Quantitative Mass Spectrometry-based Proteomics

Lennart van der Wal and Jeroen A. A. Demmers

Additional information is available at the end of the chapter

Abstract

Mass spectrometry-based proteomics, the large-scale analysis of proteins by mass spectrometry, has emerged as a powerful technology over the past decade and has become an indispensable tool in many biomedical laboratories. Many strategies for differential proteomics have been developed in recent years, which involve either the incorporation of heavy stable isotopes or are based on label-free comparisons and their statistical assessment, and each of these has specific strengths and limitations. This chapter gives an overview of the current state-of-the-art in quantitative or differential proteomics and will be illustrated by several examples.

Keywords: Mass spectrometry, quantitation, SILAC, heavy isotope labelling, chemical tagging, ^{18}O labelling

1. Introduction

Analysis of the proteome using mass spectrometry has proven to be an indispensable tool in biomedical research over the past 15 years or so. Originally, because of technical limitations, only qualitative measurements were performed for the identification of proteins in a sample. However, the need to put a quantitative label on proteomics analyses became evident rapidly. For this reason, several different technologies were developed for their use, in combination with mass spectrometry, to supply researchers with more quantitative data to investigate, e.g. the dynamics of a particular proteome. In this chapter, a brief overview of quantitative approaches in mass spectrometry-based proteomics will be given. Current protocols for quantitative analysis and software solutions for data analysis will be discussed and examples from the field (including our own laboratory) will be given to illustrate the power of these methods.

2. Mass spectrometry-based proteomics

Before the application of mass spectrometry, protein analysis was mostly based on the purification of single proteins or protein complexes, followed by the performance of experiments on these purified proteins or complexes. Usually, such biochemical experiments are quite laborious and mostly reliant on the extent to which the protein can be purified. Although mass spectrometry as a technique to study small molecules dates back to the beginning of the 20th century, the use of mass spectrometry in peptide and protein analysis is more recent. The development of both matrix-assisted laser desorption/ionization (MALDI) and electrospray ionization (ESI) in the 1980s was key to this development, as these techniques allowed the ionization of biomolecules such as peptides, proteins and nucleotides, which made their detection by mass spectrometry possible. The 2002 Nobel Prize in Chemistry was awarded to John B. Fenn and Koichi Tanaka for the development of these ionisation techniques [1]. With the possibility to analyse biomolecules, in particular peptides and proteins, using mass spectrometry, the key step towards proteomics was made. Equally important was the advent of the genomic age, supplying the databases which are instrumental for the analysis and identification of proteins, as well as the technical advances of both mass spectrometers and the (bio)informatic infrastructures that are essential for large data handling.

Mass spectrometry in itself is merely a qualitative analytical technique. The biochemical and biophysical properties of proteins and peptides are quite variable, which leads to large differences in properties such as 'sprayability' and, thus, in resulting ion intensities between different peptides, even though these may be present in equimolar amounts in the sample. In order for mass spectrometry to be useful not only for the qualitative analyses but also for the quantitative analysis, these caveats and problems need to be addressed and solved. Concerning the different types of mass spectrometers, there are several physical principles to choose from. While it goes beyond the scope of this chapter to discuss all of these in details, it is quite useful to be aware of the different possibilities available as these may influence the performance of the quantitative analysis. The type of mass spectrometers that are most widely used in proteomics are (1) time-of-flight (ToF), (2) quadrupole, (3) (Paul) ion trap, (4) FTICR, or (5) orbitrap [2,3]. In ToF analysis, the velocity of an ion is measured in order to determine the size of the particle. The quadrupole analyses the movement of an ion through an electric field, while the Paul ion trap is a type of quadrupole that uses static direct current and radio frequency oscillating electric fields to trap ions. In an FTICR mass spectrometer, ions are trapped in a strong magnetic field and the periodic movement of the ions is translated back to m/z ratios. In an orbitrap, ions are trapped in an orbital motion around a spindle, while the image current from the trapped ions is detected and converted to a high-resolution mass spectrum using Fourier transformation (for a review on orbitrap mass spectrometry, see [3]). Current mass spectrometers are usually hybrid instruments, which combine two or more of the above mentioned principles for the analysis of peptides and proteins.

3. Overview of quantitation methods

The first (semi-)quantitative approach to proteomics was achieved using '2D difference in gel electrophoresis' (2D-DIGE). With this technique, proteins are separated according to size and charge, and it includes the incorporation of fluorescent labels (CyDye) to allow the comparison of two conditions versus an internal standard [4,5]. After separation of the proteins, the spots are analysed using specialised software, measuring the relative fluorescence intensities. Spots that appear to be differentially regulated can then be excised from the gel and identified using mass spectrometry [6]. The usage of an internal standard, usually a mix of the two measured conditions, supplies this method with quantitative properties. However, the limitations of 2D-DIGE, and of 2D gel electrophoresis for complex samples in particular, have led to decreased usage of the technique. Because of limitations in the number of samples to be compared, in studying membrane-bound proteins, as well as in the relative low proteome coverage, alternative technologies have now superseded the use of 2D-DIGE as a quantitative proteomics method. Techniques currently used for quantitation are summarized in Figure 1 and discussed further below.
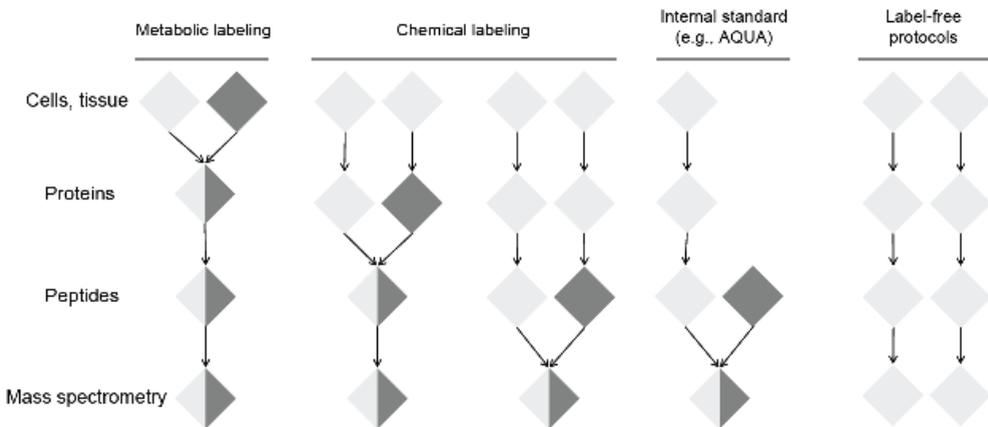

Figure 1. Overview of the stage in which incorporation of the stable isotope labels occurs using different labelling methods in quantitative proteomics. The colour of the diamonds represents the two proteins samples which are differentially labelled and compared in the workflow. (Figure adapted from [7]).

Nowadays, the techniques most frequently used to quantify proteins using mass spectrometry involve labelling proteins with isotopically labelled tags, which can be distinguished in the mass spectrometer because they differ in mass. Differential mass tags result in a (usually only small) mass difference between the 'light' and 'heavy' sample, while proteins and/or peptide properties such as the retention time on a chromatography column are not affected. This allows for the simultaneous analysis of the tagged proteins in a single mass spectrum or LC-MS run. Several methods based on the addition of labelled tags are used in modern proteomics, each with their strong and weak points. Furthermore, with the development of more sensitive and

faster mass spectrometers, methods that allow quantitation of proteins in a label-free manner have been developed, including spectral counting and the comparison of ion intensities. These techniques have the distinct advantage of requiring no (chemical) labelling of the sample, but the trade-off is the lower accuracy of the quantitation. All of these techniques will be described in this chapter, including several examples of how they are used to answer biomedical questions currently posed in the field.

4. Metabolic labelling

The use of amino acids with either of light or heavy stable N and/or C isotopes in growth medium is an approach that was introduced by the Mann lab [8]. Because the labelling takes place at the very beginning of the proteomics workflow, samples can be mixed at the earliest possible time point. Consequently, the occurrence of systematic errors that may be introduced during sample handling is reduced [8,9]. Although this method has shown to be a powerful way to perform quantitation in proteomics in many different applications, there are also several disadvantages to using metabolic labelling, most importantly the inability for application in human tissue samples. Because the samples need to be metabolically active in order to incorporate the label, this automatically precludes, e.g. blood and biopsy samples. This makes it impossible to use metabolic labelling in a diagnostic setting. Furthermore, for some metabolic labelling approaches, reliable software for data analysis is still lacking. The labelling itself is quite time-consuming, as it takes some time for cell cultures to become completely labelled. Finally, the costs of metabolic labelling approaches may be substantial due to the amount of expensive labelled reagents [10]. Below, several types of metabolic labelling will be discussed in more detail.

4.1. ^{15}N labelling

The use of heavy nitrogen (^{15}N) to label whole model organisms dates back to the 1960s, when it was applied to plants for the first time (see [10] for a review on the matter). In the late 1990s, this strategy took off for other organisms such as *E. coli*, yeast and *Drosophila* [9, 11, 12]. Protein labelling is usually achieved by adding salts containing labelled nitrogen into the medium (in the form of, e.g. NH_4Cl), which will then be metabolised by the organism and finally incorporated into proteins. The advantage of this labelling approach, apart from incorporation at the earliest possible moment, is that labelling of all peptides is guaranteed. However, it also provides increased complexity of the sample, most importantly because the mass difference between the light and the heavy counterpart peptides is basically variable for each peptide pair. This presents a major challenge for suitable bioinformatics solutions for qualitative and quantitative data analysis. ^{15}N labelling has successfully been used in the study of prokaryotes, for instance, the proteome of *S. aureus* was analysed extensively using ^{15}N labelled growth medium [13]. By tying the identified proteins to the completely sequenced genome of *S. aureus*, 80% of all expressed proteins were identified using this approach. A study by Kohlmann and coworkers [14] displayed the power of ^{15}N labelling in shotgun proteomics in analysing *R. eutropha*, a prokaryote that can grow on $^{13}CO_2$ as the sole carbon source and is

used for the industrial production of stable isotope-labelled biomolecules. *R. eutropha* has the ability to switch to lithoautotrophy, i.e. to switch to a source of reduced minerals to satisfy its energy needs. By studying the proteome of *R. eutropha* under both normal and lithoautotrophic conditions, a large upregulation of specific proteins was observed when the prokaryote had to switch its energy source, including chemotaxis-related proteins [14,15].

4.2. ^{13}C labelling

Another prime candidate for metabolic labelling is ^{13}C, as carbon is a key player in protein chemistry. ^{13}C labelling has been successfully used in the determination of protein turnover rates. For instance, by feeding *E. coli* on ^{13}C-labelled glucose, protein turnover rates using only a single culture could be measured by mass spectrometry [16]. Moreover, the method is applicable to shotgun proteomics, which allows for a broader overview of proteins and their turnover rates. In order to reduce the costs of metabolic labelling, a technique called 'subtle modification of isotope ratio proteomics' (SMIRP) was developed [17]. In SMIRP, an increase of only ~1% in isotope ratio can be used to relatively quantify proteins by calculating the ratio of isotopes and comparing it to the variability occurring in nature.

4.3. SILAC

In cultured cells, the metabolic labelling method of choice is stable isotope labelling using amino acids in culture (SILAC), which uses isotopically labelled amino acids (See Figure 2 for a typical SILAC workflow). In order for the amino acids to be incorporated into proteins, it is necessary to determine whether the studied organism is an auxotroph for said amino acid. If a cell or organism is an auxotroph for an amino acid, it cannot synthesize this amino acid itself and, therefore, the amino acid should be supplied in the food or in the growth medium [8]. Usually in SILAC, labelled lysine and arginine are used, which are particularly useful for proteins that are processed with trypsin. Since trypsin cleaves after lysine and arginine, in principle all peptides except for the C-terminal peptide are labelled. If the cells are auxotroph for the selected amino acids, all proteins in a cells are generally completely labelled after several doublings [8]. Conversely, this means that the cells must be dividing, which precludes the use of this technique on primary tissue samples. A complication that has been described in the literature that could potentially interfere with quantitation of SILAC labelled proteins is the natural occurrence of arginine-to-proline conversion. While lysine and arginine are relatively stable in the cell, it is possible for the cell to produce proline from spare arginine, which can then lead to heavy labelled proline. Obviously, this is undesirable and should be accounted for either experimentally or during data analysis (see e.g. [18]).

Labelling using SILAC can also be used to examine post-translational protein modifications such as phosphorylation and ubiquitination in a quantitative manner. An example of this is a phosphoproteomic study in yeast after the knockout of a kinase that plays a role in growth and division [19]. SILAC can in principle be used for any cultured cell type. A recent study from our lab into hormonal signalling in *Drosophila* combined SILAC mass spectrometry with transcriptome analysis [20]. *Drosophila* Kc cells were stimulated with the key insect hormone ecdysone and both mRNA expression and protein expression were studied during a time
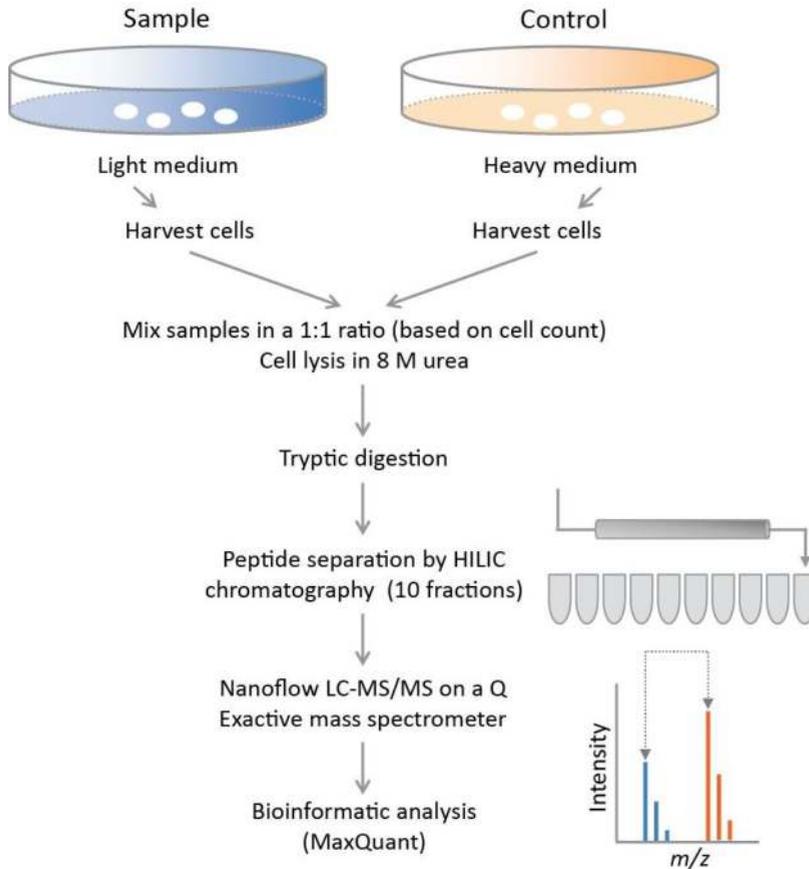
Figure 2. Typical SILAC workflow: cells representing two different biological conditions are grown in either light or heavy medium containing amino acid with stable heavy isotopes. Cells are then harvested and mixed in equal amounts and all sample preparation is performed on the mixed cell populations. In the final mass spectrum, a tryptic peptide will be observed as a peak pair, which represents the two sample conditions. By calculating the peak intensity ratio, the conditions can be compared in a quantitative fashion.

course. The results showed a correlation in the changing levels of mRNA and protein over time, although it became evident that in general there is a time delay between mRNA and protein expression. Not all mRNA–protein pairs showed this delay though, which could be attributed to post-transcriptional regulation events of mRNAs and to variable stability of proteins. Several interesting proteins linked to signalling pathways such as target of rapamycin (TOR) and Notch were identified as being regulated by ecdysone signalling, giving an indication of the scope of the ecdysone system. This study shows the applicability of SILAC in studies where a significant number of proteins are changed, and the correlation between mRNA and protein levels show the quantitative power of SILAC technology, as well as the power of this method to identify signalling networks in cellular systems. In general, there is

also a correlation, albeit weak, between steady-state levels of mRNA and protein (Figure 3). This is mainly true for products that show relatively high expression, which has also been reported in other studies. From this plot, it becomes also clear that for many mRNA products no corresponding protein was identified, illustrating the technical limitations in proteomics that still prevent very low abundant proteins to be detected. In addition, there were several protein products that could not be matched to mRNAs, indicating that, since the intensities of these proteins are generally similar to those with a matched mRNA, this could be attributed to the incomplete annotation of the *Drosophila* database.
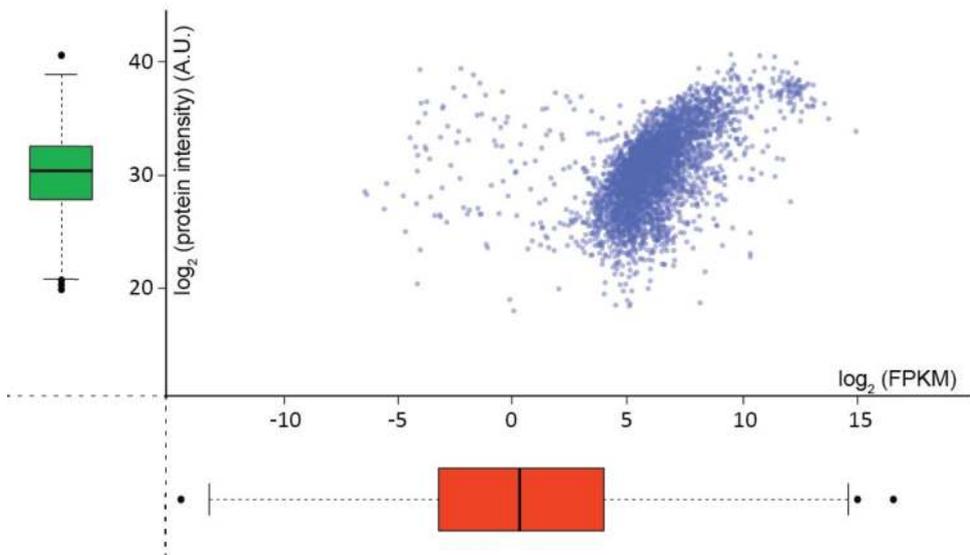

Figure 3. A scatter plot of absolute protein intensities (based on iBAQ values) versus absolute mRNA intensities (based on FPKM values) shows that steady-state levels of protein and mRNA show a weak correlation (R^2= 0.366). The intensity distribution of proteins for which no corresponding hit in the transcriptome analysis was found is represented by the green box plot. This distribution is very similar to the distribution of overlapping hits (blue data points).

An interesting technological progression in the recent years has been the emergence of fully labelled SILAC organisms, such as fruit flies, mice and rats, which allows for *in vivo* quantitative protein analysis [12,21,22]. This allows scientists to study alterations of protein levels in lab mice with as little variation possible, which in turn makes it possible to study the dynamic proteome in tissue. Currently, the generation of SILAC labelled mice is limited by cost considerations due to the expenses required to raise the mice on a diet of labelled food and this has prevented large-scale usage thus far.

Finally, the so called 'super-SILAC' standard is a pool of multiple cell lines that have been labelled using SILAC, which is then spiked into experimental samples. By spiking all the samples with this standard, quantitation becomes possible without the necessity to label the samples themselves using SILAC. This allows the application of SILAC quantitation in patient

tissue, which can evidently not be labelled using traditional SILAC. It should be noted that it is recommended to have a representative sample for the tissue to be studied in the SILAC standard, which limits the usage of this technique to tissues with a representative cell line. For a more in-depth review on this topic, see [23].

5. Chemical labelling strategies

The use of chemical labelling strategies for relative quantitation in proteomics dates back to the late 1990s [24]. The major advantages of using chemical techniques rather than metabolic labelling are the reduced cost and the higher speed of sample processing and analysis. Where labelling cells with SILAC may take up to several days [8], chemical labelling protocols are usually performed in less than an hour [25]. Chemical labelling can be applied to any protein sample, not just metabolically active samples, and some of the techniques allow for a high number of samples to be analysed simultaneously [26]. However, since chemical labelling is done either at the protein level or at the peptide level and at a relatively late stage in the sample preparation protocol, systematic errors are introduced more readily. Also, labelling at the protein level requires specific proteins such as cysteine or lysine, which makes peptides without these amino acids not quantifiable [10,24].

5.1. Labelling with an Isotope-Coded Affinity Tag (ICAT)

The first chemical labelling technique that was described for quantitative mass spectrometry was the isotope-coded affinity tag (ICAT). In ICAT, a thiol reactive group is used to conjugate the tag to cysteine residues in the protein. Apart from the reactive group, the tag has a linker and a biotin moiety. The linker has either eight hydrogen atoms for the light version or eight deuterium atoms for the heavy version, which are used to distinguish two differentially labelled conditions by the 8 Da shift in the mass spectrum [26]. The biotin moiety of the tag can be used to affinity purify the tagged peptides after trypsinisation. The weakness of ICAT lies in the requirement of cysteine residues to be present in the peptide, which leads to a limitation in the amount of peptides tagged. Furthermore, the presence of deuterium causes a shift in elution times when peptides are fractionated using HPLC, which hampers subsequent data analysis [27]. This elution time shift problem was later solved by introducing ^{13}C instead of D into the linker moiety. ICAT labelling has, for instance, been used to investigate the redox state of proteins in a study to the formation of reactive oxygen species and the way this is dealt with by the cell [28]. The ability to use ICAT in human samples has been exploited in screening cerebrospinal fluid samples of Alzheimer patients to find novel prognostic biomarkers [29].

5.2. ICPL

Labelling using isotope-coded protein labels (ICPL) is based on a similar principle as ICAT. In ICPL, lysine residues in intact proteins are labelled, which are more common than cysteine residues. The mass difference between isotope pairs of the labelled and unlabelled peptides

depends on the amount of labelled lysine residues in the peptide and can be determined fairly simply, which provides strong constraints for database searches [30]. A disadvantage of labelling lysine residues is that modifying the residue side chain makes it impossible for trypsin to cleave at this particular lysine residue. As such, this results in much longer peptides after trypsin digestion, as cleavage will only occur after arginine residues, which may lead to proteolytic peptides that cannot be detected. It is therefore recommended to either use another or an additional protease for protein digestion, or to perform the labelling at the peptide level after proteolytic cleavage. A study on tumour cell senescence in which ICPL was successfully used is a good indicator for the power of quantitative proteomics in general. Here, an effect of tumour cell senescence on several important tumourigenesis proteins such as cMYC and key metabolic enzymes such as ATP synthetases were found [31].

5.3. Isobaric tagging

Tandem mass tags (TMT) and isobaric tag for relative and absolute quantitation (iTRAQ) are based on labelling peptides with isobaric tags. Here, the label is conjugated to the N-termini and lysine residues of peptides, so that in principle every peptide is labelled (Figure 4). The various isobaric tags themselves have different masses, but are balanced by a linker moiety that ensures identical intact masses for all possible combinations of tag plus linker. As a consequence, differentially labelled peptides end up in the same precursor peak in the mass spectrum. Only when this peak is subsequently selected for fragmentation, the linkers will be cleaved first, which leads to the appearance of peaks corresponding to the different tags ('reporter ions') in the low m/z region of the spectrum. The relative peak intensities of the tags are then used for quantitation [26]. Since identical peptides end up in the same peak, the complexity of the MS spectrum is not altered as a result of the labelling procedure. Further-more, there are commercial kits available with up to 10 different tags, providing the possibility to run and compare 10 samples simultaneously. The most prominent disadvantage of this method is that the tag, just like most other chemical tags, is incorporated at the peptide level. Also, due to the low m/z values of the reporter ions, not all mass spectrometer types are suitable for detection.

Due to the high number of samples that can be measured in one run, its applicability to human tissue samples and the availability of high-resolution mass spectrometers capable of ion detection in the low m/z region, isobaric tagging has quickly become a popular method for the relative quantitation of proteins. For example, by using iTRAQ labelling for quantitation, differentially regulated phosphorylation sites could be detected that were phosphorylated by ATM/ATR, which are highly conserved kinases key in DNA damage repair [32]. iTRAQ labelling has recently been used to compare the proteome profiles of healthy brains to several prion diseased brains such as Creutzfeldt–Jakob disease [33]. This study showed that the changes in protein expression of different prion diseases are markedly similar, while most changes at the protein level were found in the cerebellum. This study provides an excellent example of biomarker research using mass spectrometry and could be a step towards defining biomarkers for different prion diseases, which are otherwise difficult to classify.
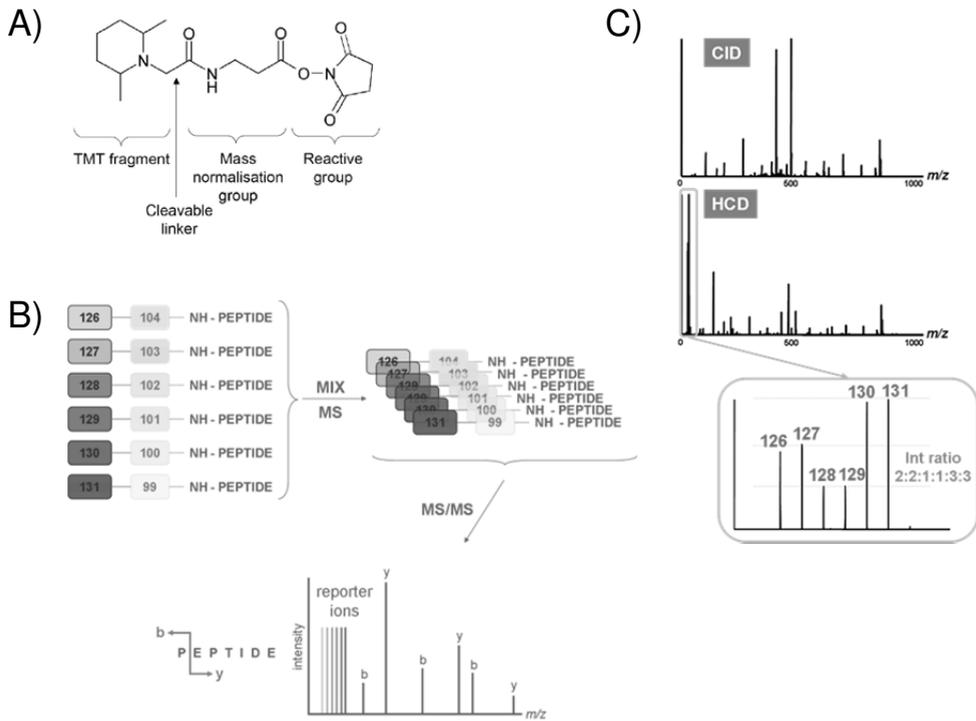
Figure 4. Principle of isobaric tagging. Peptides are tagged with chemical labels that have identical masses due to a delicate balance between individual tag and linker masses. Labelled peptides are then mixed and measured by mass spectrometry. In the mass spectrum, labelled peptides will appear as one peak, but only when these peptides are selected and fragmented by MS/MS, the mass tags will be released from the linker and will show up in the mass spectrum as differential reporter ions in the low *m/z* region. This allows for the relative abundance determination.

5.4. Dimethyl labelling

A simple method of labelling compounds at the peptide level for relative quantitation is dimethylation. Either light-labelled (with H) or heavy-labelled (with D) dimethyl groups are conjugated to the N-terminus of the peptides and to free lysine residue side chains. The advantages of dimethyl labelling include low cost, high speed and possibilities for automated sample preparation. However, since labelling occurs at the peptide level, variation between runs is still inherent to the process [10,34]. The first incarnation of dimethyl labelling was limited to only two different flavours. However, using isotopic isomers ('isotopomers') of formaldehyde with either only D or a combination of ^{13}C and D, up to three different samples can now be compared in a single run [35] (Figure 5). Although this may still be lower than the amount of different labels that can be achieved using isobaric tagging, it is significantly cheaper. Dimethyl labelling can be used for a variety of quantitative measurements, for instance, after a pulldown or immunoprecipitation enrichment protocol. Using an antibody to probe for phosphopeptides in combination with labelling allows one to quantitatively monitor

phosphorylation events [36]. Another possibility that was recently introduced is using dimethylation to study DNA–protein interactions, e.g. by using an oligonucleotide to pull down the proteins and performing the dimethylation labelling on the proteins enriched for [37]. These widely different applications show the power of dimethylation as a quantitative proteomics tool.
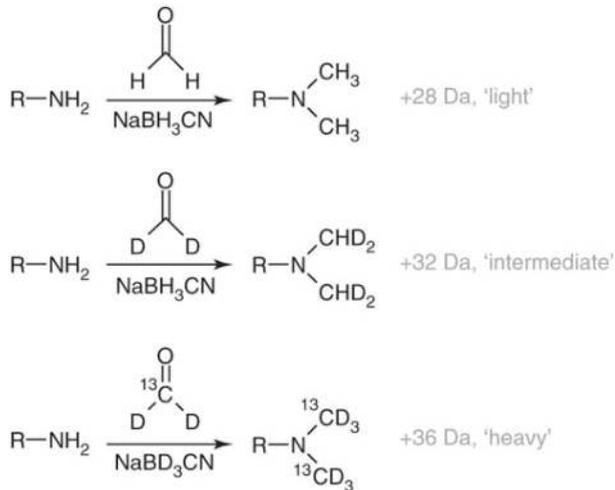

Figure 5. Labelling schemes of triplex stable isotope dimethyl labelling. R = remainder of the peptide. Figure adapted from [35].

5.5. ^{18}O labelling

Another way to differentially label samples for quantitative purposes is the use of heavy oxygen. This labelling method is different from other labelling protocols in that the label incorporation is achieved during the digestion of proteins into peptides. By performing the digestion in water that contains ^{18}O instead of ^{16}O, the carboxyl terminus of every peptide will incorporate two ^{18}O atoms. This method can be incredibly fast, with reports of labelling being achieved in 15 min [25]. A potential pitfall is that the labelling may be incomplete when not performed in a correct manner, leading to multiple peaks in the MS spectrum and therefore resulting into difficulties in quantitation [25,38]. Our lab has described a protocol to avoid incomplete labelling and to assure full incorporation of the heavy oxygen label [39]. By using immobilized trypsin under acidic conditions, all proteolytic peptides could be fully labelled with heavy oxygen with no traces of back-exchange. The labelling protocol was implemented into a protein–protein interaction analysis pipeline to differentiate between *bona fide* interaction partners of the low-level expressing cell cycle regulator cyclin-dependent kinase 9 (Cdk9) and non-specifically binding or background proteins (Figure 6). Previously known, as well as novel, interaction partners of Cdk9 were characterized, among which most notable are the Mediator complex and several other proteins involved in transcriptional regulation. It was

shown that a differential proteomics approach based on ^{18}O labelling provides a valuable method for high-confidence determination of protein interaction partners and is easily implemented in protein network analysis workflows.
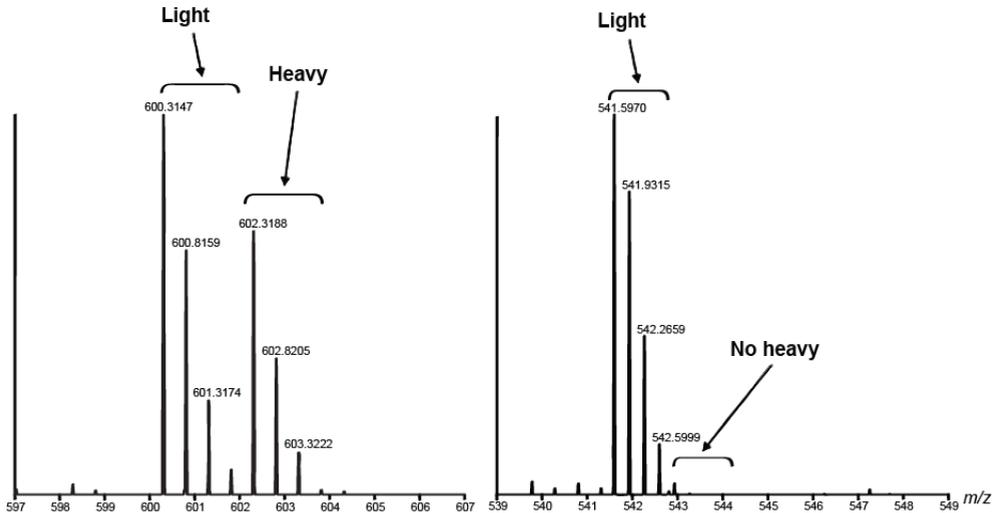

Figure 6. Example MS spectra of tryptic peptides from a 1:1 mixture of a Cdk9 co-IP experiment (light) and a control IP sample (heavy). (A) The tryptic peptide LGTPELSPTER, which originates from the contaminant acetyl-CoA carboxy-lase shows both the light and heavy forms of the peptide, and as such, is a non-specific protein. (B) The peptide GPPEETGAAVFDHPAK, of cyclin T1, can only be detected in the light sample and is therefore an interactor with Cdk9 [39].

Another method to achieve consistent labelling is to use alternative proteases besides trypsin, e.g. β-lactamase [40], which eliminates the incorporation of two heavy oxygen atoms and limits it to one atom consistently.

6. Absolute Quantitation (AQUA)

All label-based approaches described above are geared towards generating relative quantitative measurements. In many cases though, it would be interesting to measure absolute quantities of proteins instead. In order to gain absolute quantitation results, synthesized peptides or proteins containing heavy isotope labels that correspond to the target peptide or protein of interest can be spiked into the sample at a known concentration, after which the intensities of target and standard can be compared to one another. Obviously, the standard peptide can be modified with one or multiple post-translational modifications if needed [41]. Due to the fact that this spiked standard provides absolute rather than relative quantitation,

this technique has been dubbed absolute quantitation (AQUA). Spike-in components that can be used for AQUA include peptides with stable isotopes incorporated into one or several amino acids [41], a construct in which several peptides are strung together (which has the added advantage of being able to quantify multiple peptides in one run [42]), or an entirely labelled protein to quantify the amount of protein [43]. As with other quantitation techniques, the stage at which the label is incorporated largely determines the extent of the systematic quantitation error that is introduced into the sample. In studying hormonal influence on blood pressure, and more specifically angiotensin II, spiking in the synthesized heavy labelled angiotensin has been used to absolutely quantify protein levels in plasma. As such, it was shown that chronic kidney disease patients had strongly increased levels of angiotensin II [44]. These results show that AQUA can be useful in the field of biomarker research, although it has many more applications, such as in assessing the levels of enzymes in prokaryotes [45].

7. Label-free quantitation

With the development of better and faster mass spectrometers with higher sensitivity and heavier duty cycles, the number of studies that use label-free quantitation (LFQ) methods has increased over the past few years. The obvious advantage of LFQ is that no sample processing other than the standard LC-MS procedures is needed. Furthermore, there is no need for often expensive labelling kits. There are two major approaches employed in label-free quantitation: spectral counting and intensity-based quantitation. Quantitation by spectral counting is based on the observation that peptides that are more abundant will be detected and fragmented more often by the mass spectrometer, and as such the MS/MS count gives information about the abundance of the protein. However, there are several issues that should be taken into account here. In general, larger proteins generate more proteolytic peptides, which increases the chance that multiple peptides for one such protein are detected. Furthermore, in principle every peptide has different physicochemical properties, which influence the ionizability and, therefore, the detectability in the mass spectrometer. To address this, several modifications of spectral counting have been developed, which incorporate mathematical corrections, such as introducing a normalised spectral abundance factor into the equation to account for protein length variability (e.g. emPAI [46]). In intensity-based quantitation, on the other hand, the quantitation is based on the total amount of peptide that is detected in a specific retention time window for which the area under the curve in the chromatogram is accurately determined (extracted ion currents (or XICs) of peptides). LFQ has benefited greatly from recent develop‐ ments in mass spectrometer hardware as it increases the number of quantifiable features present in a given LC-MS run and allows averaging over more peptides for protein quantita‐ tion [47]. In order for the ion intensity quantitation to be reproducible, normalization steps are required as differences in the total amount of protein loaded onto the LC-MS system and instrument variances need to be accounted for. Because of this, powerful software is required and has been developed to perform this type of peptide and protein quantitation (see [48] for an in-depth review). An interesting label-free quantitation technique has been described that combines peptide counting, spectral counting and ion intensities into the so-called normalized

spectral index [49]. Using this method, the variance between multiple LC-MS runs was largely eliminated. This method shows great promise in achieving reproducible label-free quantitation.

8. Software applications for quantitative mass spectrometry

Quantitative proteomic data are typically very complex and the data analysis requires specialized software. The main challenge concerns incomplete data, as even modern advanced mass spectrometers cannot sample and fragment every peptide ion present in a complex sample. As a consequence, only a subset of peptides and proteins present in a sample can be identified. Over the past years, several strategies for mass spectrometry-based quantitative proteomics and corresponding computational methodology for the processing of quantitative data sets have been developed (reviewed in ([50,51]), as different quantitative LC-MS methods require different software solutions for data analysis. Quantitation can be achieved by comparing peak intensities in differential stable isotopic labelling, via spectral counting, or by using the ion current in label-free LC-MS measurements. Many software solutions have been published and can be used freely, with specific instrument compatibility and processing functionality which can deal with these basically different quantitation methods. The researcher has to choose the appropriate software solution for his quantitative proteomic experiments based on the experimental and analytical requirements. Since it goes beyond the scope of this chapter to discuss all of the available software tools separately, we refer the reader to an extensive and up-to-date overview of software solutions including links to websites for downloads at http://www.ms-utils.org.

9. Concluding remarks

In summary, all of the mass spectrometry-based quantitation methods have their particular strengths and weaknesses and the researcher has to choose the best method from the multitude of methods that have emerged for the analysis of simple and complex (sub-) proteomes using quantitative mass spectrometry for his specific research. This choice depends on the availability of high-resolution mass spectrometer and LC equipment, the available expertise present in the lab and the financial aspects involved. Quantitative proteomics methods have become mature and can now be applied at a large scale to the study of proteomes and their dynamics. Using the labelling methods described in this chapter, thousands of proteins can be identified and quantified in a single experiment. However, there is still room for improvements to both the experimental strategies for the quantitative analysis of very complex mixtures and of their post-translational modifications and to appropriate bioinformatics and statistical approaches in order to obtain meaningful interpretations of the results. The ultimate goal is to generate quantitative proteomic data at a scale that would allow the comprehensive investigation of a biological phenomenon.

Author details

Lennart van der Wal and Jeroen A. A. Demmers[*]

*Address all correspondence to: j.demmers@erasmusmc.nl

Erasmus University Medical Center, Rotterdam, The Netherlands

References

[1] "The Nobel Prize in Chemistry 2002". Nobelprize.org. Nobel Media AB 2014. Web. 3 Feb 2015. <http://www.nobelprize.org/nobel_prizes/chemistry/laureates/2002/>.

[2] Glish GL, Burinsky DJ. Hybrid mass spectrometers for tandem mass spectrometry. J Am Soc Mass Spectrom. 2008;19:161–72.

[3] Scigelova M, Hornshaw M, Giannakopulos A, Makarov A. Fourier transform mass spectrometry. Mol Cell Proteomics 2011;10:M111.009431.

[4] Lucitt MB, Price TS, Pizarro A, Wu W, Yocum AK, Seiler C, et al. Analysis of the zebrafish proteome during embryonic development. Mol Cell Proteomics 2008;7:981–94.

[5] Byrne JC, Downes MR, O'Donoghue N, O'Keane C, O'Neill A, Fan Y, et al. 2D-DIGE as a strategy to identify serum markers for the progression of prostate cancer. J Proteome Res. 2009;8:942–57.

[6] Chakravarty B, Gallagher SR, Chakravarty DN. Difference gel electrophoresis (DIGE). Curr Protoc Mol Biol 2005;10(23):1–8.

[7] Ong S-E, Mann M. Mass spectrometry-based proteomics turns quantitative. Nat Chem Biol [Internet]. Nature Publishing Group; 2005 Oct;1(5):252–62. Available from: http://dx.doi.org/10.1038/nchembio736

[8] Ong S-E, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, et al. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. Mol Cell Proteomics 2002;1:376–86.

[9] Oda Y, Huang K, Cross FR, Cowburn D, Chait BT. Accurate quantitation of protein expression and site-specific phosphorylation. Proc Natl Acad Sci U S A 1999;96(June): 6591–6.

[10] Gouw JW, Krijgsveld J, Heck AJR. Quantitative proteomics by metabolic labeling of model organisms. Mol Cell Proteomics 2010;9:11–24.

[11] Paša-Tolic L, Jensen P. High throughput proteome-wide precision measurements of protein expression using mass spectrometry. J Am Chem Soc [Internet]. 1999;121:7949–50. Available from: http://pubs.acs.org/doi/pdf/10.1021/ja991063o

[12] Sury MD, Chen J-X, Selbach M. The SILAC fly allows for accurate protein quantification in vivo. Mol Cell Proteomics 2010;9:2173–83.

[13] Becher D, Hempel K, Sievers S, Zühlke D, Pané-Farré J, Otto A, et al. A proteomic view of an important human pathogen: towards the quantification of the entire staphylococcus aureus proteome. PLoS One 2009;4(12): e8176.

[14] Kohlmann Y, Pohlmann A, Otto A, Becher D, Cramm R, Lütte S, et al. Analyses of soluble and membrane proteomes of Ralstonia eutropha H16 reveal major changes in the protein complement in adaptation to lithoautotrophy. J Proteome Res 2011;10:2767–76.

[15] Pohlmann A, Fricke WF, Reinecke F, Kusian B, Liesegang H, Cramm R, et al. Genome sequence of the bioplastic-producing "Knallgas" bacterium Ralstonia eutropha H16. Nat Biotechnol 2006;24(September 2006):1257–62.

[16] Cargile BJ, Bundy JL, Grunden AM, Stephenson JL. Synthesis / degradation ratio mass spectrometry for measuring relative dynamic protein turnover. Anal Chem 2004;76(1):86–97.

[17] Whitelegge JP, Katz JE, Pihakari KA, Hale R, Aguilera R, Gómez SM, et al. Subtle modification of isotope ratio proteomics; an integrated strategy for expression proteomics. Phytochemistry 2004;65:1507–15.

[18] Van Hoof D, Pinkse MWH, Oostwaard DW-V, Mummery CL, Heck AJR, Krijgsveld J. An experimental correction for arginine-to-proline conversion artifacts in SILAC-based quantitative proteomics. Nat Methods 2007;4(9):677–8.

[19] Kettenbach AN, Deng L, Wu Y, Baldissard S, Adamo ME, Gerber SA, et al. Quantitative phosphoproteomics reveals pathways for coordination of cell growth and division by the fission yeast DYRK kinase Pom1. MCP, Press. 2015.

[20] Sap KA, Bezstarosti K, Dekkers DHW, Van den Hout M, Van IJcken W, Rijkers E, et al. Global quantitative proteomics reveals novel factors in the ecdysone signaling pathway in Drosophila melanogaster. Proteomics 2014;(00):1–14.

[21] Krüger M, Moser M, Ussar S, Thievessen I, Luber CA, Forner F, et al. SILAC mouse for quantitative proteomics uncovers kindlin-3 as an essential factor for red blood cell function. Cell 2008;134:353–64.

[22] McClatchy DB, Dong MQ, Wu C, Venable JD, Yates III JR. 15N metabolic labelling of mammalian tissue with slow protein turnover. J Proteome Res 2007;6:2005–10.

[23] Shenoy A, Geiger T. Super-SILAC: current trends and future perspectives. Expert Rev Proteomics [Internet]. 2015;12(1):13–9. Available from: http://informahealth-care.com/doi/abs/10.1586/14789450.2015.982538

[24] Gygi SP, Rist B, Gerber S a, Turecek F, Gelb MH, Aebersold R. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nat Biotechnol 1999;17(October):994–9.

[25] Mirza SP, Greene AS, Olivier M. 18O labeling over a coffee break: a rapid strategy for quantitative proteomics. J Proteome Res 2008;7:3042–8.

[26] Ross PL, Huang YN, Marchese JN, Williamson B, Parker K, Hattan S, et al. Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. Mol Cell Proteomics 2004;3:1154–69.

[27] Coombs KM. Quantitative proteomics of complex mixtures. Expert Rev Proteomics 2011;8(5):659–77.

[28] Leichert LI, Gehrke F, Gudiseva H V, Blackwell T, Ilbert M, Walker AK, et al. Quantifying changes in the thiol redox proteome upon oxidative stress in vivo. Proc Natl Acad Sci U S A 2008;105:8197–202.

[29] Zhang J, Goodlett DR, Quinn JF, Peskind E, Kaye J a, Zhou Y, et al. Quantitative proteomics of cerebrospinal fluid from patients with Alzheimer disease. J Alzheimers Dis 2005;7:125–33; discussion 173–80.

[30] Schmidt A, Kellermann J, Lottspeich F. A novel strategy for quantitative proteomics using isotope-coded protein labels. Proteomics 2005;5:4–15.

[31] Biniossek ML, Lechel A, Rudolph KL, Martens UM, Zimmermann S. Quantitative proteomic profiling of tumor cell response to telomere dysfunction using isotope-coded protein labeling (ICPL) reveals interaction network of candidate senescence markers. J Proteomics [Internet]. Elsevier B.V.; 2013;91:515–35. Available from: http://dx.doi.org/10.1016/j.jprot.2013.08.007

[32] Roitinger E, Hofer M, Köcher T, Pichler P, Novatchkova M, Yang J, et al. Quantitative phosphoproteomics of the ATM and ATR dependent DNA damage response in Arabidopsis thaliana. Mol Cell Proteomics 2015;Preliminai.

[33] Shi Q, Chen L-N, Zhang B-Y, Xiao K, Zhou W, Chen C, et al. Proteomics analyses for the global proteins in the brain tissues of different human prion diseases. Mol Cell Proteomics 2015;In press.

[34] Hsu JL, Huang SY, Chow NH, Chen SH. Stable-isotope dimethyl labeling for quantitative proteomics. Anal Chem 2003;75(24):6843–52.

[35] Boersema PJ, Aye TT, Van Veen TAB, Heck AJR, Mohammed S. Triplex protein quantification based on stable isotope labeling by peptide dimethylation applied to cell and tissue lysates. Proteomics 2008;8:4624–32.

[36] Giansanti P, Stokes MP, Silva JC, Scholten A, Heck AJR. Interrogating cAMP-dependent kinase signaling in Jurkat T cells via a protein kinase A targeted immune-precipitation phosphoproteomics approach. Mol Cell Proteomics [Internet] 2013;12(18):3350–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23882029

[37] Hubner NC, Nguyen LN, Hornig NC, Stunnenberg HG. A quantitative proteomics tool to identify DNA–protein interactions in primary cells or blood. J Proteome Res 2015;14:1315–29.

[38] Zhao Y, Jia W, Sun W, Jin W, Guo L, Wei J, et al. Combination of improved 18o incorporation and multiple reaction monitoring: a universal strategy for absolute quantitative verification of serum candidate biomarkers of liver cancer. J Proteome Res 2010;9:3319–27.

[39] Bezstarosti K, Ghamari A, Grosveld FG, Demmers JAA. Differential proteomics based on 18O labeling to determine the cyclin dependent kinase 9 interactome. J Proteome Res 2010;9:4464–75.

[40] Wang M, Shen Y, Turko I V, Nelson DC, Li S. Determining carbapenemase activity with 18O labeling and targeted mass spectrometry. Anal Chem 2013;85:11014–9.

[41] Gerber S a, Rush J, Stemman O, Kirschner MW, Gygi SP. Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. Proc Natl Acad Sci U S A 2003;100:6940–5.

[42] Rivers J, Simpson DM, Robertson DHL, Gaskell SJ, Beynon RJ. Absolute multiplexed quantitative analysis of protein expression during muscle development using QconCAT. Mol Cell Proteomics 2007;6:1416–27.

[43] Brun V, Dupuis A, Adrait A, Marcellin M, Thomas D, Court M, et al. Isotope-labeled protein standards: toward absolute quantitative proteomics. Mol Cell Proteomics 2007;6:2139–49.

[44] Schulz A, Jankowski J, Zidek W, Jankowski V. Absolute quantification of endogenous angiotensin II levels in human plasma using ESI-LC-MS/MS. Clin Proteomics [Internet]. 2014;11(1):37. Available from: http://www.clinicalproteomicsjournal.com/content/11/1/37

[45] Voges R, Corsten S, Wiechert W, Noack S. Absolute quantification of Corynebacterium glutamicum glycolytic and anaplerotic enzymes by QconCAT. J Proteomics [Internet]. Elsevier B.V.; 2015;113:366–77. Available from: http://dx.doi.org/10.1016/j.jprot.2014.10.008

[46] Ishihama Y, Oda Y, Tabata T, Sato T, Nagasu T, Rappsilber J, et al. Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. Mol Cell Proteomics 2005;4:1265–72.

[47] Cox J, Hein MY, Luber CA, Paron I, Nagaraj N, Mann M. Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extrac-

tion, Termed MaxLFQ. Mol Cell Proteomics [Internet]. 2014;13:2513–26. Available from: http://www.mcponline.org/cgi/doi/10.1074/mcp.M113.031591

[48] Sandin M, Teleman J, Malmström J, Levander F. Data processing methods and quality control strategies for label-free LC-MS protein quantification. Biochim Biophys Acta [Internet]. Elsevier B.V.; 2014;1844(1):29–41. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23567904

[49] Griffin NM, Yu J, Long F, Oh P, Shore S, Li Y, et al. Label-free, normalized quantification of complex mass spectrometry data for proteomics analysis. Nat Biotech 2010;28(1):83–9.

[50] Matthiesen R, Azevedo L, Amorim A, Carvalho AS. Discussion on common data analysis strategies used in MS-based proteomics. Proteomics 2011;11:604–19.

[51] Mueller LN, Brusniak M-Y, Mani DR, Aebersold R. An assessment of software solutions for the analysis of mass spectrometry based quantitative proteomics data. J Proteome Res 2008;7:51–61.