10.1007/978-1-61779-334-9_12.
35. Jaffe, E.K. Wrangling shape-shifting morpheeins to tackle disease and approach drug discovery. *Front. Mol. Biosci.* **2020**, *7*, 582966, doi:10.3389/fmolb.2020.582966.
36. Chen, J.; Brunner, A.-D.; Cogan, J.Z.; Nuñez, J.K.; Fields, A.P.; Adamson, B.; Itzhak, D.N.; Li, J.Y.; Mann, M.; Leonetti, M.D.; et al. Pervasive functional translation of noncanonical human open reading frames. *Science* **2020**, *367*, 1140–1146, doi:10.1126/science.aay0262.
37. Frith, M.C.; Forrest, A.R.; Nourbakhsh, E.; Pang, K.C.; Kai, C.; Kawai, J.; Carninci, P.; Hayashizaki, Y.; Bailey, T.L.; Grimmond, S.M. The abundance of short proteins in the mammalian proteome. *PLoS Genet.* **2006**, *2*, e52, doi:10.1371/journal.pgen.0020052.
38. Eguen, T.; Straub, D.; Graeff, M.; Wenkel, S. MicroProteins: Small size—Big impact. *Trends Plant Sci.* **2015**, *20*, 477–482, doi:10.1016/j.tplants.2015.05.011.
39. Cao, X.; Khitun, A.; Na, Z.; Dumitrescu, D.G.; Kubica, M.; Olatunji, E.; Slavoff, S.A. Comparative proteomic profiling of unannotated microproteins and alternative proteins in human cell lines. *J. Proteome Res.* **2020**, *19*, 3418–3426, doi:10.1021/acs.jproteome.0c00254.
40. Schlesinger, D.; Elsässer, S.J. Revisiting sORFs: Overcoming challenges to identify and characterize functional microproteins. *FEBS J.* **2021**, doi:10.1111/febs.15769.
41. Chandramouly, G.; Zhao, J.; McDevitt, S.; Rusanov, T.; Hoang, T.; Borisonnik, N.; Treddinick, T.; Lopezcolorado, F.W.; Kent, T.; Siddique, L.A.; et al. Polθ reverse transcribes RNA and promotes RNA-templated DNA repair. *Sci. Adv.* **2021**, *7*, eabf1771, doi:10.1126/sciadv.abf1771.
42. Gygi, S.P.; Rochon, Y.; Franza, B.R.; Aebersold, R. Correlation between protein and mRNA abundance in yeast. *Mol. Cell. Biol.* **1999**, *19*, 1720–1730, doi:10.1128/mcb.19.3.1720.
43. Jiang, L.; Chan, J.; Fang, H.; Robinson, A.E.; Anand, S.; Gabriel, S.; Graubert, A.; Hadley, K.; Meier, S.R.; Nedzel, J.L.; et al. A quantitative proteome map of the human body. *Cell* **2020**, *183*, 269–283.e219, doi:10.1016/j.cell.2020.08.036.
44. Liu, Y.; Beyer, A.; Aebersold, R. On the dependency of cellular protein levels on mRNA abundance. *Cell* **2016**, *165*, 535–550, doi:10.1016/j.cell.2016.03.014.
45. Wanjek, C. Systems Biology as Defined by NIH. *NIH Catal.* **2011**, *19*, 10–12.
46. Hasin, Y.; Seldin, M.; Lusis, A. Multi-omics approaches to disease. *Genome Biol.* **2017**, *18*, 83, doi:10.1186/s13059-017-1215-1.
47. Argelaguet, R.; Velten, B.; Arnol, D.; Dietrich, S.; Zenz, T.; Marioni, J.C.; Buettner, F.; Huber, W.; Stegle, O. Multi-Omics Factor Analysis—a framework for unsupervised integration of multi-omics data sets. *Mol. Syst. Biol.* **2018**, *14*, e8124, doi:10.15252/msb.20178124.
48. Elliott, M.H.; Smith, D.S.; Parker, C.E.; Borchers, C. Current trends in quantitative proteomics. *J. Mass Spectrom.* **2009**, *44*, 1637–1660, doi:10.1002/jms.1692.
49. Naryzhny, S. Towards the full realization of 2DE power. *Proteomes* **2016**, *4*, 33, doi:10.3390/proteomes4040033.
50. Ponomarenko, E.A.; Poverennaya, E.V.; Ilgisonis, E.V.; Pyatritskiy, M.A.; Kopylov, A.T.; Zgoda, V.G.; Lisitsa, A.V.; Archakov, A.I. The size of the human proteome: The width and depth. *Int. J. Anal. Chem.* **2016**, *2016*, 7436849, doi:10.1155/2016/7436849.
51. Marcus, K.; Rabilloud, T. How do the different proteomic strategies cope with the complexity of biological regulations in a multi-omic world? Critical appraisal and suggestions for improvements. *Proteomes* **2020**, *8*, 23, doi:10.3390/proteomes8030023.
52. Marcus, K.; Lelong, C.; Rabilloud, T. What room for two-dimensional gel-based proteomics in a shotgun proteomics world? *Proteomes* **2020**, *8*, 17, doi:10.3390/proteomes8030017.
53. Naryzhny, S. Inventory of proteoforms as a current challenge of proteomics: Some technical aspects. *J. Proteom.* **2019**, *191*, 22–28, doi:10.1016/j.jprot.2018.05.008.
54. Lander, E.; Linton, L.; Birren, B.; Nusbaum, C.; Zody, M.; Baldwin, J.; Devon, K.; Dewar, K.; Doyle, M.; Fitzhugh, W.; et al. Initial sequencing and analysis of the human genome. *Nature* **2001**, *409*, 860–921, doi:10.1038/35057062.

55. Venter, J.; Adams, M.D.; Myers, E.; Li, P.; Mural, R.; Sutton, G.; Smith, H.; Yandell, M.; Evans, C.; Holt, R.; et al. The sequence of the human genome. *Science* **2001**, *291*, 1304–1351, doi:10.1126/science.1058040.
56. Abascal, F.; Juan, D.; Jungreis, I.; Martinez, L.; Rigau, M.; Jose, M.R.; Vazquez, J.; Tress, M.L. Loose ends: Almost one in five human genes still have unresolved coding status. *Nucleic Acids Res.* **2018**, *46*, 7070–7084, doi:10.1093/nar/gky587.
57. Sherman, R.M.; Forman, J.; Antonescu, V.; Puiu, D.; Daya, M.; Rafaels, N.; Boorgula, M.P.; Chavan, S.; Vergara, C.; Ortega, V.E.; et al. Assembly of a pan-genome from deep sequencing of 910 humans of African descent. *Nat. Genet.* **2019**, *51*, 30–35, doi:10.1038/s41588-018-0273-y.
58. Sherman, R.M.; Salzberg, S.L. Pan-genomics in the human genome era. *Nat. Rev. Genet.* **2020**, *21*, 243–254, doi:10.1038/s41576-020-0210-7.
59. Nurk, S.; Koren, S.; Rhie, A.; Rautiainen, M.; Bzikadze, A.V.; Mikheenko, A.; Vollger, M.R.; Altemose, N.; Uralsky, L.; Gershman, A.; et al. The complete sequence of a human genome. *bioRxiv* **2021**, doi:10.1101/2021.05.26.445798.
60. Clamp, M.; Fry, B.; Kamal, M.; Xie, X.; Cuff, J.; Lin, M.F.; Kellis, M.; Lindblad-Toh, K.; Lander, E.S. Distinguishing protein-coding and noncoding genes in the human genome. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 19428–19433, doi:10.1073/pnas.0709013104.
61. Cox, J.; Mann, M. Quantitative, high-resolution proteomics for data-driven systems biology. *Annu. Rev. Biochem.* **2011**, *80*, 273–299, doi:10.1146/annurev-biochem-061308-093216.
62. Patterson, S.D.; Aebersold, R.H. Proteomics: The first decade and beyond. *Nat. Genet. Suppl.* **2003**, *33*, 311–323.
63. Nesvizhskii, A.I.; Aebersold, R. Interpretation of shotgun proteomic data: The protein inference problem. *Mol. Cell. Proteom.* **2005**, *4*, 1419–1440, doi:10.1074/mcp.R500012-MCP200.
64. Coorssen, J.R.; Blank, P.S.; Albertorio, F.; Bezrukov, L.; Kolosova, I.; Backlund, P.S.; Zimmerberg, J. Quantitative femto- to attomole immunodetection of regulated secretory vesicle proteins critical to exocytosis. *Anal. Biochem.* **2002**, *307*, 54–62, doi:10.1016/S0003-2697(02)00015-5.
65. Coorssen, J.R.; Blank, P.S.; Albertorio, F.; Bezrukov, L.; Kolosova, I.; Chen, X.; Backlund, J.P.S.; Zimmerberg, J. Regulated secretion: SNARE density, vesicle fusion and calcium dependence. *J. Cell Sci.* **2003**, *116*, 2087–2097, doi:10.1242/jcs.00374.
66. Domon, B.; Aebersold, R. Options and considerations when selecting a quantitative proteomics strategy. *Nat. Biotechnol.* **2010**, *28*, 710–721, doi:10.1038/nbt.1661.
67. Borràs, E.; Sabidó, E. What is targeted proteomics? A concise revision of targeted acquisition and targeted data analysis in mass spectrometry. *Proteomics* **2017**, *17*, 1700180, doi:10.1002/pmic.201700180.
68. Moradian, A.; Kalli, A.; Sweredoski, M.J.; Hess, S. The top-down, middle-down, and bottom-up mass spectrometry approaches for characterization of histone variants and their post-translational modifications. *Proteomics* **2014**, *14*, 489–497, doi:10.1002/pmic.201300256.
69. Resing, K.A.; Ahn, N.G. Proteomics strategies for protein identification. *FEBS Lett.* **2005**, *579*, 885–889, doi:10.1016/j.febslet.2004.12.001.
70. Adhikari, S.; Nice, E.C.; Deutsch, E.W.; Lane, L.; Omenn, G.S.; Pennington, S.R.; Paik, Y.-K.; Overall, C.M.; Corrales, F.J.; Cristea, I.M.; et al. A high-stringency blueprint of the human proteome. *Nat. Commun.* **2020**, *11*, 5301–5301, doi:10.1038/s41467-020-19045-9.
71. Omenn, G.S.; Lane, L.; Overall, C.M.; Cristea, I.M.; Corrales, F.J.; Lindskog, C.; Paik, Y.-K.; Van Eyk, J.E.; Liu, S.; Pennington, S.R.; et al. Research on the human proteome reaches a major milestone: >90% of predicted human proteins now credibly detected, according to the HUPO Human Proteome Project. *J. Proteome Res.* **2020**, *19*, 4735–4746, doi:10.1021/acs.jproteome.0c00485.
72. Couso, J.P. Finding smORFs: Getting closer. *Genome Biol.* **2015**, *16*, 189–189, doi:10.1186/s13059-015-0765-3.
73. Gilmore, J.M.; Washburn, M.P. Advances in shotgun proteomics and the analysis of membrane proteomes. *J. Proteom.* **2010**, *73*, 2078–2091, doi:10.1016/j.jprot.2010.08.005.
74. Matallana-Surget, S.; Leroy, B.; Wattiez, R. Shotgun proteomics: Concept, key points and data mining. *Expert Rev. Proteom.* **2010**, *7*, 5–7, doi:10.1586/epr.09.101.
75. Carvalho, P.C.; Han, X.; Xu, T.; Cociorva, D.; Carvalho Mda, G.; Barbosa, V.C.; Yates, J.R., 3rd. XDIA: Improving on the label-free data-independent analysis. *Bioinformatics* **2010**, *26*, 847–848, doi:10.1093/bioinformatics/btq031.
76. Bell, A.J.; Deutsch, E.W.; Au, C.E.; Kearney, R.E.; Beavis, R.; Sechi, S.; Nilsson, T.; Bergeron, J.J.M. A HUPO test sample study reveals common problems in mass spectrometry-based proteomics. *Nat. Methods* **2009**, *6*, 423–430, doi:10.1038/nmeth.1333.
77. The, M.; Tasnim, A.; Käll, L. How to talk about protein-level false discovery rates in shotgun proteomics. *Proteomics* **2016**, *16*, 2461–2469, doi:10.1002/pmic.201500431.
78. Chait, B.T. Mass spectrometry: Bottom-up or top-down? *Science* **2006**, *314*, 65–66.
79. Streng, A.S.; de Boer, D.; Bouwman, F.G.; Mariman, E.C.M.; Scholten, A.; van Dieijen-Visser, M.P.; Wodzig, W.K.W.H. Development of a targeted selected ion monitoring assay for the elucidation of protease induced structural changes in cardiac troponin T. *J. Proteom.* **2016**, *136*, 123–132, doi:10.1016/j.jprot.2015.12.028.
80. Kuznetsova, K.G.; Levitsky, L.I.; Pyatnitskiy, M.A.; Ilina, I.Y.; Bubis, J.A.; Solovyeva, E.M.; Zgoda, V.G.; Gorshkov, M.V.; Moshkovskii, S.A. Cysteine alkylation methods in shotgun proteomics and their possible effects on methionine residues. *J. Proteom.* **2021**, *231*, 104022, doi:10.1016/j.jprot.2020.104022.
81. Wright, E.P.; Partridge, M.A.; Padula, M.P.; Gauci, V.J.; Malladi, C.S.; Coorssen, J.R. Top-down proteomics: Enhancing 2D gel electrophoresis from tissue processing to high-sensitivity protein detection. *Proteomics* **2014**, *14*, 872–889, doi:10.1002/pmic.201300424.

82. Yates, J.R.; Kelleher, N.L. Top down proteomics. *Anal. Chem.* **2013**, *85*, 6151, doi:10.1021/ac401484r.
83. Lobas, A.A.; Karpov, D.S.; Kopylov, A.T.; Solovyeva, E.M.; Ivanov, M.V.; Ilina, I.Y.; Lazarev, V.N.; Kuznetsova, K.G.; Ilgisonis, E.V.; Zgoda, V.G.; et al. Exome-based proteogenomics of HEK-293 human cell line: Coding genomic variants identified at the level of shotgun proteome. *Proteomics* **2016**, *16*, 1980–1991, doi:10.1002/pmic.201500349.
84. Gillet, L.C.; Navarro, P.; Tate, S.; Röst, H.; Selevsek, N.; Reiter, L.; Bonner, R.; Aebersold, R. Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: A new concept for consistent and accurate proteome analysis. *Mol. Cell. Proteom.* **2012**, *11*, O111.016717, doi:10.1074/mcp.O111.016717.
85. Jungblut, P.R. The proteomics quantification dilemma. *J. Proteom.* **2014**, *107*, 98–102, doi:10.1016/j.jprot.2014.03.015.
86. Bogdanow, B.; Zauber, H.; Selbach, M. Systematic errors in peptide and protein identification and quantification by modified peptides. *Mol. Cell. Proteom.* **2016**, *15*, 2791–2801, doi:10.1074/mcp.M115.055103.
87. Olsen, J.V.; Mann, M. Status of large-scale analysis of post-translational modifications by mass spectrometry. *Mol. Cell. Proteom.* **2013**, *12*, 3444–3452, doi:10.1074/mcp.O113.034181.
88. Alfaro, J.A.; Bohlander, P.; Dai, M.; Filius, M.; Howard, C.J.; van Kooten, X.F.; Ohayon, S.; Pomorski, A.; Schmid, S.; Aksimentiev, A.; et al. The emerging landscape of single-molecule protein sequencing technologies. *Nat. Methods* **2021**, *18*, 604–617, doi:10.1038/s41592-021-01143-1.
89. Shi, T.; Song, E.; Nie, S.; Rodland, K.D.; Liu, T.; Qian, W.J.; Smith, R.D. Advances in targeted proteomics and applications to biomedical research. *Proteomics* **2016**, *16*, 2160–2182, doi:10.1002/pmic.201500449.
90. Guo, T.; Kouvonen, P.; Koh, C.C.; Gillet, L.C.; Wolski, W.E.; Röst, H.L.; Rosenberger, G.; Collins, B.C.; Blum, L.C.; Gillessen, S.; et al. Rapid mass spectrometric conversion of tissue biopsy samples into permanent quantitative digital proteome maps. *Nat. Med.* **2015**, *21*, 407–413, doi:10.1038/nm.3807.
91. Ludwig, C.; Gillet, L.; Rosenberger, G.; Amon, S.; Collins, B.C.; Aebersold, R. Data-independent acquisition-based SWATH-MS for quantitative proteomics: A tutorial. *Mol. Syst. Biol.* **2018**, *14*, e8126, doi:10.15252/msb.20178126.
92. O'Farrell, P.H. High resolution two-dimensional electrophoresis of proteins. *J. Biol. Chem.* **1975**, *250*, 4007–4021.
93. Coorssen, J.; Yergey, A. Approaches to top-down proteomics: In honour of Prof. Patrick H. O'Farrell. *Proteomes* **2017**, *5*, 18, doi:10.3390/proteomes5030018.
94. Gygi, S.P.; Corthals, G.L.; Zhang, Y.; Rochon, Y.; Aebersold, R. Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 9390–9395, doi:10.1073/pnas.160270797.
95. Timms, J.F.; Cramer, R. Difference gel electrophoresis. *Proteomics* **2008**, *8*, 4886–4897, doi:10.1002/pmic.200800298.
96. Butt, R.H.; Coorssen, J.R. Postfractionation for enhanced proteomic analyses: Routine electrophoretic methods increase the resolution of standard 2D-PAGE. *J. Proteome Res.* **2005**, *4*, 982–991, doi:10.1021/pr050054d.
97. Butt, R.H.; Coorssen, J.R. Pre-extraction sample handling by automated frozen disruption significantly improves subsequent proteomic analyses. *J. Proteome Res.* **2006**, *5*, 437–448, doi:10.1021/pr0503634.
98. Thiede, B.; Koehler, C.J.; Strozynski, M.; Treumann, A.; Stein, R.; Zimny-Arndt, U.; Schmid, M.; Jungblut, P.R. High resolution quantitative proteomics of HeLa cells protein species using stable isotope labeling with amino acids in cell culture (SILAC), two-dimensional gel electrophoresis (2DE) and nano-liquid chromatography coupled to an LTQ-Orbitrap Mass spectrometer. *Mol. Cell. Proteom.* **2012**, *12*, 529–538, doi:10.1074/mcp.M112.019372.
99. Butt, R.H.; Coorssen, J.R. Coomassie blue as a near-infrared fluorescent stain: A systematic comparison with Sypro Ruby for in-gel protein detection. *Mol. Cell. Proteom.* **2013**, *12*, 3834–3850, doi:10.1074/mcp.M112.021881.
100. Gauci, V.J.; Padula, M.P.; Coorssen, J.R. Coomassie blue staining for high sensitivity gel-based proteomics. *J. Proteom.* **2013**, *90*, 96–106.
101. 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, doi:10.1016/j.jprot.2013.01.010.
102. Wright, E.P.; Prasad, K.A.G.; Padula, M.P.; Coorssen, J.R. Deep imaging: How much of the proteome does current top-down technology already resolve? *PLoS ONE* **2014**, *9*, e86058, doi:10.1371/journal.pone.0086058.
103. Noaman, N.; Abbineni, P.S.; Withers, M.; Coorssen, J.R. Coomassie staining provides routine (sub)femtomole in-gel detection of intact proteoforms: Expanding opportunities for genuine Top-down Proteomics. *Electrophoresis* **2017**, *38*, 3086–3099, doi:10.1002/elps.201700190.
104. Zhan, X.; Yang, H.; Peng, F.; Li, J.; Mu, Y.; Long, Y.; Cheng, T.; Huang, Y.; Li, Z.; Lu, M.; et al. How many proteins can be identified in a 2DE gel spot within an analysis of a complex human cancer tissue proteome? *Electrophoresis* **2018**, *39*, 965–980, doi:10.1002/elps.201700330.
105. Naryzhny, S.; Klopov, N.; Ronzhina, N.; Zorina, E.; Zgoda, V.; Kleyst, O.; Belyakova, N.; Legina, O. A database for inventory of proteoform profiles: “2DE-pattern”. *Electrophoresis* **2020**, *41*, 1118–1124, doi:10.1002/elps.201900468.
106. Bjellqvist, B.; Ek, K.; Giorgio Righetti, P.; Gianazza, E.; Görg, A.; Westermeier, R.; Postel, W. Isoelectric focusing in immobilized pH gradients: Principle, methodology and some applications. *J. Biochem. Biophys. Methods* **1982**, *6*, 317–339, doi:10.1016/0165-022X(82)90013-6.
107. Chevalier, F. Highlights on the capacities of “Gel-based” proteomics. (Review) (Report). *Proteome Sci.* **2010**, *8*, 23.
108. Pergande, M.R.; Cologna, S.M. Isoelectric point separations of peptides and proteins. *Proteomes* **2017**, *5*, 4, doi:10.3390/proteomes5010004.
109. Laemmli, U.K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **1970**, *227*, 680–685, doi:10.1038/227680a0.

110. Naryzhny, S.N.; Zgoda, V.G.; Maynskova, M.A.; Novikova, S.E.; Ronzhina, N.L.; Vakhrushev, I.V.; Khryapova, E.V.; Lisitsa, A.V.; Tikhonova, O.V.; Ponomarenko, E.A.; et al. Combination of virtual and experimental 2DE together with ESI LC-MS/MS gives a clearer view about proteomes of human cells and plasma. *Electrophoresis* **2016**, *37*, 302–309, doi:10.1002/elps.201500382.
111. D'Silva, A.M.; Hyett, J.A.; Coorssen, J.R. Proteomic analysis of first trimester maternal serum to identify candidate biomarkers potentially predictive of spontaneous preterm birth. *J. Proteom.* **2018**, *178*, 31–42, doi:10.1016/j.jprot.2018.02.002.
112. Kurgan, N.; Noaman, N.; Pergande, M.R.; Cologna, S.M.; Coorssen, J.R.; Klentrou, P. Changes to the human serum proteome in response to high intensity interval exercise: A sequential top-down proteomic analysis. *Front. Physiol.* **2019**, *10*, 362, doi:10.3389/fphys.2019.00362.
113. Zhu, K.; Zhao, J.; Lubman, D.M.; Miller, F.R.; Barder, T.J. Protein pI shifts due to posttranslational modifications in the separation and characterization of proteins. *Anal. Chem.* **2005**, *77*, 2745–2755, doi:10.1021/ac048494w.
114. Hsieh, E.J.; Bereman, M.S.; Durand, S.; Valaskovic, G.A.; MacCoss, M.J. Effects of column and gradient lengths on peak capacity and peptide identification in nanoflow LC-MS/MS of complex proteomic samples. *J. Am. Soc. Mass Spectrom.* **2013**, *24*, 148–153, doi:10.1007/s13361-012-0508-6.
115. Shishkova, E.; Alexander, S.H.; Coon, Joshua, J. Now, more than ever, proteomics needs better chromatography. *Cell Syst.* **2016**, *3*, 321–324, doi:10.1016/j.cels.2016.10.007.
116. Harris, L.R.; Churchward, M.A.; Butt, R.H.; Coorssen, J.R. Assessing detection methods for gel-based proteomic analyses. *J. Proteome Res.* **2007**, *6*, 1418–1425, doi:10.1021/pr0700246.
117. Gauci, V.J.; Wright, E.P.; Coorssen, J.R. Quantitative proteomics: Assessing the spectrum of in-gel protein detection methods. *J. Chem. Biol.* **2011**, *4*, 3–29, doi:10.1007/s12154-010-0043-5.
118. Neuhoﬀ, V.; Stamm, R.; Eibl, H. Clear background and highly sensitive protein staining with Coomassie Blue dyes in polyacrylamide gels: A systematic analysis. *Electrophoresis* **1985**, *6*, 427–448, doi:10.1002/elps.1150060905.
119. Colignon, B.; Raes, M.; Dieu, M.; Delaive, E.; Mauro, S. Evaluation of three-dimensional gel electrophoresis to improve quantitative profiling of complex proteomes. *Proteomics* **2013**, *13*, 2077–2082, doi:10.1002/pmic.201200494.
120. Marouga, R.; David, S.; Hawkins, E. The development of the DIGE system: 2D fluorescence difference gel analysis technology. *Anal. Bioanal. Chem.* **2005**, *382*, 669–678, doi:10.1007/s00216-005-3126-3.
121. Unlü, M.; Morgan, M.E.; Minden, J.S. Difference gel electrophoresis: A single gel method for detecting changes in protein extracts. *Electrophoresis* **1997**, *18*, 2071–2077, doi:10.1002/elps.1150181133.
122. Minden, J.S.; Dowd, S.R.; Meyer, H.E.; Stühler, K. Difference gel electrophoresis. *Electrophoresis* **2009**, *30*, S156–S161, doi:10.1002/elps.200900098.
123. McNamara, L.E.; Kantawong, F.A.; Dalby, M.J.; Riehle, M.O.; Burchmore, R. Preventing and troubleshooting artefacts in saturation labelled fluorescence 2-D difference gel electrophoresis (saturation DiGE). *Proteomics* **2011**, *11*, 4610–4621, doi:10.1002/pmic.201100135.
124. Hacker, S.M.; Backus, K.M.; Lazear, M.R.; Forli, S.; Correia, B.E.; Cravatt, B.F. Global profiling of lysine reactivity and ligandability in the human proteome. *Nat. Chem.* **2017**, *9*, 1181–1190, doi:10.1038/nchem.2826.
125. Tonge, R.; Shaw, J.; Middleton, B.; Rowlinson, R.; Rayner, S.; Young, J.; Pognan, F.; Hawkins, E.; Currie, I.; Davison, M. Validation and development of fluorescence two-dimensional differential gel electrophoresis proteomics technology. *Proteomics* **2001**, *1*, 377–396, [https://doi.org/10.1002/1615-9861\(200103\)1:3<377::AID-PROT377>3.0.CO;2-6](https://doi.org/10.1002/1615-9861(200103)1:3<377::AID-PROT377>3.0.CO;2-6).
126. Shaw, J.; Rowlinson, R.; Nickson, J.; Stone, T.; Sweet, A.; Williams, K.; Tonge, R. Evaluation of saturation labelling two-dimensional difference gel electrophoresis fluorescent dyes. *Proteomics* **2003**, *3*, 1181–1195, doi:10.1002/pmic.200300439.
127. McNamara, L.E.; Dalby, M.J.; Riehle, M.O.; Burchmore, R. Fluorescence two-dimensional difference gel electrophoresis for biomaterial applications. *J. R. Soc. Interface* **2010**, *7*, S107–S118, doi:10.1098/rsif.2009.0177.focus.
128. Herbert, B.; Galvani, M.; Hamdan, M.; Olivieri, E.; MacCarthy, J.; Pedersen, S.; Righetti, P.G. Reduction and alkylation of proteins in preparation of two-dimensional map analysis: Why, when, and how? *Electrophoresis* **2001**, *22*, 2046–2057, doi:10.1002/1522-2683(200106)22:10<2046::AID-ELPS2046>3.0.CO;2-C.
129. Smejkal, G.B.; Li, C.; Robinson, M.H.; Lazarev, A.V.; Lawrence, N.P.; Chernokalskaya, E. Simultaneous reduction and alkylation of protein disulfides in a centrifugal ultrafiltration device prior to two-dimensional gel electrophoresis. *J. Proteome Res.* **2006**, *5*, 983–987, doi:10.1021/pr050439w.
130. Kurien, B.T.; Scofield, R.H. Common artifacts and mistakes made in electrophoresis. *Methods Mol. Biol.* **2012**, *869*, 633–640, doi:10.1007/978-1-61779-821-4_58.
131. Sitek, B.; Lüttges, J.; Marcus, K.; Klöppel, G.; Schmiegel, W.; Meyer, H.E.; Hahn, S.A.; Stühler, K. Application of fluorescence difference gel electrophoresis saturation labelling for the analysis of microdissected precursor lesions of pancreatic ductal adenocarcinoma. *Proteomics* **2005**, *5*, 2665–2679, doi:10.1002/pmic.200401298.
132. Gade, D.; Thiermann, J.; Markowsky, D.; Rabus, R. Evaluation of two-dimensional difference gel electrophoresis for protein profiling. *J. Mol. Microbiol. Biotechnol.* **2003**, *5*, 240–251, doi:10.1159/000071076.
133. Lanouette, S.; Mongeon, V.; Figeys, D.; Couture, J.F. The functional diversity of protein lysine methylation. *Mol. Syst. Biol.* **2014**, *10*, 724–n/a, doi:10.1002/msb.134974.
134. Stastna, M.; Van Eyk, J.E. Posttranslational modifications of lysine and evolving role in heart pathologies-Recent developments. *Proteomics* **2015**, *15*, 1164–1180, doi:10.1002/pmic.201400312.
135. Couvertier, S.M.; Zhou, Y.; Weerapana, E. Chemical-proteomic strategies to investigate cysteine posttranslational modifications. *Biochim. Biophys. Acta* **2014**, *1844*, 2315–2330, doi:10.1016/j.bbapap.2014.09.024.

136. Shannon, D.A.; Weerapana, E. Orphan PTMs: Rare, yet functionally important modifications of cysteine. *Biopolymers* **2014**, *101*, 156–164, doi:10.1002/bip.22252.
137. Churchward, M.A.; Butt, R.H.; Lang, J.C.; Hsu, K.K.; Coorssen, J.R. Enhanced detergent extraction for analysis of membrane proteomes by two-dimensional gel electrophoresis. *Proteome Sci.* **2005**, *3*, 5, doi:10.1186/1477-5956-3-5.
138. Takemori, N.; Takemori, A.; Wongkongkathep, P.; Nshanian, M.; Loo, R.R.O.; Lermyte, F.; Loo, J.A. Top-down/bottom-up mass spectrometry workflow using dissolvable polyacrylamide gels. *Anal. Chem.* **2017**, *89*, 8244–8250, doi:10.1021/acs.analchem.7b00357.
139. Binz, P.-A.; Müller, M.; Hoogland, C.; Zimmermann, C.; Pasquarello, C.; Cortals, G.; Sanchez, J.-C.; Hochstrasser, D.F.; Appel, R.D. The molecular scanner: Concept and developments. *Curr. Opin. Biotechnol.* **2004**, *15*, 17–23, <https://doi.org/10.1016/j.cop-bio.2003.12.003>.
140. Papasotiriou, D.G.; Markoutsas, S.; Gorka, J.; Schleiff, E.; Karas, M.; Meyer, B. MALDI analysis of proteins after extraction from dissolvable ethylene glycol diacrylate cross-linked polyacrylamide gels. *Electrophoresis* **2013**, *34*, 2484–2494, doi:10.1002/elps.201300132.
141. Takemori, A.; Ishizaki, J.; Nakashima, K.; Shibata, T.; Kato, H.; Kodera, Y.; Suzuki, T.; Hasegawa, H.; Takemori, N. BAC-DROP: Rapid digestion of proteome fractionated via dissolvable polyacrylamide gel electrophoresis and its application to bottom-up proteomics workflow. *J. Proteome Res.* **2021**, *20*, 1535–1543, doi:10.1021/acs.jproteome.0c00749.
142. Toby, T.K.; Fornelli, L.; Kelleher, N.L. Progress in top-down proteomics and the analysis of proteoforms. *Annu. Rev. Anal. Chem.* **2016**, *9*, 499–519, doi:10.1146/annurev-anchem-071015-041550.
143. Huguet, R.; Mullen, C.; Srzentić, K.; Greer, J.B.; Fellers, R.T.; Zabrouskov, V.; Syka, J.E.P.; Kelleher, N.L.; Fornelli, L. Proton transfer charge reduction enables high-throughput top-down analysis of large proteoforms. *Anal. Chem.* **2019**, *91*, 15732–15739, doi:10.1021/acs.analchem.9b03925.
144. Kafader, J.O.; Durbin, K.R.; Melani, R.D.; Des Soye, B.J.; Schachner, L.F.; Senko, M.W.; Compton, P.D.; Kelleher, N.L. Individual ion mass spectrometry enhances the sensitivity and sequence coverage of top-down mass spectrometry. *J. Proteome Res.* **2020**, *19*, 1346–1350, doi:10.1021/acs.jproteome.9b00797.
145. Liang, Y.; Jin, Y.; Wu, Z.; Tucholski, T.; Brown, K.A.; Zhang, L.; Zhang, Y.; Ge, Y. Bridge hybrid monolithic column coupled to high-resolution mass spectrometry for top-down proteomics. *Anal. Chem.* **2019**, *91*, 1743–1747, doi:10.1021/acs.analchem.8b05817.
146. Melby, J.A.; Roberts, D.S.; Larson, E.J.; Brown, K.A.; Bayne, E.F.; Jin, S.; Ge, Y. Novel strategies to address the challenges in top-down proteomics. *J. Am. Soc. Mass Spectrom.* **2021**, *32*, 1278–1294, doi:10.1021/jasms.1c00099.
147. Melchior, K.; Tholey, A.; Heisel, S.; Keller, A.; Lenhof, H.-P.; Meese, E.; Huber, C.G. Protein- versus peptide fractionation in the first dimension of two-dimensional high-performance liquid chromatography-matrix-assisted laser desorption/ionization tandem mass spectrometry for qualitative proteome analysis of tissue samples. *J. Chromatogr. A* **2010**, *1217*, 6159–6168, doi:10.1016/j.chroma.2010.07.044.
148. Tran, J.C.; Zamborg, L.; Ahlf, D.R.; Lee, J.E.; Catherman, A.D.; Durbin, K.R.; Tipton, J.D.; Vellaichamy, A.; Kellie, J.F.; Li, M.; et al. Mapping intact protein isoforms in discovery mode using top-down proteomics. *Nature* **2011**, *480*, 254, doi:10.1038/nature10575.
149. Camerini, S.; Mauri, P. The role of protein and peptide separation before mass spectrometry analysis in clinical proteomics. *J. Chromatogr. A* **2015**, *1381*, 1–12, doi:10.1016/j.chroma.2014.12.035.
150. Tran, J.C.; Doucette, A.A. Gel-Eluted Liquid Fraction Entrapment Electrophoresis: An electrophoretic method for broad molecular weight range proteome separation. *Anal. Chem.* **2008**, *80*, 1568–1573, doi:10.1021/ac702197w.
151. Chrambach, A.; Jovin, T.M. Selected buffer systems for moving boundary electrophoresis on gels at various pH values, presented in a simplified manner. *Electrophoresis* **1983**, *4*, 190–204, <https://doi.org/10.1002/elps.1150040303>.
152. Model 491 Prep Cell and Mini Prep Cell. Available online: <https://www.bio-rad.com/en-ca/product/model-491-prep-cell-mini-prep-cell?ID=230a0852-ae4f-4861-b463-194663fdc7ac> (accessed on 1 July 2021).
153. Mehrotra, J.; Mittal, A.; Dhindsa, M.S.; Sinha, S. Fractionation of mycobacterial integral membrane proteins by continuous elution SDS-PAGE reveals the immunodominance of low molecular weight subunits for human T cells. *Clin. Exp. Immunol.* **1997**, *109*, 446–450, <https://doi.org/10.1046/j.1365-2249.1997.4531351.x>.
154. Krause, R.G.E.; Goldring, J.P.D. Crystal violet stains proteins in SDS-PAGE gels and zymograms. *Anal. Biochem.* **2019**, *566*, 107–115, doi:10.1016/j.ab.2018.11.015.
155. Takemori, A.; Butcher, D.S.; Harman, V.M.; Brownridge, P.; Shima, K.; Higo, D.; Ishizaki, J.; Hasegawa, H.; Suzuki, J.; Yamashita, M.; et al. PEPPI-MS: Polyacrylamide-gel-based prefractionation for analysis of intact proteoforms and protein complexes by mass spectrometry. *J. Proteome Res.* **2020**, *19*, 3779–3791, doi:10.1021/acs.jproteome.0c00303.
156. Banerjee, S.; Mazumdar, S. Electrospray ionization mass spectrometry: A technique to access the information beyond the molecular weight of the analyte. *Int. J. Anal. Chem.* **2012**, *2012*, 282574–282540, doi:10.1155/2012/282574.
157. Clarke, W. Mass Spectrometry in the Clinical Laboratory: Determining the Need and Avoiding Pitfalls. In *Mass Spectrometry for the Clinical Laboratory*; Nair, H., Clarke, W., Eds.; Academic Press: San Diego, CA, USA, 2017; pp. 1–15.
158. Krüger, R.; Karas, M. Formation and fate of ion pairs during MALDI analysis: Anion adduct generation as an indicative tool to determine ionization processes. *J. Am. Soc. Mass Spectrom.* **2002**, *13*, 1218–1226, doi:10.1016/S1044-0305(02)00450-6.
159. Susa, A.C.; Xia, Z.; Williams, E.R. Small emitter tips for native mass spectrometry of proteins and protein complexes from non-volatile buffers that mimic the intracellular environment. *Anal. Chem.* **2017**, *89*, 3116–3122, doi:10.1021/acs.analchem.6b04897.

160. Grandori, R. Origin of the conformation dependence of protein charge-state distributions in electrospray ionization mass spectrometry. *J. Mass Spectrom.* **2003**, *38*, 11–15, doi:10.1002/jms.390.
161. Compton, P.D.; Zamdborg, L.; Thomas, P.M.; Kelleher, N.L. On the scalability and requirements of whole protein mass spectrometry. *Anal. Chem.* **2011**, *83*, 6868–6874, doi:10.1021/ac2010795.
162. Comisarow, M.B.; Marshall, A.G. Theory of Fourier transform ion cyclotron resonance mass spectroscopy. I. Fundamental equations and low-pressure line shape. *J. Chem. Phys.* **1976**, *64*, 110–119, doi:10.1063/1.431959.
163. Nikolaev, E.N.; Kostyukevich, Y.I.; Vladimirov, G.N. Fourier transform ion cyclotron resonance (FT ICR) mass spectrometry: Theory and simulations. *Mass Spectrom. Rev.* **2016**, *35*, 219–258, doi:10.1002/mas.21422.
164. Smith, D.F.; Podgorski, D.C.; Rodgers, R.P.; Blakney, G.T.; Hendrickson, C.L. 21 tesla FT-ICR mass spectrometer for ultrahigh-resolution analysis of complex organic mixtures. *Anal. Chem.* **2018**, *90*, 2041–2047, doi:10.1021/acs.analchem.7b04159.
165. Marshall, A.G.; Comisarow, M.B.; Parisod, G. Relaxation and spectral line shape in Fourier transform ion resonance spectroscopy. *J. Chem. Phys.* **1979**, *71*, 4434–4444, doi:10.1063/1.438196.
166. Shaw, J.B.; Lin, T.-Y.; Leach III, F.E.; Tolmachev, A.V.; Tolić, N.; Robinson, E.W.; Koppelaar, D.W.; Paša-Tolić, L. 21 Tesla Fourier transform ion cyclotron resonance mass spectrometer greatly expands mass spectrometry toolbox. *J. Am. Soc. Mass Spectrom.* **2016**, *27*, 1929–1936, doi:10.1007/s13361-016-1507-9.
167. Nikolaev, E.; Lioznov, A. How to increase further the resolving power of the ultrahigh magnetic field FT ICR instruments? The new concept of the FT ICR cell—the open dynamically harmonized cell as a part of the vacuum system wall. *Anal. Chem.* **2021**, *93*, 1249–1253, doi:10.1021/acs.analchem.0c03237.
168. Marshall, A.G.; Hendrickson, C.L.; Jackson, G.S. Fourier transform ion cyclotron resonance mass spectrometry: A primer. *Mass Spectrom. Rev.* **1998**, *17*, 1–35, doi:10.1002/(SICI)1098-2787(1998)17:1<1::AID-MAS1>3.0.CO;2-K.
169. Perry, R.H.; Cooks, R.G.; Noll, R.J. Orbitrap mass spectrometry: Instrumentation, ion motion and applications. *Mass Spectrom. Rev.* **2008**, *27*, 661–699, doi:10.1002/mas.20186.
170. Makarov, A. Electrostatic axially harmonic orbital trapping: A high-performance technique of mass analysis. *Anal. Chem.* **2000**, *72*, 1156–1162, doi:10.1021/ac991131p.
171. Makarov, A.; Denisov, E.; Kholomeev, A.; Balschun, W.; Lange, O.; Strupat, K.; Horning, S. Performance evaluation of a hybrid linear ion trap/orbitrap mass spectrometer. *Anal. Chem.* **2006**, *78*, 2113–2120, doi:10.1021/ac0518811.
172. Fort, K.L.; van de Waterbeemd, M.; Boll, D.; Reinhardt-Szyba, M.; Belov, M.E.; Sasaki, E.; Zschoche, R.; Hilvert, D.; Makarov, A.A.; Heck, A.J.R. Expanding the structural analysis capabilities on an Orbitrap-based mass spectrometer for large macromolecular complexes. *Analyst* **2018**, *143*, 100–105, doi:10.1039/C7AN01629H.
173. Shaw, J.B.; Brodbelt, J.S. Extending the isotopically resolved mass range of orbitrap mass spectrometers. *Anal. Chem.* **2013**, *85*, 8313–8318, doi:10.1021/ac401634b.
174. Harper, C.C.; Elliott, A.G.; Oltrogge, L.M.; Savage, D.F.; Williams, E.R. Multiplexed charge detection mass spectrometry for high-throughput single ion analysis of large molecules. *Anal. Chem.* **2019**, *91*, 7458–7465, doi:10.1021/acs.analchem.9b01669.
175. Kafader, J.O.; Beu, S.C.; Early, B.P.; Melani, R.D.; Durbin, K.R.; Zabrouskov, V.; Makarov, A.A.; Maze, J.T.; Shinholt, D.L.; Yip, P.F.; et al. STORI plots enable accurate tracking of individual ion signals. *J. Am. Soc. Mass Spectrom.* **2019**, *30*, 2200–2203, doi:10.1007/s13361-019-02309-0.
176. Kafader, J.O.; Melani, R.D.; Durbin, K.R.; Ikwuagwu, B.; Early, B.P.; Fellers, R.T.; Beu, S.C.; Zabrouskov, V.; Makarov, A.A.; Maze, J.T.; et al. Multiplexed mass spectrometry of individual ions improves measurement of proteoforms and their complexes. *Nat. Methods* **2020**, *17*, 391–394, doi:10.1038/s41592-020-0764-5.
177. Li, Z.; Adams, R.M.; Chourey, K.; Hurst, G.B.; Hettich, R.L.; Pan, C. Systematic comparison of label-free, metabolic labeling, and isobaric chemical labeling for quantitative proteomics on LTQ Orbitrap Velos. *J. Proteome Res.* **2012**, *11*, 1582–1590, doi:10.1021/pr200748h.
178. Zhang, Z.; Pan, H.; Chen, X. Mass spectrometry for structural characterization of therapeutic antibodies. *Mass Spectrom. Rev.* **2009**, *28*, 147–176, doi:10.1002/mas.20190.
179. 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, doi:10.1021/acs.analchem.6b03756.
180. Ross, P.L.; Huang, Y.N.; Marchese, J.N.; Williamson, B.; Parker, K.; Hattai, S.; Khainovski, N.; Pillai, S.; Dey, S.; Daniels, S.; et al. Multiplexed protein quantification in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Mol. Cell. Proteom.* **2004**, *3*, 1154–1169, doi:10.1074/mcp.M400129-MCP200.
181. Chong, P.K.; Gan, C.S.; Pham, T.K.; Wright, P.C. Isobaric tags for relative and absolute quantitation (iTRAQ) reproducibility: Implication of multiple injections. *J. Proteome Res.* **2006**, *5*, 1232–1240, doi:10.1021/pr060018u.
182. Mahoney, D.W.; Therneau, T.M.; Heppelmann, C.J.; Higgins, L.; Benson, L.M.; Zenka, R.M.; Jagtap, P.; Nelsestuen, G.L.; Bergen, H.R.; Oberg, A.L. Relative quantification: Characterization of bias, variability and fold changes in mass spectrometry data from iTRAQ-labeled peptides. *J. Proteome Res.* **2011**, *10*, 4325–4333, doi:10.1021/pr2001308.
183. Evans, C.; Noirel, J.; Ow, S.Y.; Salim, M.; Pereira-Medrano, A.G.; Couto, N.; Pandhal, J.; Smith, D.; Pham, T.K.; Karunakaran, E.; et al. An insight into iTRAQ: Where do we stand now? *Anal. Bioanal. Chem.* **2012**, *404*, 1011–1027, doi:10.1007/s00216-012-5918-6.
184. Wang, X.; Li, Y.; Xu, G.; Liu, M.; Xue, L.; Liu, L.; Hu, S.; Zhang, Y.; Nie, Y.; Liang, S.; et al. Mechanism study of peptide GMBP1 and its receptor GRP78 in modulating gastric cancer MDR by iTRAQ-based proteomic analysis. *BMC Cancer* **2015**, *15*, 358, doi:10.1186/s12885-015-1361-3.

185. Jia, K.; Zhao, X.; Dang, X. Mass spectrometry-based iTRAQ analysis of serum markers in patients with pancreatic cancer. *Oncol. Lett.* **2020**, *19*, 4106–4114, doi:10.3892/ol.2020.11491.
186. Beati, H.; Langlands, A.; ten Have, S.; Müller, H.A.J. SILAC-based quantitative proteomic analysis of *Drosophila* gastrula stage embryos mutant for fibroblast growth factor signalling. *Fly* **2020**, *14*, 10–28, doi:10.1080/19336934.2019.1705118.
187. Schober, F.A.; Atanassov, I.; Moore, D.; Calvo-Garrido, J.; Moedas, M.F.; Wedell, A.; Freyer, C.; Wredenberg, A. Stable isotope labeling of amino acids in flies (SILAF) reveals differential phosphorylation of mitochondrial proteins upon loss of OXPHOS subunits. *Mol. Cell. Proteom.* **2021**, *20*, 100065, doi:10.1016/j.mcpro.2021.100065.
188. Issaq, H.J.; Conrads, T.P.; Janini, G.M.; Veenstra, T.D. Methods for fractionation, separation and profiling of proteins and peptides. *Electrophoresis* **2002**, *23*, 3048–3061, doi:10.1002/1522-2683(200209)23:17<3048::AID-ELPS3048>3.0.CO;2-L.
189. Fang, X.; Zhang, W.-W. Affinity separation and enrichment methods in proteomic analysis. *J. Proteom.* **2008**, *71*, 284–303, doi:10.1016/j.jprot.2008.06.011.
190. Zhang, Y.; Fonslow, B.R.; Shan, B.; Baek, M.-C.; Yates, J.R. Protein analysis by shotgun/bottom-up proteomics. *Chem. Rev.* **2013**, *113*, 2343–2394, doi:10.1021/cr3003533.
191. Pinkse, M.W.H.; Uitto, P.M.; Hilhorst, M.J.; Ooms, B.; Heck, A.J.R. Selective isolation at the femtomole level of phosphopeptides from proteolytic digests using 2D-NanoLC-ESI-MS/MS and titanium oxide precolumns. *Anal. Chem.* **2004**, *76*, 3935–3943, doi:10.1021/ac0498617.
192. Olsen, J.V.; Blagoev, B.; Gnäd, F.; Macek, B.; Kumar, C.; Mortensen, P.; Mann, M. Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell* **2006**, *127*, 635–648, doi:10.1016/j.cell.2006.09.026.
193. Nielsen, M.L.; Vermeulen, M.; Bonaldi, T.; Cox, J.; Moroder, L.; Mann, M. Iodoacetamide-induced artifact mimics ubiquitination in mass spectrometry. *Nat. Methods* **2008**, *5*, 459–460, doi:10.1038/nmeth0608-459.
194. Harlan, R.; Zhang, H. Targeted proteomics: A bridge between discovery and validation. *Expert Rev. Proteom.* **2014**, *11*, 657–661, doi:10.1586/14789450.2014.976558.
195. Parker, C.E.; Borchers, C.H. Mass spectrometry based biomarker discovery, verification, and validation—Quality assurance and control of protein biomarker assays. *Mol. Oncol.* **2014**, *8*, 840–858, doi:10.1016/j.molonc.2014.03.006.
196. Paulovich, A.G.; Whiteaker, J.R.; Hoofnagle, A.N.; Wang, P. The interface between biomarker discovery and clinical validation: The tar pit of the protein biomarker pipeline. *Proteom. Clin. Appl.* **2008**, *2*, 1386–1402, doi:10.1002/prca.200780174.
197. Hamers-Casterman, C.; Atarhouch, T.; Muyldermans, S.; Robinson, G.; Hamers, C.; Songa, E.B.; Bendahman, N.; Hamers, R. Naturally occurring antibodies devoid of light chains. *Nature* **1993**, *363*, 446–448, doi:10.1038/363446a0.
198. Cheloha, R.W.; Harmand, T.J.; Wijne, C.; Schwartz, T.U.; Ploegh, H.L. Exploring cellular biochemistry with nanobodies. *J. Biol. Chem.* **2020**, *295*, 15307–15327, doi:10.1074/jbc.REV120.012960.
199. Li, D.; Morisseau, C.; McReynolds, C.B.; Duflot, T.; Bellien, J.; Nagra, R.M.; Taha, A.Y.; Hammock, B.D. Development of improved double-nanobody sandwich ELISAs for human soluble epoxide hydrolase detection in peripheral blood mononuclear cells of diabetic patients and the prefrontal cortex of Multiple Sclerosis patients. *Anal. Chem.* **2020**, *92*, 7334–7342, doi:10.1021/acs.analchem.0c01115.
200. Lipman, N.S.; Jackson, L.R.; Trudel, L.J.; Weis-Garcia, F. Monoclonal versus polyclonal antibodies: Distinguishing characteristics, applications, and information resources. *ILAR J.* **2005**, *46*, 258–268, doi:10.1093/ilar.46.3.258.
201. Ascoli, C.A.; Aggeler, B. Overlooked benefits of using polyclonal antibodies. *BioTechniques* **2018**, *65*, 127–136, doi:10.2144/btn-2018-0065.
202. Hattori, T.; Koide, S. Next-generation antibodies for post-translational modifications. *Curr. Opin. Struct. Biol.* **2018**, *51*, 141–148, doi:10.1016/j.sbi.2018.04.006.
203. Haab, B.; Paulovich, A.G.; Anderson, N.L.; Clark, A.M.; Downing, G.J.; Hermjakob, H.; LaBaer, J.; Uhlen, M. A reagent resource to identify proteins and peptides of interest for the cancer community. *Mol. Cell. Proteom.* **2006**, *5*, 1996, doi:10.1074/mcp.T600020-MCP200.
204. Marx, V. Targeted proteomics. *Nat. Methods* **2013**, *10*, 19–22, doi:10.1038/nmeth.2285.
205. Gilda, J.E.; Ghosh, R.; Cheah, J.X.; West, T.M.; Bodine, S.C.; Gomes, A.V. Western blotting inaccuracies with unverified antibodies: Need for a Western blotting minimal reporting standard (WDBMRS). *PLoS ONE* **2015**, *10*, e0135392, doi:10.1371/journal.pone.0135392.
206. Bjerner, J.; Nustad, K.; Norum, L.F.; Olsen, K.H.; Bormer, O.P. Immunometric assay interference: Incidence and prevention. *Clin. Chem.* **2002**, *48*, 613–621.
207. Ylikotila, J.; Hellström, J.L.; Eriksson, S.; Vehniäinen, M.; Välimaa, L.; Takalo, H.; Bereznikova, A.; Pettersson, K. Utilization of recombinant Fab fragments in a cTnI immunoassay conducted in spot wells. *Clin. Biochem.* **2006**, *39*, 843–850, doi:10.1016/j.clinbiochem.2006.04.023.
208. Mann, M. Can proteomics retire the western blot? *J. Proteome Res.* **2008**, *7*, 3065, doi:10.1021/pr800463v.
209. Krastins, B.; Prakash, A.; Sarracino, D.A.; Nedelkov, D.; Niederkofer, E.E.; Kiernan, U.A.; Nelson, R.; Vogelsang, M.S.; Vadali, G.; Garces, A.; et al. Rapid development of sensitive, high-throughput, quantitative and highly selective mass spectrometric targeted immunoassays for clinically important proteins in human plasma and serum. *Clin. Biochem.* **2013**, *46*, 399–410, doi:10.1016/j.clinbiochem.2012.12.019.
210. Towbin, H.; Staehelin, T.; Gordon, J. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: Procedure and some applications. *Proc. Natl. Acad. Sci. USA* **1979**, *76*, 4350–4354, doi:10.1073/pnas.76.9.4350.

211. Butler, T.A.J.; Paul, J.W.; Chan, E.-C.; Smith, R.; Tolosa, J.M. Misleading Westerns: Common quantification mistakes in Western blot densitometry and proposed corrective measures. *BioMed Res. Int.* **2019**, 2019, 5214821, doi:10.1155/2019/5214821.
212. Hause, R.J.; Kim, H.D.; Leung, K.K.; Jones, R.B. Targeted protein-omic methods are bridging the gap between proteomic and hypothesis-driven protein analysis approaches. *Expert Rev Proteom.* **2011**, 8, 565–575, doi:10.1586/epr.11.49.
213. D'Silva, A.M.; Hyett, J.A.; Coorssen, J.R. First trimester protein biomarkers for risk of spontaneous preterm birth: Identifying a critical need for more rigorous approaches to biomarker identification and validation. *Fetal Diagn. Ther.* **2020**, 47, 497–506, doi:10.1159/000504975.
214. Mollica, J.P.; Oakhill, J.S.; Lamb, G.D.; Murphy, R.M. Are genuine changes in protein expression being overlooked? Reassessing Western blotting. *Anal. Biochem.* **2009**, 386, 270–275, doi:10.1016/j.ab.2008.12.029.
215. Taylor, S.C.; Posch, A. The design of a quantitative Western blot experiment. *BioMed Res. Int.* **2014**, 2014, 361590, doi:10.1155/2014/361590.
216. Kurien, B.T.; Scofield, R.H. Western Blotting: An Introduction. In *Western Blotting: Methods and Protocols*; Kurien, B.T., Scofield, R.H., Walker, J.M., Eds.; Methods in Molecular Biology Humana Press: New York, NY, USA, 2015; pp. 17–30.
217. Macphee, D.J. Methodological considerations for improving Western blot analysis. *J. Pharmacol. Toxicol. Methods* **2010**, 61, 171–177, doi:10.1016/j.vascn.2009.12.001.
218. Peferoen, M.; Huybrechts, R.; De Loof, A. Vacuum-blotting: A new simple and efficient transfer of proteins from sodium dodecyl sulfate—polyacrylamide gels to nitrocellulose. *FEBS Lett.* **1982**, 145, 369–372, doi:10.1016/0014-5793(82)80202-0.
219. Bass, J.J.; Wilkinson, D.J.; Rankin, D.; Phillips, B.E.; Szewczyk, N.J.; Smith, K.; Atherton, P.J. An overview of technical considerations for Western blotting applications to physiological research. *Scand. J. Med. Sci. Sports* **2017**, 27, 4–25, doi:10.1111/sms.12702.
220. Wang, G.; de Jong, R.N.; van den Bremer, E.T.J.; Parren, P.W.H.I.; Heck, A.J.R. Enhancing accuracy in molecular weight determination of highly heterogeneously glycosylated proteins by native tandem mass spectrometry. *Anal. Chem.* **2017**, 89, 4793–4797, doi:10.1021/acs.analchem.6b05129.
221. Alegria-Schaffer, A.; Lodge, A.; Vattam, K. Performing and Optimizing Western Blots with an Emphasis on Chemiluminescent Detection. Available online: <https://www.sciencedirect.com/science/article/abs/pii/S0076687909630330> (accessed on 1 July 2021).
222. Dorri, Y. Immunoblotting of Antigens: Whole, Strip, and New-Line Nitrocellulose Membrane Immunoblotting Using the Chemiluminescence Technique. In *Detection of Blotted Proteins: Methods and Protocols*, 1st ed.; Kurien, B.T., Scofield, R.H., Walker, J.M., Eds.; Methods in Molecular Biology Humana Press: New York, NY, USA, 2015; pp. 79–94.
223. Desai, S.; Dworecki, B.R.; Nlend, M.C. Direct Immunodetection of Antigens Within the Precast Polyacrylamide Gel. In *Detection of Blotted Proteins: Methods and Protocols*, 1st ed.; Kurien, B.T., Scofield, R.H., Walker, J.M., Eds.; Methods in Molecular Biology Humana Press: New York, NY, USA, 2015; pp. 101–114.
224. Gingrich, J.C.; Davis, D.R.; Nguyen, Q. Multiplex detection and quantitation of proteins on Western blots using fluorescent probes. *BioTechniques* **2000**, 29, 636–642, doi:10.2144/00293pf02.
225. Deng, D.; Hao, Y.; Xue, J.; Liu, X.; Xu, X.; Liu, L. A colorimetric enzyme-linked immunosorbent assay with CuO nanoparticles as signal labels based on the growth of gold nanoparticles in situ. *Nanomaterials* **2018**, 9, 4, doi:10.3390/nano9010004.
226. Wu, C.; Shan, Y.; Wu, X.; Wang, S.; Liu, F. Quantitative protein detection using single molecule imaging enzyme-linked immunosorbent assay (iELISA). *Anal. Biochem.* **2019**, 587, 113466, doi:10.1016/j.ab.2019.113466.
227. Rissin, D.M.; Walt, D.R. Digital concentration readout of single enzyme molecules using femtoliter arrays and Poisson statistics. *Nano Lett.* **2006**, 6, 520–523, doi:10.1021/nl060227d.
228. Rissin, D.M.; Walt, D.R. Digital readout of target binding with attomole detection limits via enzyme amplification in femtoliter arrays. *J. Am. Chem. Soc.* **2006**, 128, 6286–6287, doi:10.1021/ja058425e.
229. Rissin, D.M.; Gorris, H.H.; Walt, D.R. Distinct and long-lived activity states of single enzyme molecules. *J. Am. Chem. Soc.* **2008**, 130, 5349–5353, doi:10.1021/ja711414f.
230. Wu, D.; Milutinovic, M.D.; Walt, D.R. Single molecule array (Simoa) assay with optimal antibody pairs for cytokine detection in human serum samples. *Analyst* **2015**, 140, 6277–6282, doi:10.1039/c5an01238d.
231. Rissin, D.M.; Kan, C.W.; Campbell, T.G.; Howes, S.C.; Fournier, D.R.; Song, L.; Piech, T.; Patel, P.P.; Chang, L.; Rivnak, A.J.; et al. Single-molecule enzyme-linked immunosorbent assay detects serum proteins at subfemtomolar concentrations. *Nat. Biotechnol.* **2010**, 28, 595–599, doi:10.1038/nbt.1641.
232. Coons, A.H.; Creech, H.J.; Jones, R.N. Immunological properties of an antibody containing a fluorescent group. *Proc. Soc. Exp. Biol. Med.* **1941**, 47, 200–202, doi:10.3181/00379727-47-13084p.
233. Gatter, K.C.; Abdulaziz, Z.; Beverley, P.; Corvalan, J.R.F.; Ford, C.; Lane, E.B.; Mota, M.; Nash, J.R.G.; Pulford, K.; Stein, H.; et al. Use of monoclonal antibodies for the histopathological diagnosis of human malignancy. *J. Clin. Pathol.* **1982**, 35, 1253–1267, doi:10.1136/jcp.35.11.1253.
234. Lohse, J.; Petersen, K.H.; Woller, N.C.; Pedersen, H.C.; Skladtchikova, G.; Jørgensen, R.M. Improved catalyzed reporter deposition, iCARD. *Bioconjugate Chem.* **2014**, 25, 1036–1042, doi:10.1021/bc400311g.
235. Jensen, K.; Krusenstjerna-Hafstrøm, R.; Lohse, J.; Petersen, K.H.; Derand, H. A novel quantitative immunohistochemistry method for precise protein measurements directly in formalin-fixed, paraffin-embedded specimens: Analytical performance measuring HER2. *Mod. Pathol.* **2017**, 30, 180–193, doi:10.1038/modpathol.2016.176.
236. Anand, S.; Samuel, M.; Ang, C.-S.; Keerthikumar, S.; Mathivanan, S. Label-based and label-free strategies for protein quantification. In *Proteome Bioinformatics*; Keerthikumar, S., Mathivanan, S., Eds.; Springer: New York, NY, USA, 2017; pp. 31–43.

237. Bantscheff, M.; Schirle, M.; Sweetman, G.; Rick, J.; Kuster, B. Quantitative mass spectrometry in proteomics: A critical review. *Anal. Bioanal. Chem.* **2007**, *389*, 1017–1031, doi:10.1007/s00216-007-1486-6.
238. Gouw, J.W.; Krijgsveld, J.; Heck, A.J.R. Quantitative proteomics by metabolic labeling of model organisms. *Mol. Cell. Proteom.* **2010**, *9*, 11–24, doi:10.1074/mcp.R900001-MCP200.
239. Chahrour, O.; Cobice, D.; Malone, J. Stable isotope labelling methods in mass spectrometry-based quantitative proteomics. *J. Pharm. Biomed. Anal.* **2015**, *113*, 2–20, doi:10.1016/j.jpba.2015.04.013.
240. Ankney, J.A.; Muneer, A.; Chen, X. Relative and absolute quantitation in mass spectrometry-based proteomics. *Annu. Rev. Anal. Chem.* **2018**, *11*, 49–77, doi:10.1146/annurev-anchem-061516-045357.
241. Xie, F.; Liu, T.; Qian, W.-J.; Petyuk, V.A.; Smith, R.D. Liquid chromatography-mass spectrometry-based quantitative proteomics. *J. Biol. Chem.* **2011**, *286*, 25443–25449, doi:10.1074/jbc.r110.199703.
242. Quijada, J.V.; Schmitt, N.D.; Salisbury, J.P.; Auclair, J.R.; Agar, J.N. Heavy sugar and heavy water create tunable intact protein mass increases for quantitative mass spectrometry in any feed and organism. *Anal. Chem.* **2016**, *88*, 11139–11146, doi:10.1021/acs.analchem.6b03234.
243. Thompson, A.; Schäfer, J.; Kuhn, K.; Kienle, S.; Schwarz, J.; Schmidt, G.; Neumann, T.; Hamon, C. Tandem mass tags: A novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal. Chem.* **2003**, *75*, 1895–1904, doi:10.1021/ac0262560.
244. Rauniyar, N.; Gao, B.; McClatchy, D.B.; Yates, J.R. Comparison of protein expression ratios observed by sixplex and duplex TMT labeling method. *J. Proteome Res.* **2013**, *12*, 1031–1039, doi:10.1021/pr3008896.
245. Zecha, J.; Satpathy, S.; Kanashova, T.; Avanesian, S.C.; Kane, M.H.; Clauser, K.R.; Mertins, P.; Carr, S.A.; Kuster, B. TMT labeling for masses: A robust and cost-efficient, in-solution labeling approach. *Mol. Cell. Proteom.* **2019**, *18*, 1468–1478, doi:10.1074/mcp.TIR119.001385.
246. Megger, D.A.; Pott, L.L.; Ahrens, M.; Padden, J.; Bracht, T.; Kuhlmann, K.; Eisenacher, M.; Meyer, H.E.; Sitek, B. Comparison of label-free and label-based strategies for proteome analysis of hepatoma cell lines. *Biochim. Biophys. Acta* **2014**, *1844*, 967–976, doi:10.1016/j.bbapap.2013.07.017.
247. 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, doi:10.3390/ijms161226120.
248. Shen, X.; Shen, S.; Li, J.; Hu, Q.; Nie, L.; Tu, C.; Wang, X.; Orsburn, B.; Wang, J.; Qu, J. An IonStar experimental strategy for MS1 ion current-based quantification using ultrahigh-field orbitrap: Reproducible, in-depth, and accurate protein measurement in large cohorts. *J. Proteome Res.* **2017**, *16*, 2445–2456, doi:10.1021/acs.jproteome.7b00061.
249. Lange, V.; Picotti, P.; Domon, B.; Aebersold, R. Selected reaction monitoring for quantitative proteomics: A tutorial. *Mol. Syst. Biol.* **2008**, *4*, 222, doi:10.1038/msb.2008.61.
250. Tang, J.; Zhang, Y.; Fu, J.; Wang, Y.; Li, Y.; Yang, Q.; Yao, L.; Xue, W.; Zhu, F. Computational advances in the label-free quantification of cancer proteomics data. *Curr. Pharm. Des.* **2018**, *24*, 3842–3858, doi:10.2174/1381612824666181102125638.
251. Sticker, A.; Goeminne, L.; Martens, L.; Clement, L. Robust summarization and inference in proteome-wide label-free quantification. *Mol. Cell. Proteom.* **2020**, *19*, 1209–1219, doi:10.1074/mcp.RA119.001624.
252. Prakash, A.; Rezai, T.; Krastins, B.; Sarracino, D.; Athanas, M.; Russo, P.; Ross, M.M.; Zhang, H.; Tian, Y.; Kulasingam, V.; et al. Platform for establishing interlaboratory reproducibility of selected reaction monitoring-based mass spectrometry peptide assays. *J. Proteome Res.* **2010**, *9*, 6678–6688, doi:10.1021/pr100821m.
253. Kockmann, T.; Trachsel, C.; Panse, C.; Wahlander, A.; Selevsek, N.; Grossmann, J.; Wolski, W.E.; Schlapbach, R. Targeted proteomics coming of age—SRM, PRM and DIA performance evaluated from a core facility perspective. *Proteomics* **2016**, *16*, 2183–2192, doi:10.1002/pmic.201500502.
254. Hu, A.; Noble, W.S.; Wolf-Yadlin, A. Technical advances in proteomics: New developments in data-independent acquisition. *F1000 Res.* **2016**, *5*, 419, doi:10.12688/f1000research.7042.1.
255. Sen, M.K.; Almuslehi, M.S.M.; Shortland, P.J.; Mahns, D.A.; Coorssen, J.R. Proteomics of Multiple Sclerosis: Inherent issues in defining the pathoetiology and identifying (early) biomarkers. *Int. J. Mol. Sci.* **2021**, *22*, 7377, doi:10.3390/ijms22147377.
256. Rabilloud, T. When 2D is not enough, go for an extra dimension. *Proteomics* **2013**, *13*, 2065–2068, doi:10.1002/pmic.201300215.
257. Liu, J.; Wang, F.; Mao, J.; Zhang, Z.; Liu, Z.; Huang, G.; Cheng, K.; Zou, H. High-sensitivity N-glycoproteomic analysis of mouse brain tissue by protein extraction with a mild detergent of N-dodecyl β -D-maltoside. *Anal. Chem.* **2015**, *87*, 2054–2057, doi:10.1021/ac504700t.
258. Epstein, J.A.; Blank, P.S.; Searle, B.C.; Catlin, A.D.; Cologna, S.M.; Olson, M.T.; Backlund, P.S.; Coorssen, J.R.; Yergey, A.L. ProteinProcessor: A probabilistic analysis using mass accuracy and the MS spectrum. *Proteomics* **2016**, *16*, 2480–2490, doi:10.1002/pmic.201600137.
259. Cesnik, A.J.; Shortreed, M.R.; Schaffer, L.V.; Knoener, R.A.; Frey, B.L.; Scalf, M.; Solntsev, S.K.; Dai, Y.; Gasch, A.P.; Smith, L.M. Proteoform Suite: Software for constructing, quantifying, and visualizing proteoform families. *J. Proteome Res.* **2018**, *17*, 568–578, doi:10.1021/acs.jproteome.7b00685.
260. Donnelly, D.P.; Rawlins, C.M.; DeHart, C.J.; Fornelli, L.; Schachner, L.F.; Lin, Z.; Lippens, J.L.; Aluri, K.C.; Sarin, R.; Chen, B.; et al. Best practices and benchmarks for intact protein analysis for top-down mass spectrometry. *Nat. Methods* **2019**, *16*, 587–594, doi:10.1038/s41592-019-0457-0.

-
261. Hardman, G.; Perkins, S.; Brownridge, P.J.; Clarke, C.J.; Byrne, D.P.; Campbell, A.E.; Kalyuzhnyy, A.; Myall, A.; Evers, P.A.; Jones, A.R.; et al. Strong anion exchange-mediated phosphoproteomics reveals extensive human non-canonical phosphorylation. *EMBO J.* **2019**, *38*, e100847, doi:10.15252/embj.2018100847.
262. Meyer, J.G. Fast proteome identification and quantification from data-dependent acquisition-tandem mass spectrometry (DDA MS/MS) using free software tools. *Methods Protoc.* **2019**, *2*, 8, doi:10.3390/mps2010008.
263. Prakash, A.; Ahmad, S.; Majumder, S.; Jenkins, C.; Orsburn, B. Bolt: A new age peptide search engine for comprehensive MS/MS sequencing through vast protein databases in minutes. *J. Am. Soc. Mass Spectrom.* **2019**, *30*, 2408–2418, doi:10.1007/s13361-019-02306-3.
264. Chen, C.; Hou, J.; Tanner, J.J.; Cheng, J. Bioinformatics methods for mass spectrometry-based proteomics data analysis. *Int. J. Mol. Sci.* **2020**, *21*, 2873, doi:10.3390/ijms21082873.
265. Hubler, S.L.; Kumar, P.; Mehta, S.; Easterly, C.; Johnson, J.E.; Jagtap, P.D.; Griffin, T.J. Challenges in peptide-spectrum matching: A robust and reproducible statistical framework for removing low-accuracy, high-scoring hits. *J. Proteome Res.* **2020**, *19*, 161–173, doi:10.1021/acs.jproteome.9b00478.
266. Teclemariam, E.T.; Pergande, M.R.; Cologna, S.M. Considerations for mass spectrometry-based multi-omic analysis of clinical samples. *Expert Rev. Proteom.* **2020**, *17*, 99–107, doi:10.1080/14789450.2020.1724540.
267. Yu, F.; Haynes, S.E.; Teo, G.C.; Avtonomov, D.M.; Polasky, D.A.; Nesvizhskii, A.I. Fast quantitative analysis of timsTOF PASEF data with MSFragger and IonQuant. *Mol. Cell. Proteom.* **2020**, *19*, 1575–1585, doi:10.1074/mcp.TIR120.002048.
268. Cappelletti, V.; Hauser, T.; Piazza, I.; Pepelnjak, M.; Malinowska, L.; Fuhrer, T.; Li, Y.; Dörig, C.; Boersema, P.; Gillet, L.; et al. Dynamic 3D proteomes reveal protein functional alterations at high resolution in situ. *Cell* **2021**, *184*, 545–559.e522, doi:10.1016/j.cell.2020.12.021.
269. Deutsch, E.W.; Perez-Riverol, Y.; Carver, J.; Kawano, S.; Mendoza, L.; Van Den Bossche, T.; Gabriels, R.; Binz, P.-A.; Pullman, B.; Sun, Z.; et al. Universal Spectrum Identifier for mass spectra. *Nat. Methods* **2021**, *18*, 768–770, doi:10.1038/s41592-021-01184-6.
270. Geiszler, D.J.; Kong, A.T.; Avtonomov, D.M.; Yu, F.; Leprevost, F.D.V.; Nesvizhskii, A.I. PTM-Shepherd: Analysis and summarization of post-translational and chemical modifications from open search results. *Mol. Cell. Proteom.* **2021**, *20*, 100018, doi:10.1074/mcp.TIR120.002216.
271. Meier, F.; Park, M.A.; Mann, M. Trapped ion mobility spectrometry (TIMS) and parallel accumulation—Serial fragmentation (PASEF) in proteomics. *Mol. Cell. Proteom.* **2021**, 100138, <https://doi.org/10.1016/j.mcpro.2021.100138>.
272. Onisiforou, A.; Spyrou, G.M. Identification of viral-mediated pathogenic mechanisms in neurodegenerative diseases using network-based approaches. *Brief. Bioinform.* **2021**, doi:10.1093/bib/bbab141.
273. Riffle, M.; Hoopmann, M.R.; Jaschob, D.; Zhong, G.; Moritz, R.L.; MacCoss, M.J.; Davis, T.N.; Isoherranen, N.; Zelter, A. Discovery and visualization of uncharacterized drug-protein adducts using mass spectrometry. *bioRxiv* **2021**, doi:10.1101/2021.06.24.449838.
274. Theodorakis, E.; Antonakis, A.N.; Baltsavia, I.; Georgios, A.P.; Samiotaki, M.; Grigoris, D.A.; Theodosiou, T.; Acuto, O.; Efsthathiou, G.; Iliopoulos, I. ProteoSign v2: A faster and evolved user-friendly online tool for statistical analyses of differential proteomics. *Nucleic Acids Res.* **2021**, *49*, W573–W577, doi:10.1093/nar/gkab329.