



HAL
open science

How Do the Different Proteomic Strategies Cope with the Complexity of Biological Regulations in a Multi-Omic World? Critical Appraisal and Suggestions for Improvements

Katrin Marcus, Thierry Rabilloud

► To cite this version:

Katrin Marcus, Thierry Rabilloud. 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 (3), pp.23. 10.3390/proteomes8030023 . hal-02932136

HAL Id: hal-02932136

<https://hal.science/hal-02932136>

Submitted on 7 Sep 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Review

How Do the Different Proteomic Strategies Cope with the Complexity of Biological Regulations in a Multi-Omic World? Critical Appraisal and Suggestions for Improvements

Katrin Marcus ¹ and Thierry Rabilloud ^{2,*} 

¹ Medizinisches Proteom-Center, Medical Faculty & Medical Proteome Analysis, Center for Proteindiagnostics (PRODI) Ruhr-University Bochum Gesundheitscampus, 4 44801 Bochum, Germany; Katrin.Marcus@ruhr-uni-bochum.de

² Laboratory of Chemistry and Biology of Metals, Université Grenoble Alpes UMR 5249, CNRS, CEA 38054 Grenoble, France

* Correspondence: thierry.rabilloud@cnrs.fr

Received: 11 August 2020; Accepted: 1 September 2020; Published: 3 September 2020



Abstract: In this second decade of the 21st century, we are lucky enough to have different types of proteomic analyses at our disposal. Furthermore, other functional omics such as transcriptomics have also undergone major developments, resulting in mature tools. However, choice equals questions, and the major question is how each proteomic strategy is fit for which purpose. The aim of this opinion paper is to reposition the various proteomic strategies in the frame of what is known in terms of biological regulations in order to shed light on the power, limitations, and paths for improvement for the different proteomic setups. This should help biologists to select the best-suited proteomic strategy for their purposes in order not to be driven by raw availability or fashion arguments but rather by the best fitness for purpose. In particular, knowing the limitations of the different proteomic strategies helps in interpreting the results correctly and in devising the validation experiments that should be made downstream of the proteomic analyses.

Keywords: two-dimensional gel electrophoresis; proteomics; post-translational modifications; proteoforms; clinical proteomics

1. The “City Noir” of Proteomics: Where Is the Light?

City noir is not only a wonderful piece of 20th century music composed by John Adams; it also has the wonderful cover displayed on Figure 1:

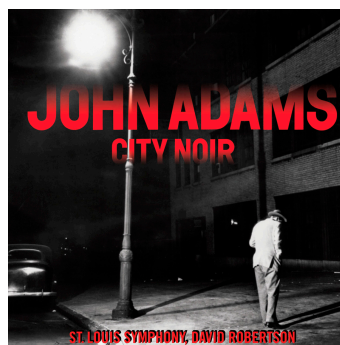


Figure 1. The city noir cover (reproduced with permission of Nonesuch records).

This cover is highly reminiscent of a well-known story, where you cross a person like the one on the cover, obviously looking for something under the lamppost. The following dialog ensues:

“Hey Sir, you seem to be looking for something, may I help you?”

“Sure, I’ve lost my keys”

After a search when both of you can detect nothing, you say:

“Well, I can’t see any keys, are you sure that you lost them around here?”

“Oh, probably not, but at least, here there is light . . . ”

This story is indeed a metaphor of a well-known cognitive bias, which is called the cognitive availability bias. This bias consists of using a piece of data to make a point, even when knowing that these data are not really relevant, because the data are available, and not questioning any further the fitness of the data for the purpose. This has happened (and still happens) countless times in science and is very prevalent in proteomics and in functional omics in general. The problem is not to use a surrogate proxy for the real object of interest, biochemically speaking, but arises when the limits of the proxy are forgotten and the results over-interpreted. In order to get a fair appraisal of where the light in proteomics really is, it is first necessary to take a step back and remind us of what types of proteomics we have at our disposal and then how cells work.

2. The Landscape

2.1. The Different Proteomic Strategies

The word “proteome,” defined as the protein complement of the genome, was first used in a publication in 1995 [1]. Remarkably, this very definition tells that the proteome comes after the genome, and indeed there is no proteomic science without genomic science. Still today, when working on a new organism, a totally sequenced and well-annotated genome is a prerequisite for doing proteomics. In a sense, all proteomic data are interpreted from the genomic book. However, there are different ways to acquire and read proteomic data. Depending on the analyte that is separated and quantified, two major types of proteomics can be described.

Top-down proteomics separates and quantifies complete proteins prior to their analysis by mass spectrometry. Depending on the main tool used, top-down proteomics comes in two very different setups.

In the first setup, two-dimensional gel electrophoresis (2DGE) is the tool used to separate and quantify proteins, and this leads to 2DGE proteomics. This is the oldest type of proteomic setup used, as exemplified in the first proteomic paper [1]. Although no longer in fashion, this proteomic setup still has a lot to offer [2,3]. In this setup, proteins are digested into peptides to carry out the mass spectrometry based identification/characterization (as first described in Henzel et al. [4]), although complete proteins have sometimes been eluted from 2D gels and analyzed directly in the mass spectrometer [5,6]. In the second setup, i.e., mass spectrometry (MS)-based top-down proteomics [7], quantification and analysis are carried out on complete proteins injected in the mass spectrometer, although a variable extent of proteins separation can be applied before injection in the mass spectrometer to deconvolute the samples [8].

In contrast to top-down proteomics, bottom-up proteomics separates, quantifies, and analyzes not complete proteins but peptides produced by the digestion of the proteins. Peptides are analyzed by mass spectrometry, and the resulting data is compared with a database using automated search programs followed by an *in silico* reassembly of the respective proteins. Intellectually speaking, the method derives from the observation that univocal identification can be obtained from a few peptides coming from the digestion of a protein [4]. Thus, if digestion peptides can be identified well enough against a genome sequence, and if these digestion peptides can be produced, handled,

separated, and analyzed reproducibly, why not use them as a representative proxy for proteins? This approach, first described in 1997 by Yates et al. [9], has met considerable success because peptides are very convenient proxies for proteins. They are easy to produce and to handle, separate well by simple (in principle) chromatography, and fragment well in the mass spectrometer (at least from the tryptic peptides that are most often used). By contrast, proteins follow the “3S rule”: they are Sticky (difficult to extract quantitatively), Soft (difficult to fragment neatly in a mass spectrometer), and Separation-unfriendly (despite what can be achieved by 2DGE).

Thus, looping back to the city noir metaphor, the bottom-up proteomics lamppost sheds bright light, but what are the keys available under this lamppost? As scientists, we know the power of Murphy’s law, which implies, among many other consequences, that convenience and relevance do not often come hand-in-hand. They are rather two sides of a coin—either you see one, or you see the other.

In fact, the cardinal and insuperable flaw of bottom-up proteomics is the loss of the connectivity that is present between peptides in a complete protein. Thus, bottom-up proteomics only collects peptide-wide information. As peptides coming from an experimentator-driven digestion have no intrinsic functional role (which complete proteins have), protein-wide information must be reconstructed *in silico* from the actual peptide-wide information, a process that is much less straightforward than it seems. What are the issues here? Peptides generated from and matching just one single protein (isoform) in the database are called unique. Peptides that cannot be clearly assigned to one only but to several proteins or proteoforms are “shared” peptides. Proteoform is defined as all different molecular forms of a protein product from a single gene including its modifications [10]. Those shared peptides do not enable an unambiguous assignment and thus no unambiguous protein identification. This fact is addressed using a process referred to as “protein inference.” Protein identification (and quantification) is usually obtained with so-called peptide search engines that match experimentally obtained peptide spectra to theoretically derived ones based on the sequences of *in silico* digested protein sequences from a database. The result is a peptide-spectrum match (PSM). Several spectra may match to only one peptide. If, in turn, one PSM does not unambiguously represent one protein, the result will be a list of proteins grouped into a protein group. Conventionally used data search algorithms and programs are SEQUEST [11], Mascot [12], X!Tandem [13] MaxQuant [14], Proteome Discoverer (Thermo Scientific), or Progenesis QIP (Nonlinear). Up-to now, these software packages can handle the protein inference issue only in parts. Over the years several attempts have been made to address this challenge and additional tools have been developed (e.g. in references [15–20]), improving the situation, but the protein inference problem is still not fully solved up to now. To make a clear point here, all search algorithms and programs will process the MS/MS data according to personally defined parameters. The result is a list of proteins or protein groups, but it shall be kept in mind that crucial information about the actual and biologically relevant proteoform might have been lost on the way.

Overall, this reconstruction process is operated against the genetic template, which means, in turn, that bottom-up proteomics operate at the gene product level, i.e., the theoretical protein sequence predicted from the gene. Thus, the final relevance of bottom-up proteomics will depend on the extent to which the gene product is a good functional proxy for the actual proteins produced in the organism of interest.

2.2. How Cells Really Work: Multiple Layers of Regulations

Whatever we may think of it, we live in an intellectual frame dominated by the primacy of the genome, as established in 1941 by Beadle and Tatum [21], where a direct causal link between genes and proteins (enzymes at that time) was established (mRNA was to be discovered two decades later [22]). This primacy of the genome lies on very solid ground. At the organismal level, if the gene is not transcribed and translated (for protein-coding genes, of course), nothing will happen at all. This does not mean that the reverse proposition (i.e., once the gene is transcribed, the die is cast) is true.

From a practical standpoint, we can define transcription *sensu lato* as all the processes going from transcription *stricto sensu* to the release of the final, mature mRNA(s). In an omics perspective, this is a

very convenient definition, as this is the operational definition of transcriptomics, which handles and measures mature mRNAs.

It shall not be forgotten that, as their very name says, mRNAs are just messenger molecules. Once they have been produced, they must be translated into proteins, and the proteins mature to become fully active. From a gene perspective, beyond transcription *sensu lato* lie two additional levels of regulations: translational control and post-translational control.

Translational control may be described as the poor cousin of regulations. From an energy management standpoint, translational control may look like nonsense. Why spend energy on making mRNAs that are not used? There are some physiological situations where it makes perfect sense, such as gametogenesis/early development, and it is no surprise that translational control has been evidenced and studied in detail in this frame [23,24]. However translational control does exist in somatic cells (e.g., in references [25–31]). In this case, translational control is the parallel of precautionary expenditures for societies: how much are you prepared to spend just in order to be ready for the difficultly predictable? The fact that translational control occurs in all types of organisms, even in prokaryotes where transcription and translation are coupled [32,33], speaks in favor of the importance of this mechanism.

Post-translational control has been studied much more extensively in the last few decades, mostly through the prism of post-translational modifications (PTMs) (*stricto sensu* the grafting/ungrafting of chemical moieties to the polypeptide backbone of proteins), although protein cleavage and degradation is also an important part of post-translational control. Literally dozens of different post-translational modifications have been described, and it is fair to say that proteomics has been instrumental to the discovery of many of them. Description of even the most common types of modifications is not in the scope of the present paper, and we would prefer to concentrate on them as a regulatory layer. The problem with PTMs is that their consequences are not univocal. When a gene is transcribed, the associated function will eventually be turned on; when transcription stops, it will eventually fade away. For translational control, the situation is also rather simple. When an RNA is a special type of RNP (ribonucleoparticle), it is translationally silent; when this particle is altered, it becomes translated [34]. This straightforwardness does not hold true for PTMs. For example phosphorylation of a protein can be an activating or a deactivating event [35], and the same holds true for acetylation [36]. Thus, the description of the modification landscape of proteins cannot be directly converted into functional inferences. Moreover, it should be kept in mind that protein function modulation can take place by many other, and even more subtle mechanisms, such as the binding of allosteric regulators of changes in loosely bound prosthetic groups, as in the example of cytosolic aconitase where the loss of a single iron atom in an iron-sulfur cluster triggers an activity change from aconitase to RNA-binding [37].

When trying to figure out the relevance of the various omics in deciphering changes in cell functioning, the question turns into weighing these different layers of regulations in the cellular life. We believe it is fair to confess that today, we cannot do such a comparison. Once again, we would like to use a metaphor to represent the authors' views, and this time, it will be a pastry metaphor. With eggs, flour, sugar, butter, baking powder, and water, you can make many different pastries: pies, cakes, muffins, sponge cakes, puff pastries, etc. The only differences lie in the proportions and handling of the ingredients. In our view, transcriptomics is the shopping list, bottom-up proteomics the actual ingredients lists, and top-down proteomics the actual recipe.

3. Position of Bottom-up Proteomics

3.1. The Lost Battle of Proteomics: Comprehensiveness

In the wake of the success of the genomic projects, all functional omics have dreamt of comprehensiveness. Because we live in this genetic-primacy world, comprehensiveness shall be understood as the fraction of the genome detectable and measurable by the functional omics, and this view is shared for at least transcriptomics and proteomics, where direct inferences and predictions

can be made from the genome. With the development of RNA sequencing, it is fair to say that transcriptomics has reached comprehensiveness [38], while proteomics cannot claim this status, even after more than a decade of extensive effort. There are a series of reasons explaining this situation.

The first reason lies in the chemical diversity issue. Chemically speaking, mRNAs are always and only a peculiar type of poly(sugar-phosphate), whatever protein they encode; thus, “one size fits all” for the chemical handling of RNA molecules in RNA sequencing. By contrast, proteins come with very different sizes, isoelectric points, hydrophobicities, foldings, etc., so that there is no universal method to handle all of them. This aspect is particularly severe for top-down proteomics, where complete proteins are to be analyzed, but much less for bottom-up proteomics, where peptides have a much lower chemical diversity. Moreover, one protein produces several peptides, so that it can always be reasonably hoped that some of these peptides will behave nicely in the proteomic analysis and thus act as a proxy for the protein. This chemical diversity also makes it difficult to develop universally applicable analytical methods for proteomics corresponding to the sequencing methods in the nucleic acid world.

The second reason lies in the dynamic range. In a mammalian cell, the dynamic expression range of mRNAs spans four orders of magnitude [39], while the protein dynamic expression range of mRNAs spans six orders of magnitude. When it comes to analyzing the needle and the haystack at the same time, these additional orders of magnitude add significant complications.

Another difficulty that shall be mentioned is linked to peptide sharing in closely related proteins that belong to multigene families. In this case, the difficulty comes from the fact that the genetic code is degenerate. Thus, different sequence strings on RNAs can code for completely identical sequence strings at the peptide level, meaning, in turn, that gene products that are easily differentiated at the RNA level may be much more difficult to differentiate at the peptide level.

Finally, a more practical (but very important) reason lies in the amplification that is possible for nucleic acids, e.g., through Reverse Transcription-Polymerase Chain Reaction (RT)-PCR, which strongly decreases the sensitivity issues, allowing investigations to reach the single cell level [40].

3.2. What Room for Bottom-up Proteomics in a Transcriptomic World?

With this obvious and salient issue of comprehensiveness, what is the use of running a proteomic experiment when RNA sequencing can do the job faster and wider? The question is more acute for bottom-up proteomics because this proteomic setup and transcriptomics deliver results at exactly the same level, i.e., the gene product. First of all, there are all the applications where the mRNA level is not relevant or absent, such as body fluids, subcellular proteomics, protein complexes, etc. Then, the question stands at the cellular level (or above) when both transcriptomics and proteomics can be carried out. In this frame, which represents the majority of proteomic papers, the only interest of bottom-up proteomics is to be more relevant, functionally speaking, than transcriptomics. By regression, this question steps back to the correlation between RNA and protein levels. This issue of correlation can be traced to the prehistory of proteomics [41]. With the development of deeper methods, and thus playing with bigger datasets, this question has been the subject of heated debate [39,42–44]. The apparent stakes are very simple: if the correlation between RNA and protein levels is high, there is no need to run bottom-up proteomic experiments. If the correlation is low, then it is justified to run proteomic experiments. It seems to us, however, that unless this correlation is the exact scientific question under study, looping back to the question of translational control, this correlation index is not the adequate indicator. In differential proteomics, where the question is to compare two different states by making quantitative comparisons of mRNA or proteins levels, each mRNA can have its own private “translation efficiency factor.” These factors may vary greatly, making a poor overall correlation, but they can be steady when conditions change, so that for each mRNA-protein pair, the change of the mRNA abundance may reflect (or not) the change in protein abundance. In a sense, this question is the first derivative of the correlation question, dealing not with heights but with slopes. The stakes therefore switch to the question of the correlation of the changes in abundances at the mRNA and

protein levels. In the literature, there are old and striking examples where the mRNA and protein clocks do not all turn at the same speed, leading, in fact, to opposite changes [45]. The question is to determine whether such discrepancies are rare or frequent. This question has been addressed by Ning et al. [46], and their results are quite clear. In approximately three quarters of the cases, the changes observed for mRNA levels and protein levels went in the same direction. Put more bluntly, this means in turn that one change call out of four made by transcriptomics is completely wrong, amply justifying doing bottom-up proteomics.

4. Escaping from the Transcription Sway: Top-Down Proteomic Returns

The paradigm that is conveyed by transcriptomics and bottom-up proteomics is indeed simple and thus popular: amounts make everything, and their interpretation is direct and simple (see Section 2.2), which is intellectually comfortable. However, a few decades of research on PTMs, not only by proteomics, has shown how important PTM can be in cellular life, going from everyday metabolic control [47] to physiologically or pathophysiologically important decisions mediated by the receptor-kinases signaling cascades (e.g., in Sasi et al. [48]). In this context, it becomes important to really take PTMs into account, not only by mapping the modification sites, but by taking into account how the modifications are integrated on complete proteins creating various proteoforms, which implies in turn to resort to top-down proteomics.

4.1. Combinatorial PTMs, A Separation Challenge

The existence of multiple modifications on proteins has been made visible since the earliest days of 2DGE proteomics, as multiple spots of different pI, same molecular weight, and coming from the same protein. As soon as a multiple modified spot can be detected, the question of PTM combination arises. It shall be reminded at this point that spot separation in the isoelectric focusing (IEF) dimension on 2D gels counts only charges. Thus, every modification that induces the same delta in charge will drive the protein at the same place on the 2D gel. As an example, lysine acetylation and serine phosphorylation both result in a -1 charge, and will drive a protein spot at the same place. With the onset of high sensitivity and resolution mass spectrometry, it has become possible to map all modifications that can be present in a single protein spot, and the results can be daunting [49,50]. Calculations made from the theoretical and observed pI render possible the determination of the number of modifications, but any combination of them in the list of detected modifications may occur. This obviously shows that the separation afforded by 2D gels is clearly suboptimal, but even though, the degree of separation possible and the ability to take into account modifications, even without knowing their mapping, can be very valuable, biologically speaking (e.g., in references [51–53]). However it is our opinion that many users of 2DGE proteomics do not use fully this ability of the setup, and tend to confuse the part (a change in one form of the protein) for the whole (a change in the total amount of the protein). Because of the importance of PTMs, there is obvious benefit to use a top-down-oriented proteomic setup when possible, either 2DGE based or top-down MS-based. This often highlights interesting PTM-based changes, which are worth validating because they do correspond to cellular regulations, but not at the total protein level.

4.2. Proteisobars, the Ultimate Nightmare in Proteomics

In terms of PTM combination, the resolution afforded in MS-based top-down proteomics goes a step further than the one afforded by 2D gels. For example, a lysine acetylation and a serine phosphorylation are not distinguished in a 2D gel but are in a top-down MS. Despite this improvement, proteomics still faces the wall of proteisobars, i.e., proteoforms that have exactly the same mass, because they bear the exact same combination of modifications, just at different sites. Proteisobars have the same mass (not separable by mass or electrophoresis), the same pI (not separable by IEF unless maybe at extremely high resolution and in a native form [54]), and similar hydrophobicity (not separable on reverse phase chromatography). When introduced simultaneously (because of this lack of separation)

in a mass spectrometer and analyzed in a top-down mode, the position of the modifications will be solved but not the relative amounts of the different isobaric variants.

5. “Mehr Licht”: How to Move Proteomic Forward

It is said that “mehr Licht” (more light) were the final words pronounced by the famous German poet and writer Johann Wolfgang von Goethe on his deathbed. Indeed, we need to bring more light to the proteomic lampposts, and especially to those positioned in very dark areas such as top-down proteomics.

5.1. Progress in Bottom-up Proteomics

For bottom-up proteomics, the progresses made on the mass spectrometers themselves have moved the frontier from mass spectrometry to peptide separation. In fact, the glass ceiling of peptide numbers and details in the characterization of the proteins can be moved upward by putting more effort into peptide separation. More separation necessarily means more time spent to gain these details. Today, in addition to time, this improved separation often also means a good deal of craftsmanship and dedication, as exemplified by giant peptide separation columns [55], but craftsmanship was also needed in the early days of nanoLC ESI MS/MS, which has become a daily routine in many proteomic platforms. Moreover, every additional separation step also holds the risk of introducing higher variability in the analysis workflow.

Another current trend in bottom-up proteomics is to increase the overall sensitivity to be able to work on very small sample sizes. Once again, this movement parallels the one observed in transcriptomics with single-cell transcriptomics [40]. In this quest, nucleic acids have the tremendous advantage of being amplifiable, which is not true in the case of proteins. In fact, the only PCR that can be experienced in proteomics is Peptide Content Reduction, which is far less positive. Nevertheless, various protocols have been developed to handle very low amounts of proteins (e.g., in references [56–58]) in order to face this challenge. In the same trend, the ability to detect proteins at very low levels is still actively investigated, sometimes by means that do not rely on mass spectrometry (e.g., in references [59,60]), and there is no doubt that bottom-up proteomics will perform routinely on a scale close to 1000 mammalian cells in the very near future.

Apart from the analytical developments, there is active research in the numerical and statistical treatment of shotgun proteomics data. In the bottom-up proteomic process, the first issue that arises is that of missing data, which by the way should be decreased if more time was devoted to cleaner peptide separations [55]. This missing data issue has been recognized to be of major importance, and various strategies have been implemented to make “reasonable imputation” of the missing data (e.g., in references [61,62]). The authors of the present paper may just betray their old-school scientist characters by stating that performing data imputation is just an elegant way of creating data from nothing and validating this process *ex post* by showing that it improves the subsequent statistical analysis. In our opinion, a better (although probably more time and resource consuming) workflow producing less missing data would be highly preferable.

When trying to make sense of proteomic datasets, whichever the proteomic setup used, statistical analyses are mandatory. This need reflects the fuzziness of the data, coming both from the biology and from the proteomic measurements themselves, which prevent us from telling outright the good, the bad, and the ugly. Proteomic data in particular and omic data in general really stretch the power of statistics, as omic experiments are systematically underpowered compared to the richness of the data, making multiple comparisons a cardinal issue. Here again, numerous strategies have been proposed to get the most of proteomic data (e.g., in references [63,64]), and many of them such as the popular volcano plot rely on a combination of fold change threshold and statistical tests results [64]. It should be recalled, however, that a fold change threshold is by definition arbitrary and cannot be applied as such [65]. Just to give an example, shall we implement the same fold change threshold for a very abundant protein that is documented in the shotgun proteomic experiment by more than 10 abundant

peptides and for a rare protein that is barely quantified by two low-abundance peptides? Besides the proteomic measurement itself, the cost function for the cell should also be considered. Increasing by 25% the level of a major metabolic protein may represent an awful lot of ATP invested, compared to doubling a rare protein that, in addition, can be easily silenced as needed by a judicious PTM. As we do not know the details of cell functioning yet, it appears safer not to play with fold changes, even if this results in longer lists of modulated proteins at the end [65].

Another trend in shotgun proteomic datasets treatment is to use a false discovery rate (FDR) threshold. This FDR is usually set at low values (5% or 1%) [63] in order to limit the number of false positives. It should, however, be kept in mind that, when used at this stage of interpretation of protein abundance changes, FDR in fact represents the likelihood with which a validation experiment based on the protein abundances changes detected by proteomics would fail. As biologists, if we try to figure out the failure rate that we accept or are ready to accept in our targeted biology experiments, it will clearly fall well above 5% and should rather be around 20–30%. As tLangley and Mayr rightly state in [63], there is always a tradeoff between true and false positives, and using a low FDR is, in a sense, accepting the throwing out of some babies with the bathwater.

In a nutshell, we believe that we shall not forget that, in our applications, statistical analysis should be viewed as nothing other than risk management for the validation experiments that should (or rather must) follow. Thus, thresholds and FDR should be used (or not) and tuned according to this perspective, and not only to reduce the size of the “positive list” coming out of a proteomic experiment.

It shall not be forgotten, however, that on top of the analytical glass ceiling, there is still the steel ceiling induced by the loss of connectivity between peptides in bottom-up proteomics, which will be an insuperable limitation when details of the cellular regulations are to be investigated. The two functional omic techniques that offer the highest numbers of genes covered are transcriptomics and bottom-up proteomics, but both operate at the gene level products, and therefore do not take into account a large part of cellular regulations. It is impressive to see how this point is neglected, which represents a very good example of cognitive availability bias. What is now presented in recent papers as “progress in proteomics” (and is in fact progress in bottom-up proteomics) only focuses on sensitivity and number issues (e.g., in references [59,60,66]) and thus only on protein identification and never on protein characterization, while this would be the only way to integrate the cellular regulations that are currently not taken into account.

5.2. Progress in 2DGE Proteomics: Do and Dare

In our opinion, 2DGE proteomics shall be viewed as a mature, efficient, and cost-efficient way to enter the top-down proteomic field, despite the known limitations of 2D gels for the analysis of hydrophobic [67,68], high molecular weight proteins, and proteins with extreme isoelectric points [69]. Although it is quite clear that the resolving power of 2DGE is insufficient compared to the diversity of proteoforms [49,70], 2DGE is an efficient tool for at least a pre-separation of isoforms. At the separation level, although it is true that narrow gradients can afford a tremendous resolution [54], their systematic use would be extremely resources-consuming and is therefore not widespread, to say the least. Furthermore, the high resolution of narrow pH gradients can deploy only if the proteins are soluble enough around their pI. Otherwise, isoelectric precipitation induces streaking that outweighs the theoretical increase in resolution. Nevertheless, relatively rare proteins can be analyzed by 2DGE proteomics, provided that adequate prefractionation and narrow pH gradients are used [71,72].

However, it seems to us [69] and to others (e.g., in Zhan et al. [70]) that the coupling of high-sensitivity mass spectrometry with 2DGE has a lot to offer in proteomics at the proteoform level. Indeed, high-sensitivity mass spectrometry offers a wealth of peptide information for each 2D spot analyzed. This information can be handled from two different perspectives.

The first perspective focuses on the fact that one spot can contain many different proteins [73]. Evaluating the real proportions of the different proteins present in a single spot from the mass spectrometry data is not as straightforward as it may seem, as the indices used to do so are

questionable [42]. Furthermore, going in that direction when performing differential proteomics may be a risky business. It should be kept in mind that proteins do not have the same exact positions from one 2D gel to another. To compensate for these migration changes, we use the staining pattern as a guide to recognize the same spot across a series of gels, and compensate for the migration changes. If we claim to analyze even the minor components of a 2D spot in a comparative proteomic experiment, we make the implicit assumption that the migration alterations of these minor components will follow exactly the migration alterations of the major components, which we detect through the staining. The authors believe that this is a rather risky bet.

The second perspective focuses on the fact that one spot contains many different modified peptides for the same protein. Up to now, this information has been used mostly to describe the modifications that can take place on a protein without the need to implement prior knowledge of the modifications and using the actual MS/MS spectra to deduce the nature and position of the modifications [49,50]. We believe that this information on modified peptides can be used in differential proteomic experiments to follow the quantitative evolution of the different proteoforms present in a 2D spot between situations of interest. We are aware that this description may lack precision. However, it may be of physiological interest, as exemplified by following cofilin phosphorylations [74]. It should be emphasized that, in this setup, the intensities of identical peptides are compared between spots, so that the peptide inhomogeneity issue present when comparing different peptides between them is absent in this configuration.

Thus, our opinion is that many (if not most) users of 2DGE proteomics do not fully use the power of this proteomic setup, especially for PTMs and in the framework of top-down proteomics. We believe that, when coupled with high sensitivity mass spectrometry, 2DGE proteomics has a lot to offer, and even better at a very reasonable price.

It also seems to us that 2DGE could be more used as a prefractionation tool upstream of MS-based top-down proteomics by eluting complete proteins and injecting them in the mass spectrometer. This is not as technically difficult as it seems, as western blotting has amply demonstrated the possibility of eluting proteins of any size out of a SDS gel (e.g., in Fritz et al. [75]). In the times of Edman sequencing, where the sensitivities were very far from what they are now, scientists have developed methods enabling to concentrate proteins from several 2D spots [76]. However, there is still a strong issue with these techniques, which is linked to the presence of sodium dodecyl sulfate (SDS). SDS is indispensable for the solubilization of proteins and for high resolution separations, and there is indeed very limited flexibility in terms of detergent choice to maintain solubilization and separation [77]. This is why SDS is so widely used for protein elution from 2D gel spots, either on a solid [76] or liquid [5,78] support, or in so called 3D electrophoresis [79,80]. However, SDS must be removed prior to introduction in the mass spectrometer and/or the chromatographic column [81]. This process is plagued by losses for the simple reason that proteins dissolved in SDS are denatured and show a high propensity to precipitate as soon as the SDS is removed, unless another solubilizer of denatured proteins is present, such as urea [82]. Thus, urea-based electrophoresis and especially acid urea electrophoresis (e.g., in Kitta et al. [83]) may be of interest as an interface between classical 2D gels and a top-down mass spectrometry setup, as an acidic pH will strongly limit protein carbamylation. It should also be kept in mind that the SDS electrophoresis process itself may induce oxidative protein modifications, although this has not been described to date in the literature [6]. However some precautions may be worth taking [84], including the replacement of persulfate-based initiator systems by photopolymerization systems [84,85].

5.3. Progress in MS-Based Top-Down Proteomics: Is Virtue in the Middle?

For this proteomic setup, this paper will focus on the separation challenges upstream of the top-down mass spectrometry itself and not on the MS data production and interpretation. More very valuable information on this can be found in references [86–90]. Indeed, for maximizing the number of proteins that can be analyzed by MS-based top-down proteomics, some protein separation must be applied upfront. As proteins are more efficiently separated by electrophoresis than by

chromatography [91], this separation is performed by a low resolution variant of 2D electrophoresis [8], which still poses the problem of SDS removal [81]. Thus, it might be of interest to reverse the order of the separations and start by the SDS electrophoresis separation, as described in rare variants of 2DGE [92,93]. For an easy interface with the Liquid Chromatography-Mass Spectrometry (LC/MS) system, a good setup for the second dimension IEF separation could be to use the so-called offgel separation [94], where proteins are collected directly into a liquid phase.

Although this may improve both the quantitative yield and the number of proteins amenable to analysis, the proteisobars issue will remain unaddressed. On the one hand, the complete protein cannot be separated from each other, and on the other hand, cutting them into small pieces for bottom-up analyses will make the reconstruction of the proteoforms impossible. So the only remaining solution is the so-called middle-down approaches [95], where proteins are cut into polypeptides in the 5–30 kDa range, which are then separated and analyzed by tandem mass spectrometry (MS/MS). The hope is that, by knowing the mass of the total protein, and thus its number and types of modifications, the different chunks of proteins analyzed by the middle-down approaches will allow the user to differentiate the various proteisobars. Ideally, at least two overlapping sets of protein chunks should be produced and analyzed, so that the complete proteins can be reconstructed, exactly as people were sequencing proteins in the first days of Edman sequencing, i.e., before the age of DNA sequencing.

The difficulty resides indeed in the production of the protein fragments of the adequate size. Ideally, this would mean a protease cutting every 100 amino acids on average, i.e., a protease tightly recognizing a pair of amino acids. Although many proteases remain to be discovered, this type of cutting does not seem to be widespread. More widespread are “digestive” proteases that cleave after one (or one class) of amino acids, such as trypsin, chymotrypsin, Lys-C, Lys-N, or the same, or specific proteases that recognize a larger site (e.g., thrombin, TEV protease) but thus do not cleave frequently enough for middle-down proteomics. Furthermore, the ideal protease should be able to work in (at least moderately) denaturing environments, as proteomics usually works with generic protocols and thus on denatured proteins. There have been efforts to find such a protease [96], but they have not yet ended in a practical protocol. Thus, one possibility could be to resort to even older methods such as limited proteolysis [97], in which the protein of interest is exposed for a short time to a low concentration of a “digestive” protease and the reaction is stopped by the electrophoretic separation of the starting protein, the fragments, and the proteases. Although the method is generic and uses reagents and separations that are commonplace in proteomics, it requires optimization for every protein to be analyzed, so that it is likely that this approach will be kept as a last resort method.

6. As a Conclusion: Is Proteomics Enlightening?

In this very short conclusive section, we would like to close the parenthesis that we opened at the beginning of this paper. It seems to us that progress in proteomics has been largely driven by the steady and spectacular improvements in the performances of mass spectrometers. This was accompanied by an increasing popularity of counting identifiable and quantifiable peptides and proteins without keeping the reality of the biological regulations in mind. However, we get the impression that it will no longer be the case, and that the bottleneck will swing back to the separation prior to the mass spectrometry, i.e., from science in the vacuum to science in water. Beside the avenues for technical improvement that we have just touched in the previous section, we strongly believe that the key issue, as we mentioned in the introduction of this paper, is to interpret the omic data (and in our case the proteomic data) at exactly the right level, without intellectual shortcuts that are often true misinterpretations. Only at this price will the proteomic literature deliver on its promises, and the best way to avoid overinterpretations will always bear the name of validation [98].

Author Contributions: Conceptualization, K.M. and T.R.; writing—original draft preparation, K.M. and T.R.; writing—review and editing, K.M. and T.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

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

References

1. Wasinger, V.C.; Cordwell, S.J.; Cerpa-Poljak, A.; Yan, J.X.; Gooley, A.A.; Wilkins, M.R.; Duncan, M.W.; Harris, R.; Williams, K.L.; Humphery-Smith, I. Progress with gene-product mapping of the Mollicutes: *Mycoplasma genitalium*. *Electrophoresis* **1995**, *16*, 1090–1094. [[CrossRef](#)] [[PubMed](#)]
2. Rogowska-Wrzesinska, A.; Le Bihan, M.C.; Thaysen-Andersen, M.; Roepstorff, P. 2D gels still have a niche in proteomics. *J. Proteom.* **2013**, *88*, 4–13. [[CrossRef](#)] [[PubMed](#)]
3. Oliveira, B.M.; Coorssen, J.R.; Martins-de-Souza, D. 2DE: The phoenix of proteomics. *J. Proteom.* **2014**, *104*, 140–150. [[CrossRef](#)] [[PubMed](#)]
4. Henzel, W.J.; Billeci, T.M.; Stults, J.T.; Wong, S.C.; Grimley, C.; Watanabe, C. Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases. *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 5011–5015. [[CrossRef](#)]
5. Claverol, S.; Burlet-Schiltz, O.; Girbal-Neuhausser, E.; Gairin, J.E.; Monsarrat, B. Mapping and structural dissection of human 20 S proteasome using proteomic approaches. *Mol. Cell. Proteom. MCP* **2002**, *1*, 567–578. [[CrossRef](#)]
6. Claverol, S.; Burlet-Schiltz, O.; Gairin, J.E.; Monsarrat, B. Characterization of Protein Variants and Post-Translational Modifications: ESI-MSⁿ Analyses of Intact Proteins Eluted from Polyacrylamide Gels. *Mol. Cell. Proteom.* **2003**, *2*, 483–493. [[CrossRef](#)]
7. Kelleher, N.L. Top-down proteomics. *Anal. Chem.* **2004**, *76*, 197A–203A. [[CrossRef](#)]
8. Tran, J.C.; Zamdborg, 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*, 254–258. [[CrossRef](#)]
9. Yates, J.R.; McCormack, A.L.; Schieltz, D.; Carmack, E.; Link, A. Direct analysis of protein mixtures by tandem mass spectrometry. *J. Protein Chem.* **1997**, *16*, 495–497. [[CrossRef](#)]
10. The Consortium for Top Down Proteomics; Smith, L.M.; Kelleher, N.L. Proteoform: A single term describing protein complexity. *Nat. Methods* **2013**, *10*, 186–187. [[CrossRef](#)]
11. Eng, J.K.; McCormack, A.L.; Yates, J.R. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **1994**, *5*, 976–989. [[CrossRef](#)]
12. Perkins, D.N.; Pappin, D.J.C.; Creasy, D.M.; Cottrell, J.S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **1999**, *20*, 3551–3567. [[CrossRef](#)]
13. Craig, R.; Beavis, R.C. TANDEM: Matching proteins with tandem mass spectra. *Bioinformatics* **2004**, *20*, 1466–1467. [[CrossRef](#)]
14. Cox, J.; Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **2008**, *26*, 1367–1372. [[CrossRef](#)] [[PubMed](#)]
15. Nesvizhskii, A.I.; Aebersold, R. Interpretation of Shotgun Proteomic Data: The Protein Inference Problem. *Mol. Cell. Proteom.* **2005**, *4*, 1419–1440. [[CrossRef](#)]
16. Uszkoreit, J.; Maerkens, A.; Perez-Riverol, Y.; Meyer, H.E.; Marcus, K.; Stephan, C.; Kohlbacher, O.; Eisenacher, M. PIA: An Intuitive Protein Inference Engine with a Web-Based User Interface. *J. Proteome Res.* **2015**, *14*, 2988–2997. [[CrossRef](#)]
17. Audain, E.; Uszkoreit, J.; Sachsenberg, T.; Pfeuffer, J.; Liang, X.; Hermjakob, H.; Sanchez, A.; Eisenacher, M.; Reinert, K.; Tabb, D.L.; et al. In-depth analysis of protein inference algorithms using multiple search engines and well-defined metrics. *J. Proteom.* **2017**, *150*, 170–182. [[CrossRef](#)]
18. Uszkoreit, J.; Perez-Riverol, Y.; Eggers, B.; Marcus, K.; Eisenacher, M. Protein Inference Using PIA Workflows and PSI Standard File Formats. *J. Proteome Res.* **2019**, *18*, 741–747. [[CrossRef](#)]
19. Miller, R.M.; Millikin, R.J.; Hoffmann, C.V.; Solntsev, S.K.; Sheynkman, G.M.; Shortreed, M.R.; Smith, L.M. Improved Protein Inference from Multiple Protease Bottom-Up Mass Spectrometry Data. *J. Proteome Res.* **2019**, *18*, 3429–3438. [[CrossRef](#)]
20. Pfeuffer, J.; Sachsenberg, T.; Dijkstra, T.M.H.; Serang, O.; Reinert, K.; Kohlbacher, O. EPIFANY: A Method for Efficient High-Confidence Protein Inference. *J. Proteome Res.* **2020**, *19*, 1060–1072. [[CrossRef](#)]

21. Beadle, G.W.; Tatum, E.L. Genetic Control of Biochemical Reactions in *Neurospora*. *Proc. Natl. Acad. Sci. USA* **1941**, *27*, 499–506. [[CrossRef](#)] [[PubMed](#)]
22. Jacob, F.; Monod, J. Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.* **1961**, *3*, 318–356. [[CrossRef](#)]
23. Richter, J.D. Translational control during early development. *BioEssays News Rev. Mol. Cell. Dev. Biol.* **1991**, *13*, 179–183. [[CrossRef](#)] [[PubMed](#)]
24. Lai, F.; King, M.L. Repressive translational control in germ cells. *Mol. Reprod. Dev.* **2013**, *80*, 665–676. [[CrossRef](#)]
25. Imaizumi-Scherrer, M.T.; Maundrell, K.; Civelli, O.; Scherrer, K. Transcriptional and post-transcriptional regulation in duck erythroblasts. *Dev. Biol.* **1982**, *93*, 126–138. [[CrossRef](#)]
26. Thomas, G.; Thomas, G. Translational control of mRNA expression during the early mitogenic response in Swiss mouse 3T3 cells: Identification of specific proteins. *J. Cell Biol.* **1986**, *103*, 2137–2144. [[CrossRef](#)]
27. Pramanik, S.K.; Bag, J. Translation of an mRNA in rat L6 muscle cells is regulated within the cell cycle. *Eur. J. Biochem.* **1987**, *170*, 59–67. [[CrossRef](#)]
28. Ranjan, M.; Tafuri, S.R.; Wolffe, A.P. Masking mRNA from translation in somatic cells. *Genes Dev.* **1993**, *7*, 1725–1736. [[CrossRef](#)]
29. Schafer, S.; Adami, E.; Heinig, M.; Rodrigues, K.E.C.; Kreuchwig, F.; Silhavy, J.; Van Heesch, S.; Simaite, D.; Rajewsky, N.; Cuppen, E.; et al. Translational regulation shapes the molecular landscape of complex disease phenotypes. *Nat. Commun.* **2015**, *6*. [[CrossRef](#)]
30. Esnault, S.; Shen, Z.-J.; Malter, J.S. Protein Translation and Signaling in Human Eosinophils. *Front. Med.* **2017**, *4*. [[CrossRef](#)]
31. Tahmasebi, S.; Amiri, M.; Sonenberg, N. Translational Control in Stem Cells. *Front. Genet.* **2019**, *9*. [[CrossRef](#)] [[PubMed](#)]
32. Storz, G.; Opdyke, J.A.; Zhang, A. Controlling mRNA stability and translation with small, noncoding RNAs. *Curr. Opin. Microbiol.* **2004**, *7*, 140–144. [[CrossRef](#)] [[PubMed](#)]
33. Picard, F.; Dressaire, C.; Girbal, L.; Coccagn-Bousquet, M. Examination of post-transcriptional regulations in prokaryotes by integrative biology. *C. R. Biol.* **2009**, *332*, 958–973. [[CrossRef](#)] [[PubMed](#)]
34. Evdokimova, V.; Ruzanov, P.; Anglesio, M.S.; Sorokin, A.V.; Ovchinnikov, L.P.; Buckley, J.; Triche, T.J.; Sonenberg, N.; Sorensen, P.H.B. Akt-Mediated YB-1 Phosphorylation Activates Translation of Silent mRNA Species. *Mol. Cell. Biol.* **2006**, *26*, 277–292. [[CrossRef](#)]
35. Humphrey, S.J.; James, D.E.; Mann, M. Protein Phosphorylation: A Major Switch Mechanism for Metabolic Regulation. *Trends Endocrinol. Metab. TEM* **2015**, *26*, 676–687. [[CrossRef](#)]
36. Xiong, Y.; Guan, K.-L. Mechanistic insights into the regulation of metabolic enzymes by acetylation. *J. Cell Biol.* **2012**, *198*, 155–164. [[CrossRef](#)]
37. Haile, D.J.; Rouault, T.A.; Tang, C.K.; Chin, J.; Harford, J.B.; Klausner, R.D. Reciprocal control of RNA-binding and aconitase activity in the regulation of the iron-responsive element binding protein: Role of the iron-sulfur cluster. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 7536–7540. [[CrossRef](#)]
38. McGettigan, P.A. Transcriptomics in the RNA-seq era. *Curr. Opin. Chem. Biol.* **2013**, *17*, 4–11. [[CrossRef](#)]
39. Schwanhäusser, B.; Busse, D.; Li, N.; Dittmar, G.; Schuchhardt, J.; Wolf, J.; Chen, W.; Selbach, M. Global quantification of mammalian gene expression control. *Nature* **2011**, *473*, 337–342. [[CrossRef](#)]
40. Macaulay, I.C.; Voet, T. Single Cell Genomics: Advances and Future Perspectives. *PLoS Genet.* **2014**, *10*, e1004126. [[CrossRef](#)]
41. Anderson, L.; Seilhamer, J. A comparison of selected mRNA and protein abundances in human liver. *Electrophoresis* **1997**, *18*, 533–537. [[CrossRef](#)] [[PubMed](#)]
42. Li, J.J.; Bickel, P.J.; Biggin, M.D. System wide analyses have underestimated protein abundances and the importance of transcription in mammals. *Peer J.* **2014**, *2*, e270. [[CrossRef](#)] [[PubMed](#)]
43. Liu, Y.; Beyer, A.; Aebersold, R. On the Dependency of Cellular Protein Levels on mRNA Abundance. *Cell* **2016**, *165*, 535–550. [[CrossRef](#)] [[PubMed](#)]
44. Fortelny, N.; Overall, C.M.; Pavlidis, P.; Freue, G.V.C. Can we predict protein from mRNA levels? *Nature* **2017**, *547*, E19–E20. [[CrossRef](#)]
45. Khochbin, S.; Gorka, C.; Lawrence, J.J. Multiple control level governing H10 mRNA and protein accumulation. *FEBS Lett.* **1991**, *283*, 65–67. [[CrossRef](#)]

46. Ning, K.; Fermin, D.; Nesvizhskii, A.I. Comparative analysis of different label-free mass spectrometry based protein abundance estimates and their correlation with RNA-Seq gene expression data. *J. Proteome Res.* **2012**, *11*, 2261–2271. [[CrossRef](#)]
47. Zhao, S.; Xu, W.; Jiang, W.; Yu, W.; Lin, Y.; Zhang, T.; Yao, J.; Zhou, L.; Zeng, Y.; Li, H.; et al. Regulation of Cellular Metabolism by Protein Lysine Acetylation. *Science* **2010**, *327*, 1000–1004. [[CrossRef](#)]
48. Sasi, M.; Vignoli, B.; Canossa, M.; Blum, R. Neurobiology of local and intercellular BDNF signaling. *Pflugers Arch.* **2017**, *469*, 593–610. [[CrossRef](#)]
49. Seo, J.; Jeong, J.; Kim, Y.M.; Hwang, N.; Paek, E.; Lee, K.-J. Strategy for Comprehensive Identification of Post-translational Modifications in Cellular Proteins, Including Low Abundant Modifications: Application to Glyceraldehyde-3-phosphate Dehydrogenase. *J. Proteome Res.* **2008**, *7*, 587–602. [[CrossRef](#)]
50. John, J.P.P.; Pollak, A.; Lubec, G. Complete sequencing and oxidative modification of manganese superoxide dismutase in medulloblastoma cells. *Electrophoresis* **2009**, *30*, 3006–3016. [[CrossRef](#)]
51. Martins-de-Souza, D.; Maccarrone, G.; Wobrock, T.; Zerr, I.; Gormanns, P.; Reckow, S.; Falkai, P.; Schmitt, A.; Turck, C.W. Proteome analysis of the thalamus and cerebrospinal fluid reveals glycolysis dysfunction and potential biomarkers candidates for schizophrenia. *J. Psychiatr. Res.* **2010**, *44*, 1176–1189. [[CrossRef](#)]
52. Sun, H.H.; Fukao, Y.; Ishida, S.; Yamamoto, H.; Maekawa, S.; Fujiwara, M.; Sato, T.; Yamaguchi, J. Proteomics Analysis Reveals a Highly Heterogeneous Proteasome Composition and the Post-translational Regulation of Peptidase Activity under Pathogen Signaling in Plants. *J. Proteome Res.* **2013**, *12*, 5084–5095. [[CrossRef](#)] [[PubMed](#)]
53. Niimori-Kita, K.; Tamamaki, N.; Koizumi, D.; Niimori, D. Matrin-3 is essential for fibroblast growth factor 2-dependent maintenance of neural stem cells. *Sci. Rep.* **2018**, *8*, 13412. [[CrossRef](#)] [[PubMed](#)]
54. Cossu, G.; Manca, M.; Strahler, J.R.; Manash, S.M.; Righetti, P.G. Detection of electrophoretically silent mutations by immobilized pH gradients. *J. Chromatogr. A* **1986**, *361*, 223–229. [[CrossRef](#)]
55. Iwasaki, M.; Sugiyama, N.; Tanaka, N.; Ishihama, Y. Human proteome analysis by using reversed phase monolithic silica capillary columns with enhanced sensitivity. *J. Chromatogr. A* **2012**, *1228*, 292–297. [[CrossRef](#)] [[PubMed](#)]
56. Hughes, C.S.; Foehr, S.; Garfield, D.A.; Furlong, E.E.; Steinmetz, L.M.; Krijgsveld, J. Ultrasensitive proteome analysis using paramagnetic bead technology. *Mol. Syst. Biol.* **2014**, *10*, 757. [[CrossRef](#)] [[PubMed](#)]
57. Zougman, A.; Selby, P.J.; Banks, R.E. Suspension trapping (STrap) sample preparation method for bottom-up proteomics analysis. *Proteomics* **2014**, *14*, 1000–1006. [[CrossRef](#)]
58. Sielaff, M.; Kuharev, J.; Bohn, T.; Hahlbrock, J.; Bopp, T.; Tenzer, S.; Distler, U. Evaluation of FASP, SP3, and iST Protocols for Proteomic Sample Preparation in the Low Microgram Range. *J. Proteome Res.* **2017**, *16*, 4060–4072. [[CrossRef](#)]
59. Swaminathan, J.; Boulgakov, A.A.; Hernandez, E.T.; Bardo, A.M.; Bachman, J.L.; Marotta, J.; Johnson, A.M.; Anslyn, E.V.; Marcotte, E.M. Highly parallel single-molecule identification of proteins in zeptomole-scale mixtures. *Nat. Biotechnol.* **2018**. [[CrossRef](#)]
60. Timp, W.; Timp, G. Beyond mass spectrometry, the next step in proteomics. *Sci. Adv.* **2020**, *6*, eaax8978. [[CrossRef](#)]
61. Webb-Robertson, B.-J.M.; Wiberg, H.K.; Matzke, M.M.; Brown, J.N.; Wang, J.; McDermott, J.E.; Smith, R.D.; Rodland, K.D.; Metz, T.O.; Pounds, J.G.; et al. Review, evaluation, and discussion of the challenges of missing value imputation for mass spectrometry-based label-free global proteomics. *J. Proteome Res.* **2015**, *14*, 1993–2001. [[CrossRef](#)]
62. Lazar, C.; Gatto, L.; Ferro, M.; Bruley, C.; Burger, T. Accounting for the Multiple Natures of Missing Values in Label-Free Quantitative Proteomics Data Sets to Compare Imputation Strategies. *J. Proteome Res.* **2016**, *15*, 1116–1125. [[CrossRef](#)]
63. Langley, S.R.; Mayr, M. Comparative analysis of statistical methods used for detecting differential expression in label-free mass spectrometry proteomics. *J. Proteom.* **2015**, *129*, 83–92. [[CrossRef](#)] [[PubMed](#)]
64. Gai Gianetto, Q.; Couté, Y.; Bruley, C.; Burger, T. Uses and misuses of the fudge factor in quantitative discovery proteomics. *Proteomics* **2016**, *16*, 1955–1960. [[CrossRef](#)] [[PubMed](#)]
65. Herrmann, A.G.; Searcy, J.L.; Le Bihan, T.; McCulloch, J.; Deighton, R.F. Total variance should drive data handling strategies in third generation proteomic studies. *Proteomics* **2013**, *13*, 3251–3255. [[CrossRef](#)] [[PubMed](#)]

66. Pappireddi, N.; Martin, L.; Wühr, M. A Review on Quantitative Multiplexed Proteomics. *ChemBioChem* **2019**, *20*, 1210–1224. [[CrossRef](#)]
67. Santoni, V.; Molloy, M.; Rabilloud, T. Membrane proteins and proteomics: Un amour impossible? *Electrophoresis* **2000**, *21*, 1054–1070. [[CrossRef](#)]
68. Rabilloud, T. Membrane proteins and proteomics: Love is possible, but so difficult. *Electrophoresis* **2009**, *30* (Suppl. 1), S174–S180. [[CrossRef](#)]
69. Marcus, K.; Lelong, C.; Rabilloud, T. What Room for Two-Dimensional Gel-Based Proteomics in a Shotgun Proteomics World? *Proteomes* **2020**, *8*, 17. [[CrossRef](#)]
70. 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. [[CrossRef](#)]
71. Pedersen, S.K.; Harry, J.L.; Sebastian, L.; Baker, J.; Traini, M.D.; McCarthy, J.T.; Manoharan, A.; Wilkins, M.R.; Gooley, A.A.; Righetti, P.G.; et al. Unseen proteome: Mining below the tip of the iceberg to find low abundance and membrane proteins. *J. Proteome Res.* **2003**, *2*, 303–311. [[CrossRef](#)] [[PubMed](#)]
72. Tang, H.-Y.; Speicher, D.W. Complex proteome prefractionation using microscale solution isoelectrofocusing. *Expert Rev. Proteom.* **2005**, *2*, 295–306. [[CrossRef](#)] [[PubMed](#)]
73. 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. [[CrossRef](#)] [[PubMed](#)]
74. Prudent, R.; Demoncheaux, N.; Diemer, H.; Collin-Faure, V.; Kapur, R.; Paublant, F.; Lafanechere, L.; Cianferani, S.; Rabilloud, T. A quantitative proteomic analysis of cofilin phosphorylation in myeloid cells and its modulation using the LIM kinase inhibitor Pyr1. *PLoS ONE* **2018**, *13*, e0208979. [[CrossRef](#)] [[PubMed](#)]
75. Fritz, J.D.; Swartz, D.R.; Greaser, M.L. Factors affecting polyacrylamide gel electrophoresis and electroblotting of high-molecular-weight myofibrillar proteins. *Anal. Biochem.* **1989**, *180*, 205–210. [[CrossRef](#)]
76. Gevaert, K.; Verschelde, J.L.; Puype, M.; Van Damme, J.; Goethals, M.; De Boeck, S.; Vandekerckhove, J. Structural analysis and identification of gel-purified proteins, available in the femtomole range, using a novel computer program for peptide sequence assignment, by matrix-assisted laser desorption ionization-reflectron time-of-flight-mass spectrometry. *Electrophoresis* **1996**, *17*, 918–924. [[CrossRef](#)]
77. Lopez, M.F.; Patton, W.F.; Utterback, B.L.; Chung-Welch, N.; Barry, P.; Skea, W.M.; Cambria, R.P. Effect of various detergents on protein migration in the second dimension of two-dimensional gels. *Anal. Biochem.* **1991**, *199*, 35–44. [[CrossRef](#)]
78. Harrington, M.G. Elution of protein from gels. *Methods Enzymol.* **1990**, *182*, 488–495. [[CrossRef](#)]
79. Butt, R.H.; Coorsen, J.R. Postfractionation for enhanced proteomic analyses: Routine electrophoretic methods increase the resolution of standard 2D-PAGE. *J. Proteome Res.* **2005**, *4*, 982–991. [[CrossRef](#)]
80. 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. [[CrossRef](#)]
81. Kachuk, C.; Stephen, K.; Doucette, A. Comparison of sodium dodecyl sulfate depletion techniques for proteome analysis by mass spectrometry. *J. Chromatogr. A* **2015**, *1418*, 158–166. [[CrossRef](#)] [[PubMed](#)]
82. Weber, K.; Kuter, D.J. Reversible denaturation of enzymes by sodium dodecyl sulfate. *J. Biol. Chem.* **1971**, *246*, 4504–4509. [[PubMed](#)]
83. Kitta, K.; Ohnishi-Kameyama, M.; Moriyama, T.; Ogawa, T.; Kawamoto, S. Detection of low-molecular weight allergens resolved on two-dimensional electrophoresis with acid-urea polyacrylamide gel. *Anal. Biochem.* **2006**, *351*, 290–297. [[CrossRef](#)] [[PubMed](#)]
84. Sun, G.; Anderson, V.E. Prevention of artifactual protein oxidation generated during sodium dodecyl sulfate-gel electrophoresis. *Electrophoresis* **2004**, *25*, 959–965. [[CrossRef](#)]
85. Lyubimova, T.; Caglio, S.; Gelfi, C.; Righetti, P.G.; Rabilloud, T. Photopolymerization of polyacrylamide gels with methylene blue. *Electrophoresis* **1993**, *14*, 40–50. [[CrossRef](#)]
86. Riley, N.M.; Sikora, J.W.; Seckler, H.S.; Greer, J.B.; Fellers, R.T.; LeDuc, R.D.; Westphall, M.S.; Thomas, P.M.; Kelleher, N.L.; Coon, J.J. The Value of Activated Ion Electron Transfer Dissociation for High-Throughput Top-Down Characterization of Intact Proteins. *Anal. Chem.* **2018**, *90*, 8553–8560. [[CrossRef](#)]
87. LeDuc, R.D.; Fellers, R.T.; Early, B.P.; Greer, J.B.; Shams, D.P.; Thomas, P.M.; Kelleher, N.L. Accurate Estimation of Context-Dependent False Discovery Rates in Top-Down Proteomics. *Mol. Cell. Proteom. MCP* **2019**, *18*, 796–805. [[CrossRef](#)]

88. 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. [[CrossRef](#)]
89. Fornelli, L.; Srzentić, K.; Toby, T.K.; Doubleday, P.F.; Huguet, R.; Mullen, C.; Melani, R.D.; Dos Santos Seckler, H.; DeHart, C.J.; Weisbrod, C.R.; et al. Thorough Performance Evaluation of 213 nm Ultraviolet Photodissociation for Top-down Proteomics. *Mol. Cell. Proteom. MCP* **2020**, *19*, 405–420. [[CrossRef](#)]
90. Ives, A.N.; Su, T.; Durbin, K.R.; Early, B.P.; Dos Santos Seckler, H.; Fellers, R.T.; LeDuc, R.D.; Schachner, L.F.; Patrie, S.M.; Kelleher, N.L. Using 10,000 Fragment Ions to Inform Scoring in Native Top-down Proteomics. *J. Am. Soc. Mass Spectrom.* **2020**, *31*, 1398–1409. [[CrossRef](#)]
91. Kellner, L.; Myere, H.E. *Microcharacterization of Proteins*, 1st ed.; Kellner, R., Lottspeich, F., Meyer, H.E., Eds.; Wiley: Weinheim, Germany, 1999; ISBN 978-3-527-30084-6.
92. Tuszyński, G.P.; Buck, C.A.; Warren, L. A two-dimensional polyacrylamide gel electrophoresis (PAGE) system using sodium dodecyl sulfate—PAGE in the first dimension. *Anal. Biochem.* **1979**, *93*, 329–338. [[CrossRef](#)]
93. Nakamura, K.; Okuya, Y.; Katahira, M.; Yoshida, S.; Wada, S.; Okuno, M. Analysis of tubulin isoforms by two-dimensional gel electrophoresis using SDS-polyacrylamide gel electrophoresis in the first dimension. *J. Biochem. Biophys. Methods* **1992**, *24*, 195–203. [[CrossRef](#)]
94. Ros, A.; Faupel, M.; Mees, H.; Oostrum, J.V.; Ferrigno, R.; Reymond, F.; Michel, P.; Rossier, J.S.; Girault, H.H. Protein purification by Off-Gel electrophoresis. *Proteomics* **2002**, *2*, 151–156. [[CrossRef](#)]
95. Garcia, B.A.; Siuti, N.; Thomas, C.E.; Mizzen, C.A.; Kelleher, N.L. Characterization of neurohistone variants and post-translational modifications by electron capture dissociation mass spectrometry. *Int. J. Mass Spectrom.* **2007**, *259*, 184–196. [[CrossRef](#)]
96. Wu, C.; Tran, J.C.; Zamdborg, L.; Durbin, K.R.; Li, M.; Ahlf, D.R.; Early, B.P.; Thomas, P.M.; Sweedler, J.V.; Kelleher, N.L. A protease for “middle-down” proteomics. *Nat. Methods* **2012**, *9*, 822–824. [[CrossRef](#)]
97. Cleveland, D.W.; Fischer, S.G.; Kirschner, M.W.; Laemmli, U.K. Peptide mapping by limited proteolysis in sodium dodecyl sulfate and analysis by gel electrophoresis. *J. Biol. Chem.* **1977**, *252*, 1102–1106.
98. Rabilloud, T.; Lescuyer, P. The proteomic to biology inference, a frequently overlooked concern in the interpretation of proteomic data: A plea for functional validation. *Proteomics* **2014**, *14*, 157–161. [[CrossRef](#)]



© 2020 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/>).