Quality control in mass spectrometry-based proteomics

Wout Bittremieux[1,2], **David L. Tabb**[3], **Francis Impens**[4,5,6], **An Staes**[4,5,6], **Evy Timmerman**[4,5,6], **Lennart Martens**[5,6,7], **Kris Laukens**[1,2]

[1]Department of Mathematics and Computer Science, University of Antwerp, 2020 Antwerp, Belgium; [2]Biomedical Informatics Research Center Antwerp (biomina), University of Antwerp/Antwerp University Hospital, 2650 Edegem, Belgium; [3]Division of Molecular Biology and Human Genetics, Stellenbosch University Faculty of Medicine and Health Sciences, Cape Town, 7505, South Africa; [4]VIB Proteomics Core, 9000 Ghent, Belgium; [5]VIB-UGent Center for Medical Biotechnology, 9000 Ghent, Belgium; [6]Department of Biochemistry, Faculty of Medicine and Health Sciences, Ghent University, 9000 Ghent, Belgium; [7]Bioinformatics Institute Ghent, Ghent University, 9052 Zwijnaarde, Belgium

Corresponding author: wout.bittremieux@uantwerpen.be, kris.laukens@uantwerpen.be

Abstract

Mass spectrometry is a highly complex analytical technique and mass spectrometry-based proteomics experiments can be subject to a large variability, which forms an obstacle to obtaining accurate and reproducible results. Therefore, a comprehensive and systematic approach to quality control is an essential requirement to inspire confidence in the generated results. A typical mass spectrometry experiment consists of multiple different phases including the sample preparation, liquid chromatography, mass spectrometry, and bioinformatics stages. We review potential sources of variability that can impact the results of a mass spectrometry experiment occurring in all of these steps, and we discuss how to monitor and remedy the negative influences on the experimental results. Furthermore, we describe how specialized quality control samples of varying sample complexity can be incorporated into the experimental workflow and how they can be used to rigorously assess detailed aspects of the instrument performance.

1 Introduction

Proteomics is a crucial domain in modern biological and biomedical research [3]. The current method of choice to identify and quantify complex protein samples is often liquid chromatography (LC) followed by mass spectrometry (MS). The importance of these techniques is exemplified by their use in large-scale research initiatives, such as the two recent attempts at providing a draft of the human proteome [74, 143] or the ongoing Human Proteome Project (HPP) by the Human Proteome Organization (HUPO) [80, 82, 92, 106, 107], where LC-MS techniques are used to identify, quantify, and characterize the human proteome.

As illustrated in figure 1, a bottom-up LC-MS experiment consists of multiple different stages. First, various sample preparation measures ensure that the biological samples are optimally suited for MS analysis. Typical steps include denaturation, reduction, and alkylation of the proteins. The denatured proteins are subsequently digested into peptides through proteolytic cleavage, typically by trypsin. Next, this peptide mixture is processed through liquid chromatography, which separates the peptides based on their hydrophobicity. After eluting from the column, the peptides are ionized to obtain a charge by electrospray ionization (ESI), and the derived spectra are generated in the mass spectrometer. Whether this is done in a data-dependent acquisition (DDA) or data-independent acquisition (DIA) manner, for a typical discovery experiment, in both approaches as many spectra as possible are generated through tandem mass spectrometry (MS/MS), whereas for a targeted experiment specific peptides of interest are exclusively monitored [114]. Finally, the generated spectral data is interpreted through various

bioinformatics means [103, 104]. Peptides can be identified from the mass spectra through sequence database searching [42], spectral library searching [56, 121], or *de novo* sequencing [97]; and the peptides can be mapped back to their originating proteins through protein inference [63]. Additionally, protein quantification [9, 55] and other advanced analyses may be performed.

As succinctly described above, performing a mass spectrometry experiment is an intricate process, and each of these different steps has to be optimized to acquire accurate and reproducible results. Unfortunately, despite the many recent technological and computational advances, the results of an experiment can still be subject to a large variability [2, 131]. As represented in figure 2, this variability can originate from multiple sources [115]: the different stages of an LC-MS experiment can each exhibit stochastic behavior and influence one another, contaminants can inadvertently be present [72], and the optimal computational interpretation is often not obvious [13]. Furthermore, instrument drift and sample degradation can introduce a longitudinal variability [14, 32]. Most notably, instrument interventions, such as a preventive maintenance, have a considerable influence upon the results [14]. Especially in regard to current large-scale studies this is of major importance, as measurements obtained at different times can only be correctly compared with each other if they were obtained under consistent and comparable conditions.

Therefore, in order to inspire confidence in the obtained results, it is of vital importance that appropriate quality assurance (QA) and quality control (QC) measures are taken to monitor and control the existing variability [4, 90, 94, 130]. A systematic approach to quality control makes it pos-
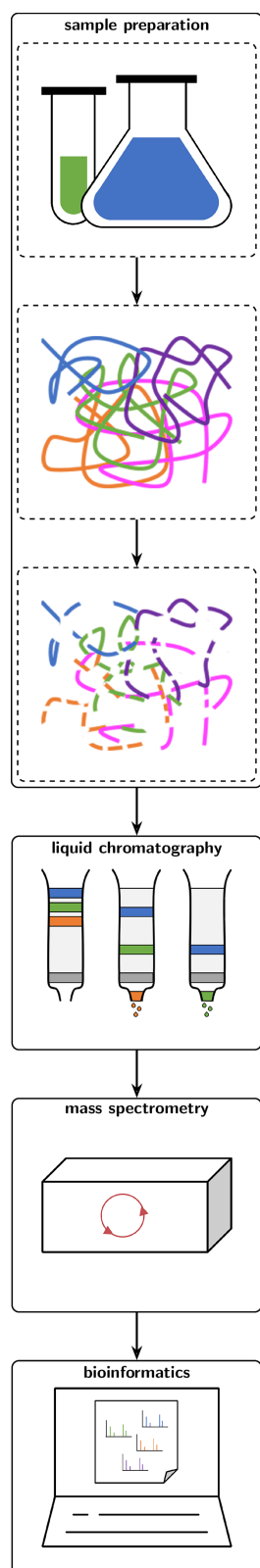
Figure 1: A typical LC-MS experiment consists of a sample preparation, a liquid chromatography, a mass spectrometry, and a bioinformatics stage. The sample preparation includes the proteolytic digestion of proteins into peptides. Next, consecutively the peptides are separated through liquid chromatography and measured through mass spectrometry. Finally, the acquired spectra are interpreted through bioinformatics means [21].
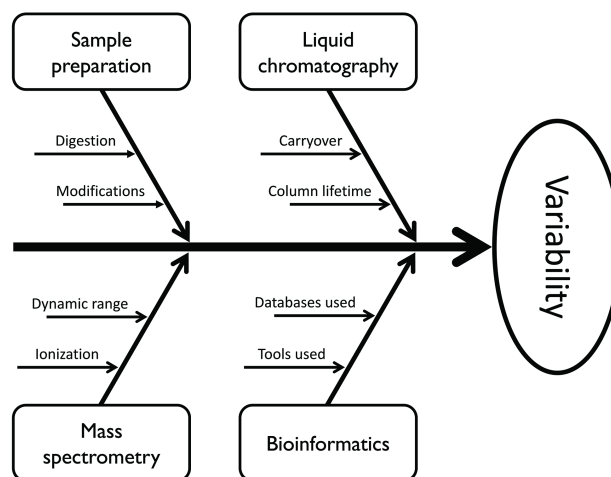
Figure 2: An Ishikawa diagram (non-exhaustively) highlighting some of the major sources of variability in each of the stages of an LC-MS experiment. These and other sources of variability will impact the results and should be considered in a comprehensive quality control workflow [20].

sible to objectively assess the quality of a mass spectrometry experiment, and empirical quantitative measures enable the intra-study, intra-laboratory, and inter-laboratory comparison of the performance of mass spectrometry runs [65]. As mentioned previously, these quality assessments are crucial to validate the results produced by long-term multi-site projects [32], such as the HUPO's Human Proteome Project [80, 82, 92, 106, 107] or the studies conducted by the National Cancer Institute (NCI): Clinical Proteomic Tumor Analysis Consortium (CPTAC) [122, 132, 150]. Also, the increased clustering of high-end research infrastructures such as mass spectrometers into institutional core facilities is pushing the need for systematic quality control [96]. Since core facilities are often accessible on a fee-for-service basis, users want their samples to be analyzed according the highest quality standards. In case results are negative or below expectations, QC data can be shared with the user to for instance exclude low instrument performance. Ideally, through systematic monitoring, proteomics core facilities can define quality thresholds for every step in their workflow(s). In case these thresholds are not met, targeted corrective measures such as replacing key reagents or cleaning the mass spectrometer can minimize down-time and avoid precious sample losses, thereby increasing the overall performance of the facility.

With so many different factors that can impact the experimental results it is important to carefully consider the various influences independently of each other. To this end, for example, a Pareto chart is a helpful visualization technique, as it can be used to represent the contribution of each individual factor to the total variability [16].

Here we will detail the origin of some common causes of variability that can influence the results of a mass spectrometry experiment and which steps should be taken to avoid them. Notably we will highlight how QC samples can be incorporated into the experimental workflow to systematically assess the instrument performance. Mass spectrom-

etry is an advanced and versatile technique, and it can be used for a wide variety of applications. As a result, there is no definitive consensus on which QC methodology to employ [93], nor is it possible to establish a single uniform approach to quality control. Instead we will broadly review some of the representative QC approaches, discuss general considerations, and show how these steps can be used to monitor the various elements of an MS workflow.

Several computational tools exist to assist in this QC monitoring. Although such tools are instrumental in being able to obtain detailed quality assessments in a longitudinal fashion we will not describe them in detail here. Instead we refer the interested reader to recent work by Bittremieux et al. [24] for an overview of available computational QC tools and an evaluation of their technical requirements and capabilities.

2 Managing LC-MS variability through quality control

First, we will briefly introduce the different types of QC samples that can be employed to monitor the performance of a mass spectrometry experiment and how these samples can be incorporated within the experimental workflow. Next we will highlight some of the problems that can arise during the different stages of an LC-MS experiment, how they negatively influence the experimental results, and how QC methodologies can be used to detect these problems.

2.1 Types of QC samples

QC samples can range from a simple peptide mixture to a single protein digest to a complex whole-cell lysate, and each of these types of samples can be employed in a specific fashion to analyze the system performance.

Common simple peptide mixtures consist of a single protein digest [76] such as bovine serum albumin (BSA), eno- lase, or cytochrome c; or a protein mixture containing a digest of a few proteins [16, 75]; for clarity we will further denote this type of QC samples as 'QC1'. Notably BSA is often used as such a sample because of its historical application in a variety of experimental procedures and its low cost. Furthermore, BSA is usually quite dissimilar from the protein content of the biological samples under consideration, which helps to minimize negative influence on the experimental results due to potential cross-contamination. QC1 samples are typically run on a very frequent basis, i.e. multiple times per day, to quickly evaluate the instrument performance, and they are especially of use to efficiently and systematically assess the LC performance based on observed peak widths and retention times (RTs). As running QC samples takes up valuable instrument time, there is a trade-off between time spent running them and time (and precious sample content) lost due to performing sample runs while the instrument was in a suboptimal state, leading to inferior results. To minimize this trade-off, the QC1 samples are typically run using a short gradient so they can be performed on a frequent basis without unduly occupying an excessive amount of instrument time.

QC samples with a higher sample complexity, denoted 'QC2', consist of a whole-cell lysate, such as a yeast lysate [12, 109], a HeLa cell lysate [76], or a *Pyrococcus furiosus* lysate [145]. QC2 samples are executed using settings equivalent to those for the biological runs to integrally simulate their performance. As this requires more instrument time than the simple QC1 runs, QC2 samples are carried out on a less frequent basis, typically once every day [113]. As instruments have been getting more powerful the importance of complex QC2 samples has increased. In contrast to the QC1 single protein digests, complex QC2 samples are used to primarily evaluate the MS performance. Given the sensitivity of current (Orbitrap) mass spectrometers, it is important to inject small amounts of QC samples (e.g. nanogram amounts of peptides; approaching the limit of detection) in order to sufficiently stress the machine and detect potential flaws [76, 113]. Using such low quantities has the additional advantage that it also helps in preventing or reducing cross-contamination of the biological samples by the QC samples. An important consideration to take into account when running QC2 samples is that ideally their characteristics should reflect those of the biological samples. For example, if phosphoproteomics experiments are conducted on a regular basis, it is important to not only perform a general quality assessment, but also to specifically evaluate the ability to detect phosphorylated peptides and proteins [76]. An example of a recent large-scale project where specialized complex QC samples were used is the CPTAC System Suitability (CompRef) Study [40], whose objective it was to validate mass spectrometry protocols by the participants. The CompRef samples were compiled for use within the CPTAC cancer studies and consisted of human-in-mouse xenograft tumor tissue to closely resemble the biological samples. These CompRef samples were first used as a preliminary validation of the workflow during the System Suitability Study, and subsequently acted as QC samples during successive CPTAC studies to characterize human colon and rectal cancer [122, 150] and to evaluate the longitudinal stability of quantitative proteomics techniques [132]. Another example of a complex sample used in a recent high-profile, multi-site study is the hybrid QC sample used by Navarro et al. [102] to benchmark software tools for label-free proteome quantification. This sample consisted of tryptic digests of human, yeast, and *Escherichia coli* proteins mixed in defined proportions to enable the evaluation of both precision and accuracy of label-free quantification, and it was used to assess and ultimately improve the performance of several software tools [102].

Mixtures of synthetic peptides, denoted 'QC3', are a slightly different type of QC samples. Depending on the complexity of their composition, these mixtures can be run and evaluated individually, similar to the QC1 samples but with an even further simplified sample content, or they can be spiked into QC2 samples or even in real samples. By spiking a well-defined mixture into the biological samples quality control can be performed in parallel with the biological analyses and a direct link between the qualitative information and the experimental data can be established. Similarly, the synthetic peptides can be spiked into one of the above QC samples, typically a complex QC2 whole cell lysate, to combine the advantages of both types of samples

into a single MS run [109]. An important consideration when spiking synthetic peptides into other samples is that these peptides should not overlap with the original sample content. This can be avoided by using artificial, synthetically modified, peptides that are dissimilar from any naturally occurring peptides [43], or by isotopically labeling the synthetic peptides so that their mass is dissimilar from the mass of their naturally occurring peptide variants [17, 110, 111]. These synthetic QC3 peptide mixtures are especially important to evaluate the performance of targeted approaches, such as selected reaction monitoring (SRM). To be able to consistently monitor the transitions of specific peptides and to optimally schedule SRM experiments chromatographic stability is an essential prerequisite, which can be evaluated using these well-characterized peptides as their transitions should exhibit minimal run-to-run variation. Synthetic mixtures can be produced in-house or purchased from commercial vendors, and they are often composed in such a fashion that they can be used to examine specific performance characteristics, as we will discuss further.

2.2 Incorporating QC samples

As illustrated in figure 3, QC samples can be combined with the biological samples in several ways [15]. This is tightly linked to the experimental design [87]: how many controls, replicates, etc., are used cannot be considered independently from the use of QC samples.

Typically, simple QC1 samples are run at the start and end of each batch of experiments, or at least once a day in case of larger batches, as shown in figure 3a. Another approach is to systematically interleave the QC samples after each fixed number of biological samples, as shown in figure 3b. This limits the amount of sample loss that can occur due to an intermediate reduction in instrument performance. For example, Zhang et al. [150] report that during their CPTAC study benchmark tumor xenograft samples were run after every five biological samples, and BSA samples were run after every ten biological samples. Further, as shown in figure 3c, a reference set detailing the expected performance might be required to statistically interpret the subsequent QC runs [16, 113]. This reference set can often be derived from historical high-quality data. In the absence of such measurements it might be necessary to run multiple QC samples successively prior to the start of an experiment. For example, when switching to a new QC standard sample or when employing a novel protocol the new data cannot be compared to the historical measurements and a reference set might need to be compiled explicitly. Likewise, if multiple (dissimilar) LC columns are combined interchangeably with the same MS instrument a separate reference set for each of the columns must be used, as performance characteristics are column- as well as instrument-dependent. Ideally, a standardized reference set would be provided by the proteomics community through analysis of the same QC sample(s) under comparable conditions. However, no consensus on QC samples currently exists and public data sources to store and share QC data are lacking. Although such information is available for a few multi-site studies [1, 2, 32, 131], as there exist numerous mass spectrometry ap-

plications no single QC strategy or reference set will likely be fully appropriate in all scenarios.

2.3 Quality control throughout the experimental workflow

As mentioned previously, a typical LC-MS experiment consists of several different stages. Broadly, this process can be divided in the four following phases [76]: (i) sample preparation, including proteolytic digestion of the proteins; (ii) peptide separation through liquid chromatography; (iii) mass spectrometry analysis; and (iv) computational data interpretation. All these steps can introduce significant variability that needs to be controlled in order to obtain reproducible results, so the ideal QC methodology should be able to assess the performance of each of these stages.

For each of these phases we will highlight potential sources of variability, and we will detail how structured quality control methodologies can be implemented to detect and control this variability.

2.3.1 Sample preparation

Sample preparation enables the analysis of complex biological samples by mass spectrometry techniques, and entails steps from the initial sample collection up to the proteolytic digestion and sample storage prior to the actual LC-MS analysis. As the results of an experiment depend on the initial sample quality this step is of vital importance to acquire trustworthy results [26].

Due to the wide variety in sample origin and experimental applications, each with their specific peculiarities and points of attention, it is impractical to cover all existing sample preparation techniques. However, appropriate sample preparation steps for a bottom-up LC-MS experiment typically include denaturation with a chaotrope, reductant, and/or alkylating agent followed by tryptic digestion of the proteins before the resulting peptides can be further processed [38, 112]. All these different steps will introduce a certain degree of variability in the output results, which needs to be monitored and controlled. As there are often multiple ways to conduct a single step to optimize the sample preparation a careful performance evaluation of the alternative options is required. To this end isobaric labeling techniques, such as isobaric tags for relative and absolute quantitation (iTRAQ), can be employed [53]. Using iTRAQ each experimental condition under investigation can be differentially labeled, multiplexed, and analyzed simultaneously, after which the relative yield of each condition can be determined based on the associated reporter ions. This strategy minimizes inter-condition variation introduced during the LC-MS process and enables the usage of all data to compare different conditions, as opposed to the derived spectrum identification and quantitation information, which can be influenced by unexpected modifications and fragmentation.

Unexpected modifications During denaturation, the secondary and tertiary structure of the proteins are removed by interrupting their non-covalent bonds. Additionally, covalent disulfide bridges are cleaved via reduction, after

A

B

C
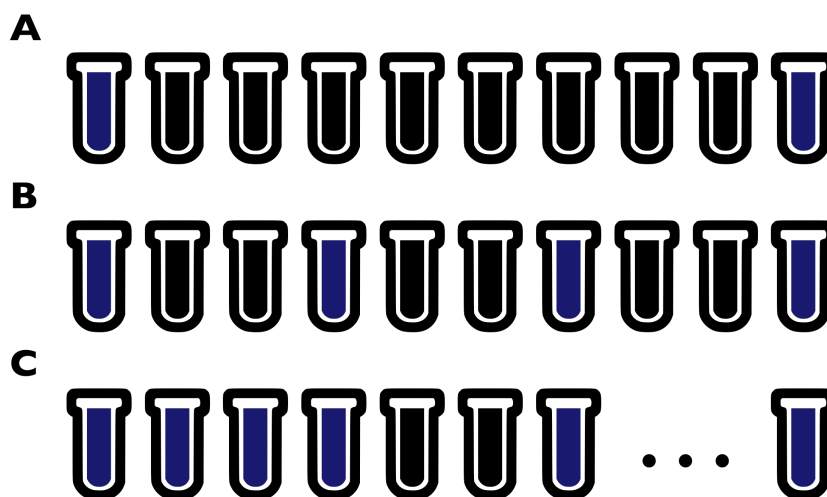
Figure 3: The experimental workflow can incorporate the QC samples (blue) through various combinations with the biological samples (black) [19]. **(a)** QC samples are run at the start and at the end of a batch to assess the batch quality. **(b)** QC samples can be interleaved with the biological samples within a single batch to detect an intermediate decrease in performance and avoid undue sample loss. **(c)** A reference set of high-quality QC measurements is used as the basis to characterize the performance.

which the proteins are alkylated to prevent the reformation of these disulfide bonds. A complete unfolding of the proteins is required to be able to achieve a full enzymatic cleavage into peptides, but these steps can also introduce unexpected post-translational modifications (PTMs) [69].

The chaotropic agent urea is often used for protein denaturation. An important consideration is that urea can cause artificial carbamylation [126]. In aqueous solutions urea dissociates upon heating and over time. One of its degradation products is isocyanate, which covalently reacts with protein N-termini and ε-amine groups of lysines (and arginines to an extremely limited extent) to form carbamyl derivates [77]. Prolonged incubation of protein samples in urea buffers can induce undesired carbamylation, which will occur at a higher rate if old urea or elevated temperatures are employed. Artificially introduced carbamylation is obviously detrimental for studies that investigate the effect of in vivo carbamylation, which has been related to protein ageing. However, general issues are that carbamylation hampers proteolytic digestion with trypsin, blocks amino groups from isotopic/isobaric labeling, and changes peptide charge states, retention times, and masses [129]. Therefore it is important to avoid the formation of urea-induced carbamylation during sample preparation. This can be done by minimizing the generation of cyanates or by removing active cyanates from the solution. Since urea only degrades in aqueous solution it should be prepared freshly [77]. Other strategies involve maintaining the sample at a low temperature [58, 91], lowering the pH [125], or using a variety of buffers [77, 84, 129]. To verify that unexpected carbamylation is not present in an excessive amount appropriate search settings during peptide identification should be used, i.e. a variable carbamylation PTM should be considered.

Another source of unexpected modifications comes from the alkylation step. Alkylation ensures that after disulfide bridges have been cleaved using a reductant the proteins remain unfolded by preventing reformation of the disul-

fide linkages. For this step, a commonly used alkylation agent is iodoacetamide (IAM). Through reaction with IAM a carbamidomethyl group is attached to cysteine residues to prevent these from reforming disulfide bridges, which results in a monoisotopic mass difference of 57.021 464 Da. A potential issue is that overalkylation with IAM will cause N-terminal carbamidomethyl modifications as well [27]. Alternatively, alkylation can be done through carboxymethylation with iodoacetic acid (IAA), which adds a monoisotopic mass of 58.005 479 Da. Similar to IAM, overalkylation with IAA will result in N-terminal modifications [146]. Therefore care has to be taken that the sample solution should not be overexposed to either IAM or IAA during alkylation, and appropriate search settings specifying the corresponding N-terminal modifications should be employed to verify this.

A prevalent modification that can easily be misinterpreted is the nonenzymatic deamidation of asparagines and glutamines to aspartates and glutamates respectively, whose rate can increase dramatically during prolonged incubations in digestion buffers at a mildly alkaline pH [59, 79]. As deamidation adds a monoisotopic mass of 0.984 016 Da, when not correctly considering this modification the ^{13}C peaks of non-deamidated peptides can be misassigned as monoisotopic peaks of the corresponding deamidated ones, although current high-resolution instruments are able to unambiguously distinguish these peaks. As before it is important to carefully perform the sample preparation and use the correct identification search settings to verify that unexpected modifications have not been introduced.

Suitable search settings are essential to correctly interpret the generated data. Importantly, any expected modifications as well as modifications that can be involuntarily introduced, as discussed previously, should be specified correctly. A recent analysis indicates that unexpected or unconsidered modifications account for missing identifications of a large proportion of unassigned spectra [33]. Similarly, an analysis of 19 million spectral clusters based
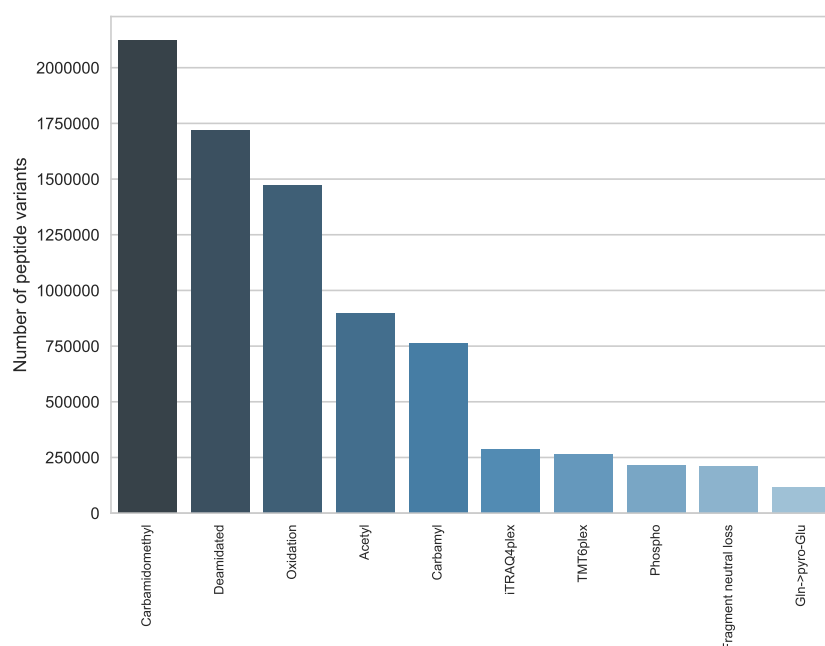
Figure 4: The most frequently observed modifications in the MassIVE database (June 26, 2017). Modification counts are based on the number of distinct peptide variants on which each PTM was detected, out of 10 773 204 peptide variants in total [22].

on previously unidentified spectra deposited in the PRIDE database [139] illustrates the extent to which unidentified spectra can be traced back to unexpected or unconsidered PTMs [57]. As an example of which PTMs commonly occur figure 4 shows the most prevalent modifications detected in the MassIVE database.

Proteolytic digestion stability   In a bottom-up LC-MS proteomics experiment proteins are not analyzed directly, instead they are cleaved into peptides through proteolytic digestion. For this task trypsin is currently the most frequently used protease [137]. Advantages of trypsin are its low cost and its high cleavage specificity and activity. Furthermore, tryptic peptides have various desirable characteristics: their mass is within the preferred mass range for mass spectrometry analysis (based on an *in silico* digestion of all proteins in UniProtKB/Swiss-Prot [134] unique fully tryptic peptides have a median length of 12 amino acids and an interquartile range between 8 and 20 amino acids) and they are ideally suited to carry at least two defined positive charges [127].

At its most basic level, trypsin cleaves exclusively and systematically C-terminal of arginine and lysine, unless followed by a proline [70]. Nevertheless, the formation of semitryptic and nonspecific peptides during protein digestion can still happen due to multiple reasons, although these peptides show a decreased repeatability [131] and they are often not considered during the subsequent bioinformatics analysis, resulting in missing or incorrect identifications [71, 86]. Moreover, most notably for targeted and other quantification experiments consistency of the detectable peptides is of crucial importance.

There are many factors that can influence the digestion stability. One of these is the manner in which the preceding

sample preparation steps were performed, and Proc et al. [116] have shown that the choice of chaotropic agents, surfactants, and solvents significantly influences the digestion reproducibility. Other factors that have an influence are the temperature and the pH at which the digestion is carried out, the enzyme-to-substrate ratio, and the duration of the digestion. At a higher temperature the thermal denaturation of trypsin results in a loss of tryptic activity and autolysis [48, 85], while a lower pH improves trypsin stability over an extended digestion period [85]. Meanwhile, although enzyme-to-substrate ratios reported in the literature range from 1 : 100 to as high as 1 : 2.5, Loziuk et al. [85] have shown that at excessive enzyme-to-substrate ratios an "overdigestion" of peptides caused by increased tryptic autolysis occurs, which may lead to the generation of nonspecific and very small peptides. Similarly, Hildonen, Halvorsen, and Reubsaet [60] recommend a limited digestion time, to avoid a complete digestion as this leads to an increased number of small peptides that are not LC-MS detectable. Furthermore, not all trypsin is created equally, with the origin of the trypsin an important source of variability. Comparisons have shown that the number of missed cleavages, semitryptic peptides, and nontryptic peptides can vary significantly based on whether the trypsin is of bovine or porcine origin [140] and between different commercial trypsins [29, 31].

To assess the digestion performance it is important to monitor the extent of missed cleavages, semitryptic peptides, and nontryptic peptides. Ideally fully tryptic peptides should be preferred as their formation is more reproducible when the trypsin digestion is able to proceed to a state of equilibrium [131]. In some cases semitryptic peptides might be desired as well to generate more detectable peptides and increase the protein sequence coverage [60]. Furthermore it is important to take into account that digestion efficiency

is protein- and sample-dependent [46]. Therefore, there is no one-size-fits-all optimal digestion procedure; specialized protocols might be required to for example optimally monitor specific transitions in a targeted experiment.

DIGESTIF is a commercially available compound QC sample that can be used to evaluate the tryptic digestion efficiency [81]. The *DIGESTIF* standard is assembled from a protein scaffold and artificial peptides, with the amino acids flanking the cleavage sites of these peptides selected to either favor or hinder proteolytic cleavage. This allows to progressively monitor the digestion performance by checking which peptides are effectively generated compared to their theoretical cleavage specificity. Alternatively, to monitor the digestion performance Domon et al. [28, 49] inserted QC samples at various moments during the experimental process. Prior to any sample preparation steps they start with a well-defined QC mixture of a few proteins, insert a first set of isotopically labeled peptides representing a subset of tryptic peptides of these proteins prior to digestion, and a second set of isotopically labeled peptides (with the same amino acid sequences but a different isotope pattern) prior to the LC-MS analysis. By comparing the relative intensities of the unlabeled peptides, originating from the initial proteins, and the labeled peptides from the first set of labeled peptides the digestion efficiency can be assessed. Further, through comparison with the intensities of the labeled peptides from the second set the overall recovery of the full sample preparation workflow can be evaluated.

Although trypsin is by far the most popular protease, employing another protease or performing a multi-protease protein digestion can have specific advantages [51, 135]. A common alternative to trypsin is the combination of Lys-C and trypsin, which generates similar peptides and significantly reduces the number of missed cleavages [52]. Less frequently used proteases can be beneficial as well, for example to generate longer peptides for "middle-down" proteomics, although these proteases are usually not as thoroughly characterized as trypsin is, so care has to be taken [135]. In these situations a consistent and systematic QC methodology assumes even greater importance.

Sample loss Differential recovery of peptides due to non-specific adsorption is a potential source of unexpected sample loss during sample preparation, leading to a reduced reproducibility [88]. This sample loss can occur in all steps of a proteomics workflow, and care should be taken that suitable sample handling material is employed at all times. It should be taken into account that adsorption is peptide-specific [54], so optimized protocols might be required for specific situations. Furthermore, a systematic analysis of well-characterized QC samples can highlight signal loss.

The type of sample tubes that are used for peptide storage can result in a large variation in the results, with low-adsorption plastic tubes more suitable than regular plastic tubes or glass tubes [10, 78]. In contrast, hydrophobic peptides exhibit an increase in recoverability for glass tubes [136]. Furthermore, the addition of other compounds to the sample solution can be used to reduce sample loss due to competition of adsorption with the peptides [54, 128].

Adsorption does not happen exclusively to sample tubes; for example some peptides, including all sulfur-containing peptides, adsorb on the stainless steel injection needle as well [136]. As a rule, the more sample handling steps are undertaken, the more loss due to surface adsorption occurs [89]. Therefore, online and automated methods can help to reduce potential sample loss.

Contaminants Another important source of variability is the presence of contaminants in the sample [61, 72, 142]. Contaminants will compete with the spectra of interest during MS measurements and can cause ion suppression of low abundant peptides. Contaminants can often have seemingly innocuous origins, such as a lab member using a new perfume or wearing a wool sweater [72]. It is important to be aware of potential sources of contaminants during all sample preparation steps to avoid undue contamination.

Some contaminants can be traced back to a prior sample preparation step. For example, trypsin autolysis artifacts can be generated during protein digestion, or polymeric interferences can leak from plastics employed in the laboratory. Other contaminants can be involuntarily introduced into the sample. One of the most prevalent contaminants is keratin, which is omnipresent and can originate from skin, hair, dust, etc.

Total elimination of all contaminants is virtually impossible, but suitable procedures can help to minimize contamination. To prevent contaminants as much as possible it is important to always work in a clean lab environment, wear suitable lab clothes, and use specialized equipment for a single task exclusively. To be able to detect contaminants it is necessary to specify them in the identification search settings. A recent analysis of public data deposited in the PRIDE repository [139] indicates that a majority of commonly incorrectly identified spectra corresponds to contaminants such as albumin, trypsin, and keratin [57]. The MaxQuant software [34] has functionality to automatically include a built-in database of contaminant sequences during sequence database searching [35]. Otherwise, lists of commonly observed contaminants are publicly available. The common Repository of Adventitious Proteins (cRAP) (http://www.thegpm.org/crap/) provides a resource of contaminant proteins, sourced from the Global Proteome Machine (GPM) [37]. Both a fasta file for use in sequence database searching and a spectral library in the X! Hunter format [36] are available. In addition, the PRIDE database [139] provides a spectral library of contaminants, generated through PRIDE Cluster [57].

2.3.2 Liquid chromatography

Prior to MS analysis peptides are typically processed using liquid chromatography to separate them based on their hydrophobicity. This adds a time dimension to the subsequently recorded MS data, which enables the mass measurement of individual peptides by spreading out the dense information content of a complex sample over the range of the LC gradient, and which provides orthogonal information for the peptide identification [108].

The LC phase is typically subject to more variable influences than any other component of the LC-MS system [11], and consequently it is the most common culprit of variability in the results of an experiment [118]. A rigorous

monitoring of the chromatographic performance is therefore essential. Useful QC metrics include the peak shape (width and height), as sharper peaks generate higher signal intensities and can reduce oversampling [118]. A disproportionate level of signal intensity early or late in the gradient can indicate that the column should be serviced or replaced. An early signal can be caused by sample bleed, and a late signal can arise from peak tailing of either overloaded peptides or peptides with poor chromatographic behavior [118]. The RT of known peptides and their elution order can be used to measure differences between early (hydrophilic) and late (hydrophobic) peptides in the chromatographic gradient [1, 118]. Leakages are a problem that can be very hard to diagnose as tiny droplets can evaporate before they are spotted and localized [105]. If a leakage occurs before the column it can be detected by examining the column pressure profile. However, if a leakage occurs between the column and the ion source the column pressure will remain similar to the standard pressure while there might still be a sensitivity loss and/or unstable operation of the ion spray as a smaller amount of liquid will exit from the column [105].

It is of vital importance to avoid cross-contamination due to sample carryover. Carryover happens when an analyte originating from a previously analyzed sample reappears during a subsequent injection, which will result in interference with the active measurements. Carryover can occur because of interactions between the sample and various materials it comes into contact with, as mentioned previously [64]. Other potential sources of carryover are mixing chambers, which are empty spaces within the column where consecutive segments of the sample mix together instead of being separated, and dead volumes, which are spaces in the LC system that are not swept by the mobile phase [105]. Even the smallest irregularities in tubing connections can lead to the generation of dead volumes where sample residues can be trapped [98]. The presence of dead volumes can be diagnosed using the UV detector, which can be a useful tool for system diagnostics and troubleshooting [99, 105]. Furthermore, both dead volumes and mixing chambers hamper the performance of chromatographic systems, as besides sample carryover, they can also give rise to peak broadening and prolonged elution times. To minimize or avoid carryover suitable column washing steps should be employed [39, 100, 144]. The level of carryover and the cleaning effectiveness can be tested by using blank injections between runs of different samples. Quality recommendations typically mandate that the response of any background component in a blank sample should be less than 20 % of the lower limit of quantification (LLOQ) [66].

Specialized QC samples can help to thoroughly monitor the performance of the LC system. By composing QC samples so that they contain peptides with varying hydrophobicities the elution profile of the LC gradient can be characterized and evaluated [30], as illustrated in figure 5. Notable are so-called indexed retention time (iRT) peptide standards. These peptides have standardized RTs spanning a wide gradient and can be used to normalize the RT of individual experiments [43]. Although RTs can be predicted through computational modeling [101], these predictions have a somewhat limited accuracy [117]. Instead, the reference RTs of the iRT peptides can be used to correct

for variations in the RT of the other peptides detected in a single experiment or to align RTs across multiple experiments. Several QC standards containing iRT peptides have been proposed [17, 30, 44, 62, 81]. These standards mostly vary slightly in the range of the LC gradient they can cover, but some standards have further advanced properties. For example, the previously mentioned *DIGESTIF* standard can additionally be used to evaluate the tryptic digestion performance [81], while the *RePLiCal* standard consists of a synthetic protein that exclusively contains lysine-terminating peptides, which ensures that proteolytic digestion by both trypsin and Lys-C can be evaluated analogously [62].

2.3.3 Mass spectrometry

As peptides elute from the LC column their mass over charge is measured in the mass spectrometer.

Prior to the mass measurement peptides are ionized through ESI. The spray stability can be checked by monitoring for drops in the ion current, which can indicate spray sputter [118, 119]. Tryptic digests are expected to generate mainly peptides containing a 2+ charge, and a high rate of differently charged peptides can indicate ionization issues and will likely impact identification rates [118]. Besides due to an unstable proteolytic digestion, as mentioned previously, partially tryptic peptides can also originate from in-source peptide fragmentation [73]. It is possible to differentiate partially tryptic peptides originating from in-source fragmentation from other proteolytic-derived partially tryptic peptides as the former will have the same LC elution time as their parent peptides [73]. To measure high-quality spectra sufficient signal should be present. Various parameters can influence the internal instrument behavior and these should be carefully optimized [6, 67, 68, 148, 151]. For example, for trapping instruments the maximum ion injection time and the automatic gain control (AGC) are interrelated instrument parameters that influence the signal-to-noise ratio, and the effective ion injection time can be monitored to detect problems with sample load [67]. By comparing the measured masses of known compounds, which can either be explicitly added reference standards or systematically observed contaminants, the mass accuracy can be evaluated [118]. If the option is available, these known masses can further be used as lock mass during mass calibration if excessive mass deviations are observed [124]. The dynamic range can be monitored if peptides are present in varying concentrations, as illustrated in figure 5. QC samples can contain distinct peptides in different concentrations [30] or isotopically labeled variants of the same peptide at different ratios [17]. While the concentrations typically span two to four orders of magnitude, the ability to detect even the smallest concentrations indicates the capacity to detect low-abundant peptides over the observed extensive proteome dynamic range [152]. Furthermore, the sensitivity of the MS instrument can be evaluated by employing only small amounts of QC samples, as mentioned previously.

2.3.4 Bioinformatics data interpretation

Although the wet laboratory workflow is often considered to contribute the most variability to the results of an MS
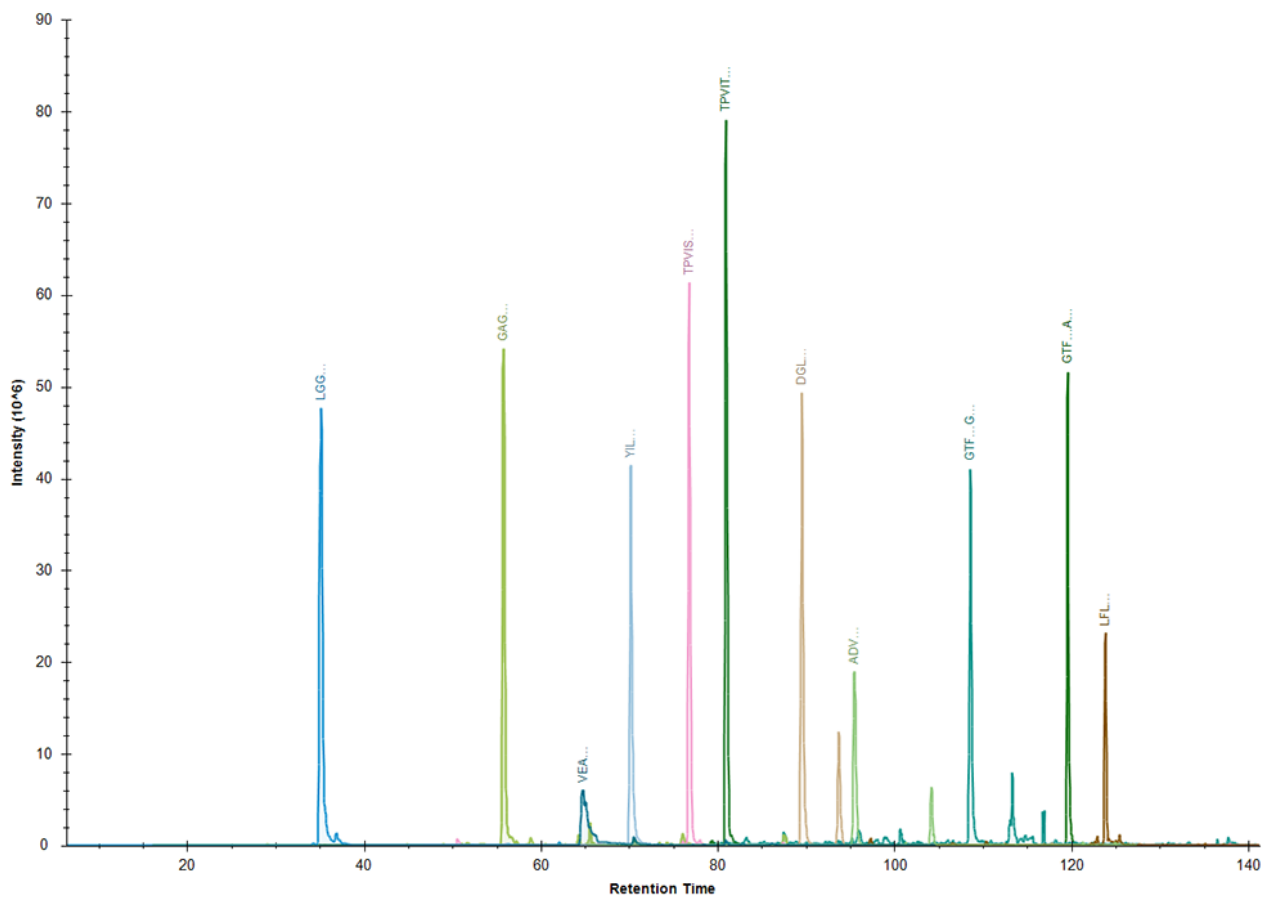
Figure 5: Depending on the composition of the QC samples the LC performance can be monitored using peptides that elute over the entire gradient, and the dynamic range can be monitored if peptides are present in varying concentrations. Depicted here is the Biognosys iRT standard which consists of eleven peptides with varying chromatographic retention [18, 43].

experiment and multiple studies have aimed to improve and standardize existing protocols, the bioinformatics data interpretation can likewise introduce major errors that are a cause of irreproducibility [13]. Already for the most fundamental task, mapping peptide sequences to spectra, there exist dozens of different search engines, each using a unique methodology, (possibly silent) assumptions, and peculiarities. Furthermore, even when using the same tool often different versions or parameter combinations can yield significantly dissimilar results. Although a careful evaluation can indicate the optimal search settings for a single tool [47, 138], the high volume of the data generated by MS techniques and the complexity of the bioinformatics tools is a barrier for a mutual objective assessment [50, 149]. The 'ground truth' for evaluation is typically not known and the introductions of novel tools regularly lack a sufficient comparison to the state-of-the-art methodology. Nevertheless, to inspire confidence in the acquired results a robust computational and statistical interpretation according to community best practices should always be performed before reporting novel biological findings [120].

In the previous sections we have already mentioned several evaluation criteria that should be investigated to detect specific problems. A benchmark of overall performance that is often monitored is the identification rate in terms

of peptide-spectrum matches (PSMs), identified peptides, and identified proteins. This gives a quick insight into the performance of the whole experimental set-up and can indicate whether more detailed quality assessments are required. Whereas for complex QC2 samples, such as a whole-cell lysate, the number of proteins is an often reported metric, for simple QC1 samples, consisting of only a single to a few proteins, the sequence coverage is usually more relevant. The appeal of these high-level QC metrics is that they give a quality assessment of the whole system in a single, easily interpretable metric. However, an MS experiment consists of multiple complex steps that are interrelated, and it might not be possible to identify the source of a decrease in performance based on only a single metric. Instead, sets of detailed QC metrics can be computed [118], highlighting individual performance aspects of the chromatography, the charge state distribution, the spectrum acquisition, etc. A disadvantage of these advanced QC metrics is that, unlike for the number of identifications or the sequence coverage where a higher value is usually better, their interpretation is often not straightforward and requires expert knowledge. Therefore, to establish value intervals of acceptable performance a high-quality reference set might be used, as described in section 2.3 [16, 113]. Furthermore, analyzing multiple metrics simultaneously requires a multivariate

approach. Although this increases the complexity of the data analysis, recent research has shown some promising approaches for informed and automated decision-making based on multivariate sets of advanced QC metrics [5, 23, 141]. Additionally, it does not suffice to investigate QC metrics for a single experiment in isolation. Instead, the longitudinal performance should be examined [14]. Through extensive monitoring of operation over time the technological passport of a mass spectrometer can be established, and based on these highly detailed and instrument-specific insights the reliability of the experimental results can be diagnosed. Although not necessarily related to their biological relevancy, this constitutes the bare essentials required to inspire solid confidence in novel scientific findings. Finally, currently QC metrics are mostly analyzed post hoc after the data acquisition has been completed. Instead, descriptive metrics should be monitored in a real-time fashion over the course of an experiment and spectrum acquisition should be halted automatically when an intermediate decrease in quality is observed to avoid the loss of precious sample content [119]. This requires advanced functionality to allow QC tools to directly interface with the MS instruments, which currently is largely missing.

3 Conclusion

Performing an LC-MS experiment is a highly complex activity and there exist a multitude of potential sources of variability that can influence the results and impact repeatability and reproducibility. We have tried to give an overview of some prevalent issues that can arise and how to detect them, but nevertheless we have only managed to cover the tip of the iceberg. Instituting a thorough QC methodology might initially seem like it requires a lot of effort and it occupies valuable instrument time without any immediate gains, but a systematic quality assessment pays off in the long run and is an indispensable prerequisite to inspire confidence in the acquired results. Especially in order to advance mass spectrometry techniques and use them as routine applications in a clinical setting a consistent analytical performance is a fundamental requirement [94, 123, 147].

Developments on both the experimental and computational front are needed to improve current QC methodologies, for which core facilities can act as an important driver [83]. As proteomics technologies have matured, core facilities have concentrated the cutting-edge technical expertise necessary to obtain high-quality results, and they form an essential means of providing this in an affordable manner [96]. Maybe even more than research laboratories, core facilities have an incentive to support and develop robust quality assurance practices to demonstrate the quality of the generated data to their clients and stakeholders, and through their expert knowledge on a broad aspect of MS-based applications they are at the forefront of developing standardized QC workflows. Significant bioinformatics work is needed as well. All too often laboratories still only monitor detailed QC metrics in an empirical fashion when a malfunctioning is suspected, instead of on a systematic basis. This can be partly attributed to the relative absence of user-friendly tools and software suites that facilitate and encourage a methodical QC workflow. Although a few tools to compute advanced QC metrics exist [15, 24], they remain underused in part due to their limited ease-of-use. Nevertheless, to make further progress objective metrics rooted in a solid bioinformatics foundation are mandatory. The end goal should not be to merely understand QC issues retrospectively, but also to prevent them from happening by timely suggesting solutions. Eventually the QC tools should ideally be tightly coupled to the MS instrumentation to make automated decisions on the fly, avoiding subjective and time-consuming manual quality assessments to increase the throughput.

Finally, because of these obvious advantages we expect that the importance of quality control will only increase in the (near) future. Currently QC information is often not included in scientific publications, which might lead to uncertainty on the conducted methodology. Instead, in the future reporting this information might become formalized, similar to existing guidelines mandated by journals [133], and the QC metrics might become an integral part of a data submission to public data repositories [41, 95], with current work ongoing to provide the necessary technical basis for this goal [25]. Coupling comprehensive QC information to the experimental data will enable assessing the reliability of an experiment at a glance. Especially in light of some historical occasions where claims turned out to be exaggerated [7, 45] and recent reports of the general reproducibility crisis in various scientific fields [8], an innate approach to quality control is mandatory to inspire confidence in and to advance the field of mass spectrometry-based proteomics.

Acknowledgments

References

[1] Abbatiello, S. E., Mani, D. R., Schilling, B., MacLean, B., et al. "Design, Implementation and Multisite Evaluation of a System Suitability Protocol for the Quantitative Assessment of Instrument Performance in Liquid Chromatography-Multiple Reaction Monitoring-MS (LC-MRM-MS)." In: *Mol. Cell. Proteomics* 12.9 (Sept. 1, 2013), pp. 2623–2639. DOI: 10.1074/mcp.M112.027078.

[2] Addona, T. A., Abbatiello, S. E., Schilling, B., Skates, S. J., et al. "Multi-Site Assessment of the Precision and Reproducibility of Multiple Reaction Monitoring–based Measurements of Proteins in Plasma." In: *Nat. Biotechnol.* 27.7 (July 2009), pp. 633–641. DOI: 10.1038/nbt.1546.

[3] Aebersold, R. and Mann, M. "Mass-Spectrometric Exploration of Proteome Structure and Function." In: *Nature* 537.7620 (Sept. 14, 2016), pp. 347–355. DOI: 10.1038/nature19949.

[4] Albar, J.-P. and Canals, F. "Standardization and Quality Control in Proteomics." In: *J. Proteomics* 95 (Dec. 16, 2013), pp. 1–2. DOI: 10.1016/j.jprot.2013.11.002.

[5] Amidan, B. G., Orton, D. J., LaMarche, B. L., Monroe, M. E., et al. "Signatures for Mass Spectrometry Data Quality." In: *J. Proteome Res.* 13.4 (Apr. 4, 2014), pp. 2215–2222. DOI: 10.1021/pr401143e.

[6] Andrews, G. L., Dean, R. A., Hawkridge, A. M., and Muddiman, D. C. "Improving Proteome Coverage on a LTQ-Orbitrap Using Design of Experiments." In: *J. Am. Soc. Mass Spectrom.* 22.4 (Apr. 2011), pp. 773–783. DOI: 10.1007/s13361-011-0075-2.

[7] Baggerly, K. A., Morris, J. S., Edmonson, S. R., and Coombes, K. R. "Signal in Noise: Evaluating Reported Reproducibility of Serum Proteomic Tests for Ovarian Cancer." In: *JNCI J. Natl. Cancer Inst.* 97.4 (Feb. 16, 2005), pp. 307–309. DOI: 10.1093/jnci/dji008.

[8] Baker, M. "1,500 Scientists Lift the Lid on Reproducibility." In: *Nature* 533.7604 (May 25, 2016), pp. 452–454. DOI: 10.1038/533452a.

[9] Bantscheff, M., Lemeer, S., Savitski, M. M., and Kuster, B. "Quantitative Mass Spectrometry in Proteomics: Critical Review Update from 2007 to the Present." In: *Anal. Bioanal. Chem.* 404.4 (Sept. 2012), pp. 939–965. DOI: 10.1007/s00216-012-6203-4.

[10] Bark, S. J. and Hook, V. "Differential Recovery of Peptides from Sample Tubes and the Reproducibility of Quantitative Proteomic Data." In: *J. Proteome Res.* 6.11 (Nov. 1, 2007), pp. 4511–4516. DOI: 10.1021/pr070294o.

[11] Barwick, V. J. "Sources of Uncertainty in Gas Chromatography and High-Performance Liquid Chromatography." In: *J. Chromatogr. A* 849.1 (July 16, 1999), pp. 13–33.

[12] Beasley-Green, A., Bunk, D., Rudnick, P., Kilpatrick, L., et al. "A Proteomics Performance Standard to Support Measurement Quality in Proteomics." In: *PROTEOMICS* 12.7 (Apr. 2012), pp. 923–931. DOI: 10.1002/pmic.201100522.

[13] Bell, A. W., Deutsch, E. W., Au, C. E., Kearney, R. E., et al. "A HUPO Test Sample Study Reveals Common Problems in Mass Spectrometry–based Proteomics." In: *Nat. Methods* 6.6 (June 29, 2009), pp. 423–430. DOI: 10.1038/nmeth.1333.

[14] Bennett, K. L., Wang, X., Bystrom, C. E., Chambers, M. C., et al. "The 2012/2013 ABRF Proteomic Research Group Study: Assessing Longitudinal Intralaboratory Variability in Routine Peptide Liquid Chromatography Tandem Mass Spectrometry Analyses." In: *Mol. Cell. Proteomics* 14.12 (Dec. 1, 2015), pp. 3299–3309. DOI: 10.1074/mcp.O115.051888.

[15] Bereman, M. S. "Tools for Monitoring System Suitability in LC MS/MS Centric Proteomic Experiments." In: *PROTEOMICS* 15 (5-6 Mar. 2015), pp. 891–902. DOI: 10.1002/pmic.201400373.

[16] Bereman, M. S., Johnson, R., Bollinger, J., Boss, Y., et al. "Implementation of Statistical Process Control for Proteomic Experiments via LC MS/MS." In: *J. Am. Soc. Mass Spectrom.* 25.4 (Apr. 2014), pp. 581–587. DOI: 10.1007/s13361-013-0824-5.

[17] Beri, J., Rosenblatt, M. M., Strauss, E., Urh, M., et al. "Reagent for Evaluating Liquid Chromatography–tandem Mass Spectrometry (LC-MS/MS) Performance in Bottom-up Proteomic Experiments." In: *Anal. Chem.* 87.23 (Dec. 1, 2015), pp. 11635–11640. DOI: 10.1021/acs.analchem.5b04121.

[18] Bittremieux, W. *Example run of the Biognosys iRT standard for liquid chromatography - mass spectrometry.* June 2017. DOI: 10.5281/zenodo.815429.

[19] Bittremieux, W. *Incorporation of QC samples within the experimental workflow.* June 2017. DOI: 10.5281/zenodo.815427.

[20] Bittremieux, W. *Ishikawa diagram of sources of variability impacting a liquid chromatography - mass spectrometry experiment.* June 2017. DOI: 10.5281/zenodo.815421.

[21] Bittremieux, W. *Liquid chromatography - mass spectrometry workflow.* June 2017. URL: https://doi.org/10.5281/zenodo.815415.

[22] Bittremieux, W. *Most frequently observed modifications in the MassIVE database.* June 2017. DOI: 10.5281/zenodo.819129.

[23] Bittremieux, W., Meysman, P., Martens, L., Valkenborg, D., et al. "Unsupervised Quality Assessment of Mass Spectrometry Proteomics Experiments by Multivariate Quality Control Metrics." In: *J. Proteome Res.* 15.4 (Apr. 1, 2016), pp. 1300–1307. DOI: 10.1021/acs.jproteome.6b00028.

[24] Bittremieux, W., Valkenborg, D., Martens, L., and Laukens, K. "Computational Quality Control Tools for Mass Spectrometry Proteomics." In: *PROTEOMICS* 17 (3-4 Feb. 2017), p. 1600159. DOI: 10.1002/pmic.201600159.

[25] Bittremieux, W., Walzer, M., Tenzer, S., Zhu, W., et al. "The Human Proteome Organization-Proteomics Standards Initiative Quality Control Working Group: Making Quality Control More Accessible for Biological Mass Spectrometry." In: *Anal. Chem.* 89.8 (Apr. 18, 2017), pp. 4474–4479. DOI: 10.1021/acs.analchem.6b04310.

[26] Bodzon-Kulakowska, A., Bierczynska-Krzysik, A., Dylag, T., Drabik, A., et al. "Methods for Samples Preparation in Proteomic Research." In: *J. Chromatogr. B* 849 (1-2 Apr. 15, 2007), pp. 1–31. DOI: 10.1016/j.jchromb.2006.10.040.

[27] Boja, E. S. and Fales, H. M. "Overalkylation of a Protein Digest with Iodoacetamide." In: *Anal. Chem.* 73.15 (Aug. 1, 2001), pp. 3576–3582. DOI: 10.1021/ac0103423.

[28] Bourmaud, A., Gallien, S., and Domon, B. "A Quality Control of Proteomic Experiments Based on Multiple Isotopologous Internal Standards." In: *EuPA Open Proteomics* 8 (Sept. 2015), pp. 16–21. DOI: 10.1016/j.euprot.2015.07.010.

[29] Bunkenborg, J., Espadas, G., and Molina, H. "Cutting Edge Proteomics: Benchmarking of Six Commercial Trypsins." In: *J. Proteome Res.* 12.8 (Aug. 2, 2013), pp. 3631–3641. DOI: 10.1021/pr4001465.

[30] Burkhart, J. M., Premsler, T., and Sickmann, A. "Quality Control of Nano-LC-MS Systems Using Stable Isotope-Coded Peptides." In: *PROTEOMICS* 11.6 (Mar. 2011), pp. 1049–1057. DOI: 10.1002/pmic.201000604.

[31] Burkhart, J. M., Schumbrutzki, C., Wortelkamp, S., Sickmann, A., et al. "Systematic and Quantitative Comparison of Digest Efficiency and Specificity Reveals the Impact of Trypsin Quality on MS-Based Proteomics." In: *J. Proteomics* 75.4 (Feb. 2, 2012), pp. 1454–1462. DOI: 10.1016/j.jprot.2011.11.016.

[32] Campos, A., Díaz, R., Martínez-Bartolomé, S., Sierra, J., et al. "Multicenter Experiment for Quality Control of Peptide-Centric LC−MS/MS Analysis — A Longitudinal Performance Assessment with nLC Coupled to Orbitrap MS Analyzers." In: *J. Proteomics* 127, Part B (Sept. 8, 2015), pp. 264–274. DOI: 10.1016/j.jprot.2015.05.012.

[33] Chick, J. M., Kolippakkam, D., Nusinow, D. P., Zhai, B., et al. "A Mass-Tolerant Database Search Identifies a Large Proportion of Unassigned Spectra in Shotgun Proteomics as Modified Peptides." In: *Nat. Biotechnol.* 33.7 (June 15, 2015), pp. 743–749. DOI: 10.1038/nbt.3267.

[34] Cox, J. and Mann, M. "MaxQuant Enables High Peptide Identification Rates, Individualized p.p.b.-Range Mass Accuracies and Proteome-Wide Protein Quantification." In: *Nat. Biotechnol.* 26.12 (Nov. 30, 2008), pp. 1367–1372. DOI: 10.1038/nbt.1511.

[35] Cox, J., Matic, I., Hilger, M., Nagaraj, N., et al. "A Practical Guide to the MaxQuant Computational Platform for SILAC-Based Quantitative Proteomics." In: *Nat. Protoc.* 4.5 (Apr. 16, 2009), pp. 698–705. DOI: 10.1038/nprot.2009.36.

[36] Craig, R., Cortens, J., Fenyo, D., and Beavis, R. "Using Annotated Peptide Mass Spectrum Libraries for Protein Identification." In: *J. Proteome Res.* 5.8 (Aug. 4, 2006), pp. 1843–1849. DOI: 10.1021/pr0602085.

[37] Craig, R., Cortens, J. P., and Beavis, R. C. "Open Source System for Analyzing, Validating, and Storing Protein Identification Data." In: *J. Proteome Res.* 3.6 (Dec. 1, 2004), pp. 1234–1242. DOI: 10.1021/pr049882h.

[38] Dittrich, J., Becker, S., Hecht, M., and Ceglarek, U. "Sample Preparation Strategies for Targeted Proteomics via Proteotypic Peptides in Human Blood Using Liquid Chromatography Tandem Mass Spectrometry." In: *PROTEOMICS - Clin. Appl.* 9 (1-2 Feb. 2015), pp. 5–16. DOI: 10.1002/prca.201400121.

[39] Dolman, S., Eeltink, S., Vaast, A., and Pelzing, M. "Investigation of Carryover of Peptides in Nano-Liquid Chromatography/Mass Spectrometry Using Packed and Monolithic Capillary Columns." In: *J. Chromatogr. B* 912 (Jan. 1, 2013), pp. 56–63. DOI: 10.1016/j.jchromb.2012.11.016.

[40] Edwards, N. J., Oberti, M., Thangudu, R. R., Cai, S., et al. "The CPTAC Data Portal: A Resource for Cancer Proteomics Research." In: *J. Proteome Res.* 14.6 (June 5, 2015), pp. 2707–2713. DOI: 10.1021/pr501254j.

[41] Eisenacher, M., Schnabel, A., and Stephan, C. "Quality Meets Quantity - Quality Control, Data Standards and Repositories." In: *PROTEOMICS* 11.6 (Mar. 2011), pp. 1031–1036. DOI: 10.1002/pmic.201000441.

[42] Eng, J. K., Searle, B. C., Clauser, K. R., and Tabb, D. L. "A Face in the Crowd: Recognizing Peptides through Database Search." In: *Mol. Cell. Proteomics* 10.11 (Nov. 1, 2011), R111.009522–R111.009522. DOI: 10.1074/mcp.R111.009522.

[43] Escher, C., Reiter, L., MacLean, B., Ossola, R., et al. "Using iRT, a Normalized Retention Time for More Targeted Measurement of Peptides." In: *PROTEOMICS* 12.8 (Apr. 2012), pp. 1111–1121. DOI: 10.1002/pmic.201100463.

[44] Eyers, C. E., Simpson, D. M., Wong, S. C. C., Beynon, R. J., et al. "QCAL—a Novel Standard for Assessing Instrument Conditions for Proteome Analysis." In: *J. Am. Soc. Mass Spectrom.* 19.9 (Sept. 2008), pp. 1275–1280. DOI: 10.1016/j.jasms.2008.05.019.

[45] Ezkurdia, I., Calvo, E., Del Pozo, A., Vázquez, J., et al. "The Potential Clinical Impact of the Release of Two Drafts of the Human Proteome." In: *Expert Rev. Proteomics* 12.6 (Nov. 2, 2015), pp. 579–593. DOI: 10.1586/14789450.2015.1103186.

[46] Fang, P., Liu, M., Xue, Y., Yao, J., et al. "Controlling Nonspecific Trypsin Cleavages in LC-MS/MS-Based Shotgun Proteomics Using Optimized Experimental Conditions." In: *Analyst* 140.22 (2015), pp. 7613–7621. DOI: 10.1039/C5AN01505G.

[47] Fenyö, D., Eriksson, J., and Beavis, R. "Mass Spectrometric Protein Identification Using the Global Proteome Machine." In: *Computational Biology*. Ed. by Fenyö, D. Vol. 673. Methods in Molecular Biology. Totowa, NJ: Humana Press, Aug. 17, 2010, pp. 189–202.

[48] Finehout, E. J., Cantor, J. R., and Lee, K. H. "Kinetic Characterization of Sequencing Grade Modified Trypsin." In: *PROTEOMICS* 5.9 (June 2005), pp. 2319–2321. DOI: 10.1002/pmic.200401268.

[49] Gallien, S., Bourmaud, A., and Domon, B. "A Simple Protocol to Routinely Assess the Uniformity of Proteomics Analyses." In: *J. Proteome Res.* 13.5 (May 2, 2014), pp. 2688–2695. DOI: 10.1021/pr4011712.

[50] Gatto, L., Hansen, K. D., Hoopmann, M. R., Hermjakob, H., et al. "Testing and Validation of Computational Methods for Mass Spectrometry." In: *J. Proteome Res.* 15.3 (Mar. 4, 2016), pp. 809–814. DOI: 10.1021/acs.jproteome.5b00852.

[51] Giansanti, P., Tsiatsiani, L., Low, T. Y., and Heck, A. J. R. "Six Alternative Proteases for Mass Spectrometry–based Proteomics beyond Trypsin." In: *Nat. Protoc.* 11.5 (Apr. 28, 2016), pp. 993–1006. DOI: 10.1038/nprot.2016.057.

[52] Glatter, T., Ludwig, C., Ahrné, E., Aebersold, R., et al. "Large-Scale Quantitative Assessment of Different in-Solution Protein Digestion Protocols Reveals Superior Cleavage Efficiency of Tandem Lys-C/Trypsin Proteolysis over Trypsin Digestion." In: *J. Proteome Res.* 11.11 (Nov. 2, 2012), pp. 5145–5156. DOI: 10.1021/pr300273g.

[53] Glibert, P., Van Steendam, K., Dhaenens, M., and Deforce, D. "iTRAQ as a Method for Optimization: Enhancing Peptide Recovery after Gel Fractionation." In: *PROTEOMICS* 14.6 (Mar. 2014), pp. 680–684. DOI: 10.1002/pmic.201300444.

[54] Goebel-Stengel, M., Stengel, A., Taché, Y., and Reeve, J. R. "The Importance of Using the Optimal Plasticware and Glassware in Studies Involving Peptides." In: *Anal. Biochem.* 414.1 (July 1, 2011), pp. 38–46. DOI: 10.1016/j.ab.2011.02.009.

[55] Gonzalez-Galarza, F. F., Lawless, C., Hubbard, S. J., Fan, J., et al. "A Critical Appraisal of Techniques, Software Packages, and Standards for Quantitative Proteomic Analysis." In: *OMICS J. Integr. Biol.* 16.9 (Sept. 10, 2012), pp. 431–442. DOI: 10.1089/omi.2012.0022.

[56] Griss, J. "Spectral Library Searching in Proteomics." In: *PROTEOMICS* 16.5 (Mar. 2016), pp. 729–740. DOI: 10.1002/pmic.201500296.

[57] Griss, J., Perez-Riverol, Y., Lewis, S., Tabb, D. L., et al. "Recognizing Millions of Consistently Unidentified Spectra across Hundreds of Shotgun Proteomics Datasets." In: *Nat. Methods* 13.8 (June 27, 2016), pp. 651–656. DOI: 10.1038/nmeth.3902.

[58] Hagel, P., Gerding, J. J. T., Fieggen, W., and Bloemendal, H. "Cyanate Formation in Solutions of Urea." In: *Biochim. Biophys. Acta BBA - Protein Struct.* 243.3 (Sept. 28, 1971), pp. 366–373. DOI: 10.1016/0005-2795(71)90003-1.

[59] Hao, P., Ren, Y., Alpert, A. J., and Sze, S. K. "Detection, Evaluation and Minimization of Nonenzymatic Deamidation in Proteomic Sample Preparation." In: *Mol. Cell. Proteomics* 10.10 (Oct. 1, 2011), O111.009381–O111.009381. DOI: 10.1074/mcp.O111.009381.

[60] Hildonen, S., Halvorsen, T. G., and Reubsaet, L. "Why Less Is More When Generating Tryptic Peptides in Bottom-up Proteomics." In: *PROTEOMICS* 14 (17-18 Sept. 2014), pp. 2031–2041. DOI: 10.1002/pmic.201300479.

[61] Hodge, K., Have, S. T., Hutton, L., and Lamond, A. I. "Cleaning up the Masses: Exclusion Lists to Reduce Contamination with HPLC-MS/MS." In: *J. Proteomics* 88 (Aug. 2, 2013), pp. 92–103. DOI: 10.1016/j.jprot.2013.02.023.

[62] Holman, S. W., McLean, L., and Eyers, C. E. "RePLiCal: A QconCAT Protein for Retention Time Standardization in Proteomics Studies." In: *J. Proteome Res.* 15.3 (Mar. 4, 2016), pp. 1090–1102. DOI: 10.1021/acs.jproteome.5b00988.

[63] Huang, T., Wang, J., Yu, W., and He, Z. "Protein Inference: A Review." In: *Brief. Bioinform.* 13.5 (Sept. 1, 2012), pp. 586–614. DOI: 10.1093/bib/bbs004.

[64] Hughes, N. C., Wong, E. Y. K., Fan, J., and Bajaj, N. "Determination of Carryover and Contamination for Mass Spectrometry-Based Chromatographic Assays." In: *AAPS J.* 9.3 (Sept. 2007), E353–E360. DOI: 10.1208/aapsj0903042.

[65] Ivanov, A. R., Colangelo, C. M., Dufresne, C. P., Friedman, D. B., et al. "Interlaboratory Studies and Initiatives Developing Standards for Proteomics." In: *PROTEOMICS* 13.6 (Mar. 2013), pp. 904–909. DOI: 10.1002/pmic.201200532.

[66] Jenkins, R., Duggan, J. X., Aubry, A.-F., Zeng, J., et al. "Recommendations for Validation of LC-MS/MS Bioanalytical Methods for Protein Biotherapeutics." In: *AAPS J.* 17.1 (Jan. 2015), pp. 1–16. DOI: 10.1208/s12248-014-9685-5.

[67] Kalli, A. and Hess, S. "Effect of Mass Spectrometric Parameters on Peptide and Protein Identification Rates for Shotgun Proteomic Experiments on an LTQ-Orbitrap Mass Analyzer." In: *PROTEOMICS* 12.1 (Jan. 2012), pp. 21–31. DOI: 10.1002/pmic.201100464.

[68] Kalli, A., Smith, G. T., Sweredoski, M. J., and Hess, S. "Evaluation and Optimization of Mass Spectrometric Settings during Data-Dependent Acquisition Mode: Focus on LTQ-Orbitrap Mass Analyzers." In: *J. Proteome Res.* 12.7 (July 5, 2013), pp. 3071–3086. DOI: 10.1021/pr3011588.

[69] Karty, J. A., Ireland, M. M. E., Brun, Y. V., and Reilly, J. P. "Artifacts and Unassigned Masses Encountered in Peptide Mass Mapping." In: *J. Chromatogr. B* 782 (1-2 Dec. 25, 2002), pp. 363–383. DOI: 10.1016/S1570-0232(02)00550-0.

[70] Keil, B. *Specificity of Proteolysis.* Berlin, Heidelberg: Springer Berlin Heidelberg, 1992.

[71] Keller, A., Nesvizhskii, A. I., Kolker, E., and Aebersold, R. "Empirical Statistical Model to Estimate the Accuracy of Peptide Identifications Made by MS/MS and Database Search." In: *Anal. Chem.* 74.20 (Oct. 15, 2002), pp. 5383–5392. DOI: 10.1021/ac025747h.

[72] Keller, B. O., Sui, J., Young, A. B., and Whittal, R. M. "Interferences and Contaminants Encountered in Modern Mass Spectrometry." In: *Anal. Chim. Acta* 627.1 (Oct. 3, 2008). Ed. by Lubman, D. M., Limbach, P. A., and Li, L., pp. 71–81. DOI: 10.1016/j.aca.2008.04.043.

[73] Kim, J.-S., Monroe, M. E., Camp, D. G., Smith, R. D., et al. "In-Source Fragmentation and the Sources of Partially Tryptic Peptides in Shotgun Proteomics." In: *J. Proteome Res.* 12.2 (Feb. 1, 2013), pp. 910–916. DOI: 10.1021/pr300955f.

[74] Kim, M.-S., Pinto, S. M., Getnet, D., Nirujogi, R. S., et al. "A Draft Map of the Human Proteome." In: *Nature* 509.7502 (May 28, 2014), pp. 575–581. DOI: 10.1038/nature13302.

[75] Klimek, J., Eddes, J. S., Hohmann, L., Jackson, J., et al. "The Standard Protein Mix Database: A Diverse Data Set to Assist in the Production of Improved Peptide and Protein Identification Software Tools." In: *J. Proteome Res.* 7.1 (Jan. 1, 2008), pp. 96–103. DOI: 10.1021/pr070244j.

[76] Köcher, T., Pichler, P., Swart, R., and Mechtler, K. "Quality Control in LC-MS/MS." In: *PROTEOMICS* 11.6 (Mar. 6, 2011), pp. 1026–1030. DOI: 10.1002/pmic.201000578.

[77] Kollipara, L. and Zahedi, R. P. "Protein Carbamylation: In Vivo Modification or in Vitro Artefact?" In: *PROTEOMICS* 13.6 (Mar. 2013), pp. 941–944. DOI: 10.1002/pmic.201200452.

[78] Kraut, A., Marcellin, M., Adrait, A., Kuhn, L., et al. "Peptide Storage: Are You Getting the Best Return on Your Investment? Defining Optimal Storage Conditions for Proteomics Samples." In: *J. Proteome Res.* 8.7 (July 6, 2009), pp. 3778–3785. DOI: 10.1021/pr900095u.

[79] Krokhin, O. V., Antonovici, M., Ens, W., Wilkins, J. A., et al. "Deamidation of -Asn-Gly- Sequences during Sample Preparation for Proteomics: Consequences for MALDI and HPLC-MALDI Analysis." In: *Anal. Chem.* 78.18 (Sept. 1, 2006), pp. 6645–6650. DOI: 10.1021/ac061017o.

[80] Lane, L., Bairoch, A., Beavis, R. C., Deutsch, E. W., et al. "Metrics for the Human Proteome Project 2013–2014 and Strategies for Finding Missing Proteins." In: *J. Proteome Res.* 13.1 (Jan. 3, 2014), pp. 15–20. DOI: 10.1021/pr401144x.

[81] Lebert, D., Louwagie, M., Goetze, S., Picard, G., et al. "DIGESTIF: A Universal Quality Standard for the Control of Bottom-up Proteomics Experiments." In: *J. Proteome Res.* 14.2 (Feb. 6, 2015), pp. 787–803. DOI: 10.1021/pr500834z.

[82] Legrain, P., Aebersold, R., Archakov, A., Bairoch, A., et al. "The Human Proteome Project: Current State and Future Direction." In: *Mol. Cell. Proteomics* 10.7 (July 1, 2011), p. M111.009993. DOI: 10.1074/mcp.M111.009993.

[83] Lilley, K. S., Deery, M. J., and Gatto, L. "Challenges for Proteomics Core Facilities." In: *PROTEOMICS* 11.6 (Mar. 2011), pp. 1017–1025. DOI: 10.1002/pmic.201000693.

[84] Lin, M.-F., Williams, C., Murray, M. V., Conn, G., et al. "Ion Chromatographic Quantification of Cyanate in Urea Solutions: Estimation of the Efficiency of Cyanate Scavengers for Use in Recombinant Protein Manufacturing." In: *J. Chromatogr. B* 803.2 (Apr. 25, 2004), pp. 353–362. DOI: 10.1016/j.jchromb.2004.01.017.

[85] Loziuk, P. L., Wang, J., Li, Q., Sederoff, R. R., et al. "Understanding the Role of Proteolytic Digestion on Discovery and Targeted Proteomic Measurements Using Liquid Chromatography Tandem Mass Spectrometry and Design of Experiments." In: *J. Proteome Res.* 12.12 (Dec. 6, 2013), pp. 5820–5829. DOI: 10.1021/pr4008442.

[86] Ma, Z.-Q., Dasari, S., Chambers, M. C., Litton, M. D., et al. "IDPicker 2.0: Improved Protein Assembly with High Discrimination Peptide Identification Filtering." In: *J. Proteome Res.* 8.8 (Aug. 7, 2009), pp. 3872–3881. DOI: 10.1021/pr900360j.

[87] Maes, E., Kelchtermans, P., Bittremieux, W., De Grave, K., et al. "Designing Biomedical Proteomics Experiments: State-of-the-Art and Future Perspectives." In: *Expert Rev. Proteomics* 13.5 (Apr. 25, 2016), pp. 495–511. DOI: 10.1586/14789450.2016.1172967.

[88] Maes, K., Smolders, I., Michotte, Y., and Van Eeckhaut, A. "Strategies to Reduce Aspecific Adsorption of Peptides and Proteins in Liquid Chromatography-Mass Spectrometry Based Bioanalyses: An Overview." In: *J. Chromatogr. A* 1358 (Sept. 5, 2014), pp. 1–13. DOI: 10.1016/j.chroma.2014.06.072.

[89] Magdeldin, S., Moresco, J. J., Yamamoto, T., and Yates, J. R. "Off-Line Multidimensional Liquid Chromatography and Auto Sampling Result in Sample Loss in LC/LC–MS/MS." In: *J. Proteome Res.* 13.8 (Aug. 1, 2014), pp. 3826–3836. DOI: 10.1021/pr500530e.

[90] Mann, M. "Comparative Analysis to Guide Quality Improvements in Proteomics." In: *Nat. Methods* 6.10 (Oct. 2009), pp. 717–719. DOI: 10.1038/nmeth1009-717.

[91] Marier, J. R. and Rose, D. "Determination of Cyanate, and a Study of Its Accumulation in Aqueous Solutions of Urea." In: *Anal. Biochem.* 7.3 (Mar. 1964), pp. 304–314. DOI: 10.1016/0003-2697(64)90135-6.

[92] Marko-Varga, G., Omenn, G. S., Paik, Y.-K., and Hancock, W. S. "A First Step toward Completion of a Genome-Wide Characterization of the Human Proteome." In: *J. Proteome Res.* 12.1 (Jan. 4, 2013), pp. 1–5. DOI: 10.1021/pr301183a.

[93] Martens, L. "A Report on the ESF Workshop on Quality Control in Proteomics." In: *Mol. Biosyst.* 6.6 (2010), pp. 935–938. DOI: 10.1039/c003912h.

[94] Martens, L. "Bringing Proteomics into the Clinic: The Need for the Field to Finally Take Itself Seriously." In: *PROTEOMICS - Clin. Appl.* 7 (5-6 June 2013), pp. 388–391. DOI: 10.1002/prca.201300020.

[95] Martens, L. "Public Proteomics Data: How the Field Has Evolved from Sceptical Inquiry to the Promise of in Silico Proteomics." In: *EuPA Open Proteomics* 11 (June 2016), pp. 42–44. DOI: 10.1016/j.euprot.2016.02.005.

[96] Meder, D., Morales, M., Pepperkok, R., Schlapbach, R., et al. "Institutional Core Facilities: Prerequisite for Breakthroughs in the Life Sciences." In: *EMBO Rep.* 17.8 (Aug. 1, 2016), pp. 1088–1093. DOI: 10.15252/embr.201642857.

[97] Medzihradszky, K. F. and Chalkley, R. J. "Lessons in de Novo Peptide Sequencing by Tandem Mass Spectrometry." In: *Mass Spectrom. Rev.* 34.1 (Jan. 2015), pp. 43–63. DOI: 10.1002/mas.21406.

[98] Mitulović, G. and Mechtler, K. "HPLC Techniques for Proteomics Analysis–a Short Overview of Latest Developments." In: *Brief. Funct. Genomic. Proteomic.* 5.4 (May 10, 2006), pp. 249–260. DOI: 10.1093/bfgp/ell034.

[99] Mitulović, G., Smoluch, M., Chervet, J.-P., Steinmacher, I., et al. "An Improved Method for Tracking and Reducing the Void Volume in Nano HPLC–MS with Micro Trapping Columns." In: *Anal. Bioanal. Chem.* 376.7 (Aug. 2003), pp. 946–951. DOI: 10.1007/s00216-003-2047-2.

[100] Mitulović, G., Stingl, C., Steinmacher, I., Hudecz, O., et al. "Preventing Carryover of Peptides and Proteins in Nano LC-MS Separations." In: *Anal. Chem.* 81.14 (July 15, 2009), pp. 5955–5960. DOI: 10.1021/ac900696m.

[101] Moruz, L. and Käll, L. "Peptide Retention Time Prediction." In: *Mass Spectrom. Rev.* (Early view Jan. 22, 2016). DOI: 10.1002/mas.21488.

[102] Navarro, P., Kuharev, J., Gillet, L. C., Bernhardt, O. M., et al. "A Multicenter Study Benchmarks Software Tools for Label-Free Proteome Quantification." In: *Nat. Biotechnol.* 34 (Oct. 3, 2016), pp. 1130–1136. DOI: 10.1038/nbt.3685.

[103] Nesvizhskii, A. I., Vitek, O., and Aebersold, R. "Analysis and Validation of Proteomic Data Generated by Tandem Mass Spectrometry." In: *Nat. Methods* 4.10 (Sept. 27, 2007), pp. 787–797. DOI: 10.1038/nmeth1088.

[104] Noble, W. S. and MacCoss, M. J. "Computational and Statistical Analysis of Protein Mass Spectrometry Data." In: *PLoS Comput. Biol.* 8.1 (Jan. 26, 2012). Ed. by Bourne, P. E., e1002296. DOI: 10.1371/journal.pcbi.1002296.

[105] Noga, M., Sucharski, F., Suder, P., and Silberring, J. "A Practical Guide to Nano-LC Troubleshooting." In: *J. Sep. Sci.* 30.14 (Aug. 7, 2007), pp. 2179–2189. DOI: 10.1002/jssc.200700225.

[106] Omenn, G. S., Lane, L., Lundberg, E. K., Beavis, R. C., et al. "Metrics for the Human Proteome Project 2015: Progress on the Human Proteome and Guidelines for High-Confidence Protein Identification." In: *J. Proteome Res.* 14.9 (Sept. 4, 2015), pp. 3452–3460. DOI: 10.1021/acs.jproteome.5b00499.

[107] Omenn, G. S., Lane, L., Lundberg, E. K., Beavis, R. C., et al. "Metrics for the Human Proteome Project 2016: Progress on Identifying and Characterizing the Human Proteome, Including Post-Translational Modifications." In: *J. Proteome Res.* 15.11 (Nov. 4, 2016), pp. 3951–3960. DOI: 10.1021/acs.jproteome.6b00511.

[108] Pasa-Tolić, L., Masselon, C., Barry, R. C., Shen, Y., et al. "Proteomic Analyses Using an Accurate Mass and Time Tag Strategy." In: *BioTechniques* 37.4 (Oct. 2004), 621–624, 626–633, 636 passim.

[109] Paulovich, A. G., Billheimer, D., Ham, A.-J. L., Vega-Montoto, L., et al. "Interlaboratory Study Characterizing a Yeast Performance Standard for Benchmarking LC-MS Platform Performance." In: *Mol. Cell. Proteomics* 9.2 (Feb. 1, 2010), pp. 242–254. DOI: 10.1074/mcp.M900222-MCP200.

[110] Percy, A. J., Chambers, A. G., Smith, D. S., and Borchers, C. H. "Standardized Protocols for Quality Control of MRM-Based Plasma Proteomic Workflows." In: *J. Proteome Res.* 12.1 (Jan. 4, 2013), pp. 222–233. DOI: 10.1021/pr300893w.

[111] Percy, A. J., Chambers, A. G., Yang, J., Jackson, A. M., et al. "Method and Platform Standardization in MRM-Based Quantitative Plasma Proteomics." In: *J. Proteomics* 95 (Dec. 16, 2013), pp. 66–76. DOI: 10.1016/j.jprot.2013.07.026.

[112] Percy, A. J., Parker, C. E., and Borchers, C. H. "Pre-Analytical and Analytical Variability in Absolute Quantitative MRM-Based Plasma Proteomic Studies." In: *Bioanalysis* 5.22 (Nov. 2013), pp. 2837–2856. DOI: 10.4155/bio.13.245.

[113] Pichler, P., Mazanek, M., Dusberger, F., Weilnböck, L., et al. "SIMPATIQCO: A Server-Based Software Suite Which Facilitates Monitoring the Time Course of LC-MS Performance Metrics on Orbitrap Instruments." In: *J. Proteome Res.* 11.11 (Nov. 2, 2012), pp. 5540–5547. DOI: 10.1021/pr300163u.

[114] Picotti, P. and Aebersold, R. "Selected Reaction Monitoring–based Proteomics: Workflows, Potential, Pitfalls and Future Directions." In: *Nat. Methods* 9.6 (May 30, 2012), pp. 555–566. DOI: 10.1038/nmeth.2015.

[115] Piehowski, P. D., Petyuk, V. A., Orton, D. J., Xie, F., et al. "Sources of Technical Variability in Quantitative LC−MS Proteomics: Human Brain Tissue Sample Analysis." In: *J. Proteome Res.* 12.5 (May 3, 2013), pp. 2128–2137. DOI: 10.1021/pr301146m.

[116] Proc, J. L., Kuzyk, M. A., Hardie, D. B., Yang, J., et al. "A Quantitative Study of the Effects of Chaotropic Agents, Surfactants, and Solvents on the Digestion Efficiency of Human Plasma Proteins by Trypsin." In: *J. Proteome Res.* 9.10 (Oct. 1, 2010), pp. 5422–5437. DOI: 10.1021/pr100656u.

[117] Reimer, J., Spicer, V., and Krokhin, O. V. "Application of Modern Reversed-Phase Peptide Retention Prediction Algorithms to the Houghten and DeGraw Dataset: Peptide Helicity and Its Effect on Prediction Accuracy." In: *J. Chromatogr. A* 1256 (Sept. 21, 2012), pp. 160–168. DOI: 10.1016/j.chroma.2012.07.092.

[118] Rudnick, P. A., Clauser, K. R., Kilpatrick, L. E., Tchekhovskoi, D. V., et al. "Performance Metrics for Liquid Chromatography-Tandem Mass Spectrometry Systems in Proteomics Analyses." In: *Mol. Cell. Proteomics* 9.2 (Feb. 1, 2010), pp. 225–241. DOI: 10.1074/mcp.M900223-MCP200.

[119] Scheltema, R. A. and Mann, M. "SprayQc: A Real-Time LC-MS/MS Quality Monitoring System to Maximize Uptime Using off the Shelf Components." In: *J. Proteome Res.* 11.6 (June 1, 2012), pp. 3458–3466. DOI: 10.1021/pr201219e.

[120] Serang, O. and Käll, L. "Solution to Statistical Challenges in Proteomics Is More Statistics, Not Less." In: *J. Proteome Res.* 14.10 (Oct. 2, 2015), pp. 4099–4103. DOI: 10.1021/acs.jproteome.5b00568.

[121] Shao, W. and Lam, H. "Tandem Mass Spectral Libraries of Peptides and Their Roles in Proteomics Research." In: *Mass Spectrom. Rev.* (Early view July 12, 2016). DOI: 10.1002/mas.21512.

[122] Slebos, R. J. C., Wang, X., Wang, X., Zhang, B., et al. "Proteomic Analysis of Colon and Rectal Carcinoma Using Standard and Customized Databases." In: *Sci. Data* 2 (June 23, 2015), p. 150022. DOI: 10.1038/sdata.2015.22.

[123] Smit, N. P. M., van den Broek, I., Romijn, F. P. H. T. M., Haex, M., et al. "Quality Requirements for Quantitative Clinical Chemistry Proteomics." In: *Transl. Proteomics* 2 (Mar. 2014), pp. 1–13. DOI: 10.1016/j.trprot.2013.10.001.

[124] Staes, A., Vandenbussche, J., Demol, H., Goethals, M., et al. "Asn3, a Reliable, Robust, and Universal Lock Mass for Improved Accuracy in LC–MS and LC–MS/MS." In: *Anal. Chem.* 85.22 (Nov. 19, 2013), pp. 11054–11060. DOI: 10.1021/ac4027093.

[125] Stark, G. R. "Reactions of Cyanate with Functional Groups of Proteins. III. Reactions with Amino and Carboxyl Groups." In: *Biochemistry (Mosc.)* 4.6 (June 1965), pp. 1030–1036. DOI: 10.1021/bi00882a008.

[126] Stark, G. R., Stein, W. H., and Moore, S. "Reactions of the Cyanate Present in Aqueous Urea with Amino Acids and Proteins." In: *J. Biol. Chem.* 235.11 (Nov. 1960), pp. 3177–3181.

[127] Steen, H. and Mann, M. "The Abc's (and Xyz's) of Peptide Sequencing." In: *Nat. Rev. Mol. Cell Biol.* 5.9 (Sept. 2004), pp. 699–711. DOI: 10.1038/nrm1468.

[128] Stejskal, K., Potěšil, D., and Zdráhal, Z. "Suppression of Peptide Sample Losses in Autosampler Vials." In: *J. Proteome Res.* 12.6 (June 7, 2013), pp. 3057–3062. DOI: 10.1021/pr400183v.

[129] Sun, S., Zhou, J.-Y., Yang, W., and Zhang, H. "Inhibition of Protein Carbamylation in Urea Solution Using Ammonium-Containing Buffers." In: *Anal. Biochem.* 446 (Feb. 1, 2014), pp. 76–81. DOI: 10.1016/j.ab.2013.10.024.

[130] Tabb, D. L. "Quality Assessment for Clinical Proteomics." In: *Clin. Biochem.* 46.6 (Apr. 2013), pp. 411–420. DOI: 10.1016/j.clinbiochem.2012.12.003.

[131] Tabb, D. L., Vega-Montoto, L., Rudnick, P. A., Variyath, A. M., et al. "Repeatability and Reproducibility in Proteomic Identifications by Liquid Chromatography-Tandem Mass Spectrometry." In: *J. Proteome Res.* 9.2 (Feb. 5, 2010), pp. 761–776. DOI: 10.1021/pr9006365.

[132] Tabb, D. L., Wang, X., Carr, S. A., Clauser, K. R., et al. "Reproducibility of Differential Proteomic Technologies in CPTAC Fractionated Xenografts." In: *J. Proteome Res.* 15.3 (Mar. 4, 2016), pp. 691–706. DOI: 10.1021/acs.jproteome.5b00859.

[133] Taylor, C. F., Paton, N. W., Lilley, K. S., Binz, P.-A., et al. "The Minimum Information about a Proteomics Experiment (MIAPE)." In: *Nat. Biotechnol.* 25.8 (Aug. 8, 2007), pp. 887–893. DOI: 10.1038/nbt1329.

[134] The UniProt Consortium. "UniProt: A Hub for Protein Information." In: *Nucleic Acids Res.* 43 (D1 Jan. 28, 2015), pp. D204–D212. DOI: 10.1093/nar/gku989.

[135] Tsiatsiani, L. and Heck, A. J. R. "Proteomics beyond Trypsin." In: *FEBS J.* 282.14 (July 2015), pp. 2612–2626. DOI: 10.1111/febs.13287.

[136] Van Midwoud, P. M., Rieux, L., Bischoff, R., Verpoorte, E., et al. "Improvement of Recovery and Repeatability in Liquid Chromatography-Mass Spectrometry Analysis of Peptides." In: *J. Proteome Res.* 6.2 (Feb. 1, 2007), pp. 781–791. DOI: 10.1021/pr0604099.

[137] Vandermarliere, E., Mueller, M., and Martens, L. "Getting Intimate with Trypsin, the Leading Protease in Proteomics." In: *Mass Spectrom. Rev.* 32.6 (Nov. 2013), pp. 453–465. DOI: 10.1002/mas.21376.

[138] Vaudel, M., Burkhart, J. M., Sickmann, A., Martens, L., et al. "Peptide Identification Quality Control." In: *PROTEOMICS* 11.10 (May 2011), pp. 2105–2114. DOI: 10.1002/pmic.201000704.

[139] Vizcaíno, J. A., Csordas, A., del-Toro, N., Dianes, J. A., et al. "2016 Update of the PRIDE Database and Its Related Tools." In: *Nucleic Acids Res.* 44 (D1 Jan. 4, 2016), pp. D447–D456. DOI: 10.1093/nar/gkv1145.

[140] Walmsley, S. J., Rudnick, P. A., Liang, Y., Dong, Q., et al. "Comprehensive Analysis of Protein Digestion Using Six Trypsins Reveals the Origin of Trypsin as a Significant Source of Variability in Proteomics." In: *J. Proteome Res.* 12.12 (Dec. 6, 2013), pp. 5666–5680. DOI: 10.1021/pr400611h.

[141] Wang, X., Chambers, M. C., Vega-Montoto, L. J., Bunk, D. M., et al. "QC Metrics from CPTAC Raw LC-MS/MS Data Interpreted through Multivariate Statistics." In: *Anal. Chem.* 86.5 (Mar. 4, 2014), pp. 2497–2509. DOI: 10.1021/ac4034455.

[142] Weber, R. J. M., Li, E., Bruty, J., He, S., et al. "MaConDa: A Publicly Accessible Mass Spectrometry Contaminants Database." In: *Bioinformatics* 28.21 (Nov. 1, 2012), pp. 2856–2857. DOI: 10.1093/bioinformatics/bts527.

[143] Wilhelm, M., Schlegl, J., Hahne, H., Gholami, A. M., et al. "Mass-Spectrometry-Based Draft of the Human Proteome." In: *Nature* 509.7502 (May 28, 2014), pp. 582–587. DOI: 10.1038/nature13319.

[144] Williams, J. S., Donahue, S. H., Gao, H., and Brummel, C. L. "Universal LC–MS Method for Minimized Carryover in a Discovery Bioanalytical Setting." In: *Bioanalysis* 4.9 (May 2012), pp. 1025–1037. DOI: 10.4155/bio.12.76.

[145] Wong, C. C. L., Cociorva, D., Miller, C. A., Schmidt, A., et al. "Proteomics of Pyrococcus Furiosus (Pfu): Identification of Extracted Proteins by Three Independent Methods." In: *J. Proteome Res.* 12.2 (Feb. 1, 2013), pp. 763–770. DOI: 10.1021/pr300840j.

[146] Woods, A. G., Sokolowska, I., and Darie, C. C. "Identification of Consistent Alkylation of Cysteine-Less Peptides in a Proteomics Experiment." In: *Biochem. Biophys. Res. Commun.* 419.2 (Mar. 9, 2012), pp. 305–308. DOI: 10.1016/j.bbrc.2012.02.016.

[147] Wright, I. and Van Eyk, J. E. "A Roadmap to Successful Clinical Proteomics." In: *Clin. Chem.* (Nov. 2016). DOI: 10.1373/clinchem.2016.254664.

[148] Xu, P., Duong, D. M., and Peng, J. "Systematical Optimization of Reverse-Phase Chromatography for Shotgun Proteomics." In: *J. Proteome Res.* 8.8 (Aug. 7, 2009), pp. 3944–3950. DOI: 10.1021/pr900251d.

[149] Yates, J. R. I., Park, S. K. R., Delahunty, C. M., Xu, T., et al. "Toward Objective Evaluation of Proteomic Algorithms." In: *Nat. Methods* 9.5 (Apr. 27, 2012), pp. 455–456. DOI: 10.1038/nmeth.1983.

[150] Zhang, B., Wang, J., Wang, X., Zhu, J., et al. "Proteogenomic Characterization of Human Colon and Rectal Cancer." In: *Nature* 513.7518 (July 20, 2014), pp. 382–387. DOI: 10.1038/nature13438.

[151] Zhou, M., Gucinski, A. C., and Boyne, M. T. "Performance Metrics for Evaluating System Suitability in Liquid Chromatography-Mass Spectrometry Peptide Mass Mapping of Protein Therapeutics and Monoclonal Antibodies." In: *mAbs* 7.6 (July 28, 2015), pp. 1104–1117. DOI: 10.1080/19420862.2015.1074364.

[152] Zubarev, R. A. "The Challenge of the Proteome Dynamic Range and Its Implications for In-Depth Proteomics." In: *PROTEOMICS* 13.5 (Mar. 2013), pp. 723–726. DOI: 10.1002/pmic.201200451.