Chapter 2
What are the Common Mass Spectrometry-Based Analyses Used in Biology?

Abstract Mass spectrometry is used in many field of research, such as biology, chemistry, geology, etc. The focus of this chapter is the common methods, requiring mass spectrometry, in biology related researches. Proteomics for example is a field of research focusing on proteins for which mass spectrometry plays a pivotal role. However, proteins are not the unique target, lipids and small compounds such as metabolites are also studied using mass spectrometry. As we are now in the era of 'omics' the field of research studying lipids is called lipidomics and the analysis of metabolites is known as metabolomics. Through the example of the main methods used in proteomics analysis, this chapter summarizes the advantages of mass spectrometry in biology research. The analyses of small biomolecules, lipids and nucleotides are also presented.

Keywords Mass spectrometry · Proteomics · Proteins · Metabolomics · Metabolites · Lipidomics · Lipids · Nucleotides

As mentioned in Chap. 1, this work focuses on studies of proteins because they represent an important part of the biology of cells and tissues. They are also a common target for the development of techniques involving mass spectrometry. Proteins are macromolecules that represent the functional element of any living organism. Proteins are complex molecules with four structure levels, the primary structure is the amino acid sequence coded by the genome, enclosed in the DNA. The primary structure or sequence of the protein contributes to the secondary structure of the molecule. This structure represents the partial folding of the molecule. Some portions of the proteins adopt a specific conformation in space; these conformations are the secondary structure of the protein. Then, along the sequence of proteins secondary structures appear. The whole protein is not linear but will be folded to form, in general, a globular structure considered as the ternary structure. Finally, the last level of conformation of the protein is the quaternary structure, which is the formation of multi-molecular structures including two or more folded proteins. The understanding of proteins structure and function is then essential to study biology.

Measuring the molecular weight or mass of proteins does not provide sufficient information to allow the protein identification. Considering that the mass of a protein may correspond to a large number of possibilities corresponding to different

G. Pottiez, *Mass Spectrometry: Developmental Approaches to Answer Biological Questions,* SpringerBriefs in Bioengineering, DOI 10.1007/978-3-319-13087-3_2

proteins, it would be impossible, only with a single mass value, to obtain an identification. Also, the molecular weight of proteins can vary between few thousands of Daltons to millions of Daltons. Then, measuring the entire range of mass would considerably reduce the precision of the measurement. As powerful as a mass spectrometer could be, it is not possible to measure several proteins with a mass range from less than a thousand Dalton to a million Dalton. Thus, protein samples need to be processed in order to reduce the mass range. To do so, proteins are cleaved into pieces called peptides. This process modifies the range of mass, because after cleavage, most of the peptides have a mass between 500 and 2000. This process of cleavage of proteins could be performed chemically, *i.e.,* the proteins are incubated in acid for 2 h (see box 2.1.), which could be accelerated by the use of microwave. It is also possible to use enzymes to cleave proteins. However, it is not viable to use exoproteases, *exo* refers to the external part of the protein, and thus such proteases cleave the amino acids one by one starting at one end of the primary structure of the protein. There is exopeptidase specific for the N-terminal and exopeptidase specific for C-terminal. The cleavage of the amino acids one after another obliterates the sequence of the protein and thus the information leading to the protein identification. In certain circumstances the successive elimination of amino acid residues could be informative. It was the principle of Edmann sequencing (see Box 2.2). In order to keep the protein sequence intact, or at least some parts of the sequence unchanged, it is then necessary to use endopeptidases. Those latter target specific amino acids within the sequence of the protein, leaving unmodified the amino acid sequence of the peptides generated. However, the choice of protease is also important, the criteria to take into consideration are (i) the cleavage site (ii) the frequency of the targeted cleavage amino acids (iii) the specificity of the enzyme and (iv) the efficiency of the enzyme. Table 2.1 summarizes the proteases most commonly used in the sample preparation for mass spectrometry analysis. The choice of enzyme obviously depends on the design of the study, the samples and the protocol. When required in this work the choice of enzyme will be explicated.

Box 2.1. Protocols for protein digestion

Protocol for chemical degradation of proteins	Protocol for enzymatic degradation of proteins
For gel pieces containing proteins	*For gel pieces containing proteins*
Freshly prepared 2% formic acid (approximate pH = 2)	Cover the gel pieces with the enzyme solution
	The volume depends on the size of the gel pieces
Working with dried destained pieces of gel	Once the gel pieces are rehydrated, discard the supernatant
Cover the gel pieces with 2% Formic acid solution	Cover the gel with 20 mM ammonium bicarbonate solution

Protocol for chemical degradation of proteins	Protocol for enzymatic degradation of proteins
The volume depends on the size of the gel pieces	Incubate overnight at 37 °C
Once the gel pieces are rehydrated, incubate for at least 2 h at 100 °C	Collect the supernatant in a clean microcentrifuge tube
Recommended 108 °C	Cover the gel pieces with a solution of 20 mM ammonium bicarbonate and 50 % acetonitrile
After incubation, allow the samples to cool down at room temperature	Collect the supernatant
Collect the supernatant in a clean microcentrifuge tube	Cover the gel pieces with acetonitrile
Dry the peptide samples	Collect the supernatant
	Dry the peptide samples
The peptides could be used for mass spectrometry analysis	The peptides could be used for mass spectrometry analysis
For samples of proteins	*For samples of proteins*
Either in solution	*Either in solution*
Add formic acid to the solution containing the peptide	Add ammonium bicarbonate 100 mM
Final concentration of 2 %	Final concentration of 20 mM
Or dried proteins	*For dried proteins*
Dissolve the proteins with 2 % Formic acid solution	Dissolve the proteins with 20 mM ammonium bicarbonate solution
50–100 µL	50–100 µL
Incubate for at least 2 h at 100 °C	Add the solution contain the trypsin
Recommended 108 °C	Recommended ≈ 12 ng of trypsin for 1 g of proteins
	Incubate overnight at 37 °C
Dry the peptide samples	Dry the peptide samples
The peptides could be used for mass spectrometry analysis	The peptide samples could be used for mass spectrometry

Box 2.2. Edmann Sequencing

Method of sequencing of proteins or peptide. For this method, the N-terminal residues are removed from the proteins one by one.

The N-terminal amino acid of the protein is labeled by reacting with Phenylisothiocyanate, is cyclized and is then cleaved from the protein leaving a free amino acid N+1. The amino acid is released as derived form. Each

derived amino acid has specific physicochemical properties allowing their identification, using for example chromatography or electrophoresis. With cycles of acid/base conditions this procedure could be repeated several times in order to identify the first amino acid residues of a sequence.

The limitations of this techniques are, first, the sequencing is limited, in the best conditions to approximately 30 amino acid residues being sequenced. Second, the N-terminal amino groups must be free in order to perform the initial reaction for the sequencing.

Trypsin is the enzyme the most frequently utilized in proteomics for the sample preparation. The reasons for this choice are, first, this enzyme is highly specific, its targets are lysine and arginine residues, cleaving at the C-terminal end of those residues. It is difficult to obtain a hundred percent efficiency in vitro with an enzyme, however trypsin is stable and highly efficient and targets most of the lysine and arginine residues. In addition, trypsin has a low rate of non-specific cleavages and only one exception is known, trypsin does not cleave lysine or arginine residues when followed by a proline residue. Finally, by cleaving at the carboxyl end of lysine and arginine residues, the newly formed peptide has an amino acid residue with a basic side chain at its C-terminal end. This increases the ability of the peptides to be positively ionized, which is an advantage for mass spectrometry analysis. The enzymes Lys-C and Arg-C have the same properties but as they are restricted to lysine and arginine respectively, the rate of peptide per protein is lower, which leads to less and longer peptides.

Mass spectrometry is a powerful technique able to provide the mass information of any biological compounds with high precision, but this technique is also limited. For example, the "Dream" instrument would allow the mass measurement of all compounds in any unprocessed sample. However, biological samples such as blood sample or tissues contain hundreds of thousands of proteins, carbohydrates, lipids, salts as well as chemicals. Taking such samples, mass spectrometry analysis, without prior processing, would lead to an unreadable set of data, if any results could be obtained. This is why, most of the research projects involving mass spectrometry require to perform a sample preparation before the actual measurement. It is also important to know that many research projects consist in method development based on sample preparation. It is impossible to precisely determine the level of influence of the sample preparation because it changes according to the experiment but in any case will influence the results.

As previously explained, protein samples are complex; they contain thousands of proteins, all different in size, structure and properties. By cleaving the proteins with endoproteases the range of mass has been reduced. But, every protein will be represented by several peptides. On average, proteins will have a mass of $\approx$ 50,000 Da, considering that such protein will be cleaved in peptides weighing on average 1500 Da, then, such protein would produce more than 30 different peptides. So, on one hand the protein cleavage simplifies the samples, by reducing the range of

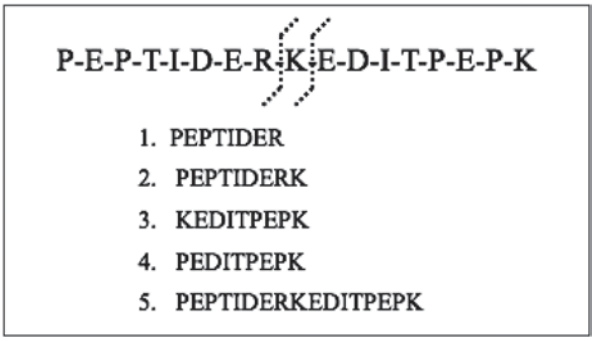

Fig. 2.1 Example of the sequence of a peptide including two cleavage sites for trypsin (arginine R and lysine K). After digestion this peptide produces five peptide fragments.

mass, but, on the other hand, multiplies the number of molecules for a single protein. As an example, considering twenty of the most abundant proteins in plasma samples and cell cytoplasm and performing a theoretical digestion with trypsin, plasma proteins [2, 3] are cleaved into 1648 peptides, with a mass range from 146 to 6655 Da. On the other hand, the theoretical digest of the proteins of the cytoplasm [4] lead to 649 peptides with a mass range from 131 to 6097 Da. This shows first how widely spread the mass range can be after digest and second, how many peptides will be produced by enzymatic digestion. However, this is an example, which considers that trypsin cleaves at all the lysine and arginine residues. Practically, trypsin, as well as other enzymes, does not necessarily reach every lysine and arginine residues and therefore partially cleaves the proteins. Then a single fraction could be represented with different forms. For example, as shown in Fig. 2.1. a hypothetical peptide with 2 cleavage sites could be present in five different forms. The site in the peptide that is not cleaved is called missed cleavage. They are frequently seen in processed samples and add a level of complexity to the digested sample.

Digesting a sample containing proteins leads to the production of a sample more homogeneous but also more complex. As shown in the example above, enzymatic digestion provides at least thirty times more compounds than the starting sample. Nonetheless, this drawback has been turned into an advantage for protein identification. Indeed, instead of looking for only one compound to identify a protein, or in the case of mass spectrometry one mass to identify the protein, the researcher seek for several pieces of a protein and combine the different information to finally obtain the identification of a single protein. Using software, the mass information and/or sequence information are used to, *in silico* "*rebuild*" the protein sequence, presenting the level of certainty for the identification. Briefly, the peptides with a mass or a sequence corresponding to one protein are gathered. Then, information such as the percentage of coverage can be determined (see Fig. 2.2). This is described in more details below.

Prior to present the some strategies available for the sample preparation it is worth knowing that among the evolutions of mass spectrometry, there was first the coupling of gas chromatography to mass spectrometers. But then, the coupling of liquid chromatography to the mass spectrometer which has made the analysis of peptides samples easier and more accessible (see Chap. 1 and Fig. 1.2.). As

a

```
                10         20         30         40         50
001  KWVTFISLL LLFSSAYSRG VFRRDTHKSE IAHRFKDLGE EHFKGLVLIA

051 FSQYLQQCPF DEHVKLVNEL TEFAKTCVAD ESHAGCEKSL HTLFGDELCK

101 VASLRETYGD MADCCEKQEP ERNECFLSHK DDSPDLPKLK PDPNTLCDEF

151 KADEKKFWGK YLYEIARRHP YFYAPELLYY ANKYNGVFQE CCQAEDKGAC

201 LLPKIETMRE KVLASSARQR LRCASIQKFG ERALKAWSVA RLSQKFPKAE

251 FVEVTKLVTD LTKVHKECCH GDLLECADDR ADLAKYICDN QDTISSKLKE

301 CCDKPLLEKS HCIAEVEKDA IPENLPPLTA DFAEDKDVCK NYQEAKDAFL

351 GSFLYEYSRR HPEYAVSVLL RLAKEYEATL EECCAKDDPH ACYSTVFDKL

401 KHLVDEPQNL IKQNCDQFEK LGEYGFQNAL IVRYTRKVPQ VSTPTLVEVS

451 RSLGKVGTRC CTKPESERMP CTEDYLSLIL NRLCVLHEKT PVSEKVTKCC

501 TESLVNRRPC FSALTPDETY VPKAFDEKLF TFHADICTLP DTEKQIKKQT

551 ALVELLKHKP KATEEQLKTV MENFVAFVDK CCAADDKEAC FAVEGPKLVV

601 STQTALA
```

b

mass	position	peptide sequence	1014.6193	549-557	QTALVELLK
2435.2427	45-65	GLVLIAFSQYLQQCPFDEHVK	1011.42	413-420	QNCDQFEK
1955.9596	319-336	DAIPENLPPLTADFAEDK	1002.583	598-607	LVVSTQTALA
1888.9268	169-183	HPYFYAPELLYYANK	977.4509	123-130	NECFLSHK
1850.8993	529-544	LFTFHADICTLPDTEK	974.4577	37-44	DLGEEHFK
1823.8996	508-523	RPCFSALTPDETYVPK	927.4934	161-167	YLYEIAR
1667.8131	469-482	MPCTEDYLSLILNR	922.488	249-256	AEFVEVTK
1633.6621	184-197	YNGVFQECCQAEDK	886.4152	131-138	DDSPDLPK
1578.5981	267-280	ECCHGDLLECADDR	841.46	483-489	LCVLHEK
1567.7427	347-359	DAFLGSFLYEYSR	818.4254	562-568	ATEEQLK
1519.7461	139-151	LKPDPNTLCDEFK	789.4716	257-263	LVTDLTK
1511.8427	438-451	VPQVSTPTLVEVSR	752.3573	341-346	NYQEAK
1497.6314	387-399	DDPHACYSTVFDK	725.2593	581-587	CCAADDK
1479.7954	421-433	LGEYGFQNALIVR	712.3736	29-34	SEIAHR
1399.6926	569-580	TVMENFVAFVDK	703.4097	212-218	VLASSAR
1388.5708	375-386	EYEATLEECCAK	701.4014	198-204	GACLLPK
1386.6206	286-297	YICDNQDTISSK	689.3729	236-241	AWSVAR
1364.4803	106-117	ETYGDMADCCEK	660.3563	490-495	TPVSEK
1362.6722	89-100	SLHTLFGDELCK	658.3155	118-122	QEPER
1349.546	76-88	TCVADESHAGCEK	649.3338	205-209	IETMR
1305.7161	402-412	HLVDEPQNLIK	649.3338	223-228	CASIQK
1283.7106	361-371	HPEYAVSVLLR	609.2878	524-528	AFDEK
1177.5591	300-309	ECCDKPLLEK	545.3405	101-105	VASLR
1163.6306	66-75	LVNELTEFAK	537.282	157-160	FWGK
1052.4499	460-468	CCTKPESER	517.298	281-285	ADLAK
1050.4924	588-597	EACFAVEGPK	509.3194	558-561	HKPK
1024.455	499-507	CCTESLVNR	508.2514	229-232	FGER
1015.4877	310-318	SHCIAEVEK	500.2463	25-28	DTHK

Fig. 2.2 **a** Sequence of the protein Bovine Serum Albumin, the cleavage sites of trypsin are highlighted in *blue* and the lysine and arginine residues followed by a proline (*inhibited site for trypsin cleavage*) are highlighted in *red*. The peptides of a theoretical digest of BSA with trypsin are underlined. **b** The theoretical digest was perform with the algorithm PeptideMass (*http://web.expasy.org/peptide_mass/*) with the options: Trypsin, No Missed cleavage, No modifications, Mass peptides 500—2500 Da. This example shows that 100 % digestion of BSA does not produce a 100 % coverage, here the coverage is 88 %, considering that all peptides created are measured by mass spectrometry, which is rarely correct.

described in the introduction, late 1980's ESI source was used for macromolecules. This ion source gives the opportunity to ionize peptides. It also has the advantage of allowing a direct coupling of the end of the chromatography column to the analyzer. The mass spectrometer performs the analysis directly on the fractions of the samples eluted from the chromatography column. Such coupling of the liquid chromatography system to the mass spectrometer is called on line. On the other hand, off-line method exists and qualifies the fractionation of the samples, the collection of some or all the fractions and the subsequent analysis of the collected fractions. This is for example the case for liquid chromatography couple to MALDI ion source, where the fraction are collected and immediately co-crystalized on a MALDI plate. Then, the mass analysis is performed separately. This represent one method of separation and simplification of the sample. In the article [5] the study presented the analysis by liquid chromatography coupled to a mass spectrometer of the protein digest of a tissue. However, the complexity of the sample led to the identification, at first, of a few proteins. Only the most abundant were identified. On the other hand, if prior to the liquid chromatography the sample was simplified or fractionated, it was possible to identify hundreds of compounds. This demonstrates more sample did not mean more identification. In this particular example, additional separation steps were required. It also shows how critical the sample preparation could be. For protein samples, another possibility to simplify the samples, prior to the mass spectrometry analysis is the in-gel separation of proteins, which has been tremendously improved by the development of the two dimensional gel electrophoresis by O'Farrell in 1975 [6]. This technique allows the separation of proteins according to two of their physicochemical properties, leading to a gel in two dimensions representing a map of the proteome of the studied tissue.

Before any mass spectrometry analysis, sample need to be simplified. Sample simplification could be performed at two levels, the protein or the peptide level. Several factors influence the choice, nevertheless, it could be imposed by the equipment available in the laboratory. Two notions are in balance while preparing a sample for mass spectrometry analysis. On one hand, in a mass spectrometry point of view, the simpler the better. In other words, more a sample in simplified, more the information obtained is accurate. On the other hand, in order to obtain simplified samples, it may be required to perform several steps of simplification or purification. It is however well known that each step of simplification/purification generates losses in the sample, and thus loss of information. This is an important notion which need to be taken into consideration while designing the sample preparation strategy.

When the separation is performed at the protein level, three options are available (i) mono-dimensional electrophoresis, (ii) bi-dimensional electrophoresis and (iii) liquid phase. With electrophoresis methods, subsequently to the separation, in order to visualize the protein, the entrapped proteins are stained. Commercially available methods exist, Coomassie brilliant blue, Silver Nitrate or fluorescent molecules, to name only a few. In mono-dimensional gels, the proteins are visible as bands. The band(s) of interests, or depending on the procedure all the bands, are cut, the entrapped proteins are digested (see Box 2.1.) and then analyzed. The concern with

Table 2.1 Endopeptidase of enzyme frequently used in study involving mass spectrometry and their characteristics

	Cleavage site	Optimum pH	Specific conditions
Arg-C	C-terminal R	8.5	
Asp-N	N-terminal D and C	6–8.5	
Chymotrypsin	C-terminal Y, F, W, and L. Secondary C-terminal M, I, S, T, V, H, G, and A.	7.8	Stabilized and activated by Ca^{2+}
Glu-C	C-terminal E and D*	3.5–9.5 Maximum 4-7.8	*Cleavage after D depends on buffer, (phosphate buffer pH 7.8)
Lys-C	C-terminal K	8.5	
Lys-N	N-terminal K	9.5	
Pepsin	C-terminal F, L and E	2	Inactivated by pH > 6
Proline-endopeptidase	C-terminal P and A	7.5	
Trypsin	C-terminal K and R	7–9	

such method is that each band of the gel may contain several proteins and the complexity of the sample remains elevated. Thus, the analysis by mass spectrometry could lead to a lack or a loss of information. An additional separation step could be performed on the peptides with liquid chromatography coupled to mass spectrometry. Protein could also be separated in two dimensions (2D-PAGE). With this technique, first proteins are separated according to their isoelectric point (p*I*), which correspond to the pH where the sum of the charges of the protein is equal to zero. Second, proteins are separated by molecular weight [7]. Both separations combined provide a better separation and a better resolution with spots of proteins containing one or just a few proteins. As for the band from mono-dimensional gels, the spots can then be digested and analyzed by mass spectrometry. The simplification of proteins in liquid phase or liquid chromatography uses different chromatographic methods such as ion exchange, utilizing the charge of the proteins, reverse phase, for a separation depending on the hydrophobicity of the protein or size exclusion chromatography. Some strategies employ a combination of methods in a single column for a better separation. It is as well possible to multiply the steps, performing different chromatography one after another.

When the simplification is made by two-dimensional gel, proteins are presented in the gel as spots containing one, or only a few proteins. Then, after enzymatic cleavage of the entrapped proteins, the mass spectrum contains only peptides from the mixture of a limited amount of proteins. As showed in Table 2.1, enzymes are specific to amino acids, therefore a protein, once digested provides a pattern of peptides. This is true as long as the experimental conditions stay the same. The measurement, by mass spectrometry, of a single protein digested then provides a mass spectrum called: *Peptide Mass Fingerprinting* or *PMF*. The notion of fingerprint implies a link between the mass spectrum and the identity of the protein, as a fin-

gerprint would be linked to the identity of a person. As shown in Fig. 2.2a a protein is identified by its sequence. The cleavage of the protein by an enzyme leads to the multiplication of the sequence information, without changing the sequence. Indeed, each peptide contains a part of the sequence of the original protein. Taken together, the peptide information covers the whole sequence of the protein. However, after digest, some peptides may still be too large or too small to be measured in the selected mass range, then some information will still be missing. Nonetheless, a large set of the sequence information could be retrieved in a mass spectrum. The set of identified sequences represents a percentage of the entire protein called *sequence coverage* (See Fig. 2.2).

The other option is to start with a sample of proteins, digest all proteins and separate the peptides generated. Again, different strategies are available, two dimensional chromatography (often ion exchange followed by reverse phase chromatography). Usually the chromatography method performed directly before mass spectrometry is reverse phase because the eluted fractions are in solvents that are compatible with ionization methods. The first dimension of separation could be replaced by a gel-based method, separating peptide according to their isoelectric point. The final set of data for such strategy correspond to the detection of peptides with similar properties gathered in mass spectra. However, as there is no information regarding the original protein (e.g., pI or molecular weight), the simple mass of peptides is not enough for the identification of the proteins. Indeed, the mass of a peptide is only a part of the information. Then, an additional measurement is necessary, such as the determination of the sequence of the peptide. This method is based on the fragmentation of the peptides. For the different analyzer found in the mass spectrometers, the technology differs but the principle stay similar, following their mass measurement, the peptides are broken or dissociated. The fragments produced are then measured. This technique is called tandem mass spectrometry or MS/MS. Some studies require more than one fragmentation, in such cases it is possible to perform MS^n (n corresponding to the number of dissociation performed).

In a peptide, the weakest bond is the backbone, indeed, while the side chains are linked to the backbone by a carbon-carbon bond, the peptide bond is an amide, which allows easier dissociation/fragmentation. Then, as shown in Fig. 2.3a, the dissociation may appear at different place around the peptide bond, before the carbonyl group, between the carbonyl and the amine and after the amine function. In order to label the fragments produced, a nomenclature has been determined [8, 9]. In order to name the fragments this nomenclature takes into consideration the place the charge is retained. If the charge is on the fragment containing the N-terminal portion of the peptide, the ions are called a, b or c, when dissociated before, within or after the amide respectively. If the charge is on the C-terminal portion, the ions are labelled x, y and z. The position and the frequency of the dissociation depend on several factors such as the amino acid sequence and the dissociation energy applied. However, the bond between the carbonyl and the amine function is more fragile and thus the dissociation is more frequent, producing mainly *b* and *y* ions. Figure 2.3b presents a theoretical peptide with the following sequence P-E-P-T-I-D-E. The fragmentation of such a seven amino acids peptide, generate in theory six

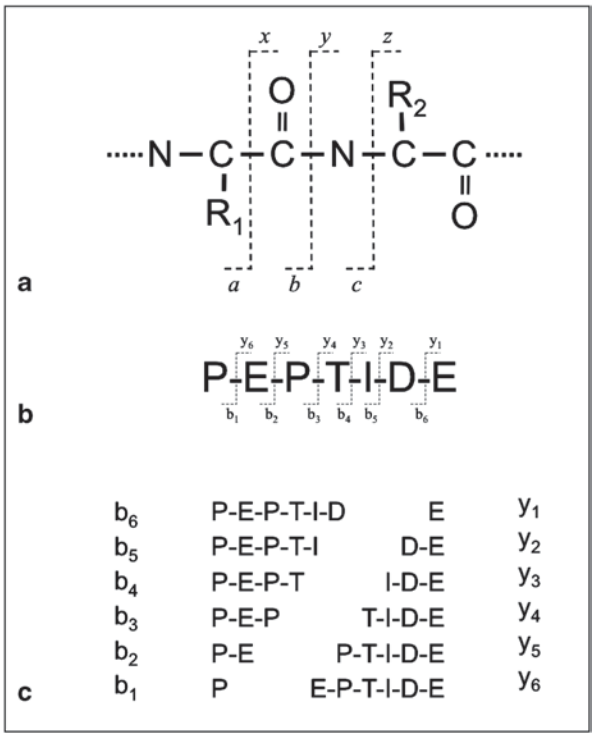

Fig. 2.3 Labeling of the fragmentation of peptides around the peptide bond. **a** A dissociation between the carbon α1 and the carboxyl group produces a fragments (*charge on the N-terminus*) and x fragments (*charge on the C-terminus*). A dissociation between the carboxyl group and the amino group produces b fragments (*charge on the N-terminus*) and y fragments (*charge on the C-terminus*). A dissociation between the amino group and the carbon α2 produces c fragments (*charge on the N-terminus*) and z fragments (*charge on the C-terminus*). **b** The numbering of the fragment starts at the amino acid residue with the charge, C-terminus for y ions and N-terminus for b ions. **c** Example of fragments formed form the peptide with the sequence P-P-E-T-I-D-E with their nomenclature

b ions and six *y* ions (Fig. 2.3c). Thereafter, each fragment can also be dissociated, producing additional ions. Finally, each peptide has a specific fragmentation pattern dictated by the properties of the peptide. In spite of the common basic structure of the peptide, called backbone, the side chains of each amino acid have an effect on the backbone and its dissociation. The electro-negativity of the atoms constituting both side chains surrounding the peptide bound either reinforce the bound or make it more fragile. The number of fragments produced reflects the effects of the amino acids on the dissociation, in other words, amino acids influence the rate of fragmentation. Then, more dissociation leads to more molecules produced. The specificity of the fragmentation, depending on the peptide sequence, leads to the *Peptide Fragmentation Fingerprinting* or *PFF*, in opposition to the PMF. There is many kinds of dissociation methods, among them we can cite three that are used in mass spectrom-

etry approaches presented in this work. Collision induced dissociation (CID), which depends on the collision of the analytes with an inert gas to break down the analytes and then measure the fragment produces. Electron-capture dissociation [10] (ECD) and electron-transfer dissociation [11] (ETD), are similar and the results of the fragmentation are comparable, the only difference is the feature and the instruments used for this method. These dissociation methods perform the destabilization of the electronic cloud of the analytes inducing their dissociation. Finally, metastable atom-activated dissociation (MAD) is as well based on the destabilization of the electronic cloud of the analyte [12].

At this stage, with mass spectrometry analysis, we are able to provide a sequence information for a part of any protein. But we still need to know the sequence of the original protein to compare the information obtained by mass spectrometry with the sequence of known proteins. That is where the sequencing of the genome and at the medical level, the sequencing of the human genome, helped greatly. While decades ago, known sequences of proteins and/or genes could be presented as print-outs or subsequently saved in disks [13], now the amount of information is too high for such support. Today, thanks to the worldwide web, the knowledge of the genome and the proteins sequence is accessible by everyone. The main databases are, in the alphabetical order, the database hosted by the *National Center for Biotechnology Information* commonly known as NCBI and the database hosted by UniProt. This latter is composed of two sections; the first section, called TrEMBL, contains automatically processed and annotated data of the coding sequences from the nucleotide sequences databases. The second section, named SwissProt, contains manually reviewed information, which make them less exhaustive but of a higher quality. Using those free of charge, available to anyone, databases, it is now easy to compare a peptide sequence obtained by mass spectrometry to the more than forty millions amino acids sequences stored in those databases. In summary, starting from any kind of tissues and using mass spectrometry it is now possible to identify the proteins contained in the biological sample. In addition, the information enclosed in UniProt KnowledgeBase (UniProtKB) it is easier to link a protein name to its structure, function, tissue and sub-cellular location and so on, as the pages on this website contain all those information and a lot more (for more details visit http://www.uniprot.org/). This effort of sharing the knowledge requires, still, the participation of researchers who are constantly increasing the knowledge we have about biology and biological elements. Then, in spite of the great advantage that was the human genome project, additional information is necessary to decipher the complex systems that represent all living organisms.

2.1 Protein Quantification

The Holy Grail of proteomics studies would be that one protein is the undoubted marker of a particular disease. Such marker could for example be found in blood, cerebrospinal fluid (CSF) or urine for diagnostic or it could be a cellular marker

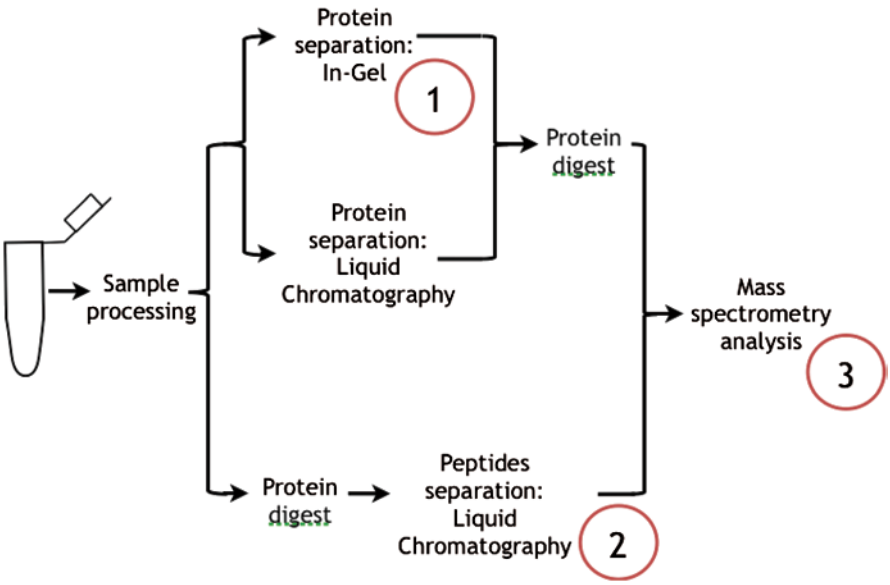

Fig. 2.4 The sample preparation starts with the processing (e.g., depletion, precipitation, subproteome purification, etc). The analysis may then be performed on proteins (top of the panel). The protein separation could be either in gel, which allows a protein quantification (**a**), or by chromatography. The proteins are then digested and analyzed. The analysis may also start with the digestion of the proteins (bottom of the panel). The peptides generated could then be separated by liquid chromatography, which could be used as part of the quantification (**b**), Finally, the peptides are measured by mass spectrometry, which is the last step where the quantification could be performed for the proteins of the sample (**c**)

which could be used as a target for therapy. In such cases the protein is only present or absent when the pathology occurs. But this is rare and generally studies show high similarities between the protein expression from a control samples and the pathology samples. It is then necessary to measure the changes in the level of expression of each protein. Any significant changes would help to provide the identity of the proteins for the prognosis, the diagnosis or the therapy. Once again, the complexity of the protein samples make them impossible to analyze in a single step process, so the protein quantification may be performed at different stages of the analysis. Figure 2.4 presents the different possibilities to quantify the proteins in a complex protein sample.

Taking the example of human blood samples, this tissue is complex at many levels first it is a liquid made of the plasma, in which float cells. The plasma is mainly made of water used as transporting solution for salt, proteins, carbohydrate, lipids and other chemicals. Focusing only on proteins the most abundant protein is Serum Albumin, which represent more than 57% of the total proteins. The second class of proteins highly abundant in the plasma is the class of Globulines that are classified in families: α_1 and α_2 Globulines, β Globulines and the most abundant family γ Golublines, also known as immunoglobulines or antibodies. Finally, the plasma is constituted of coagulation factors that are numerous and highly expressed.

In conclusion, analyzing a gel of plasma proteins would reveal the proteins previously cited and the others remain hardly detectable. If the plasma is analyzed using a liquid chromatography method the outcome would be the identification of peptides from the most abundant proteins and few peptides from other less abundant proteins. This limitation is mainly due to technical restrictions. In-gel studies are constrained by the total amount of proteins that could be loaded in the gel, as for liquid chromatography, which is limited by the column capacity. Then, an additional processing of the sample could be performed. During the preparation steps, the added step, called depletion, consists in removing the most abundant proteins from the plasma leaving the rest of the plasma proteome more visible for any proteomics techniques. Indeed, in gels, the absence of the most abundant proteins will allow the spots of the other proteins to be revealed. As well, in liquid chromatography separations, as the peptides from abundant proteins are absent, the peptides from the rest of the proteome are more visible. Thus, in both cases, more proteins can be detected and identified, which leads to more information for the plasma samples. The depletion of plasma sample is becoming more and more common within the proteomics studies. It has been described as presenting drawbacks, because depleted proteins may form complexes with the other proteins in the plasma. However, it is worth measuring the ratio benefits/risks of such additional step and depletion could lead to a small loss in valuable information, while unprocessed samples would allow the identification of the abundant proteins and very little, if not any, from the rest of the proteome.

2.1.1 Label Free Quantification

Mass spectrometry is not a quantitative technique *per se*. In a mass spectrum of peptides, the intensity of each peaks does not reflect their respective quantity but rather their intensity of ionization. It would be impossible to quantify a peptide using only its signal intensity in one spectrum. Nonetheless, one method uses the raw mass spectrometry data to relatively quantify the peptides. This technique combines the information from liquid chromatography and mass spectrometry to quantify the peptides in a sample [14–16]. During a study using liquid chromatography coupled to an electrospray ion source, the elution of a peptide has an aspect close to a bell-shape. Focusing on the peak of a single peptide, the intensity of the peak increases along with the gradient of elution (Fig. 2.5). This increase corresponds to the augmentation of the number of ions detected in the mass spectrometer. Then, when comparing the intensity and the area of the peak between two samples, it is possible to estimate the relative expression of a peptide between samples. Also, in general, the intensity will increase quickly to reach a maximum and then decrease with a steep slope. Another technique takes into consideration the fact that the abundance of a peptide is reflected by its number of fragmentation spectra [17]. The goal is then to count the number of fragmentation spectra of each peptide and compare the numbers between different conditions, in order to estimate the increase or the decrease of the peptide expression (see the review from Bantscheff et al. [18] for

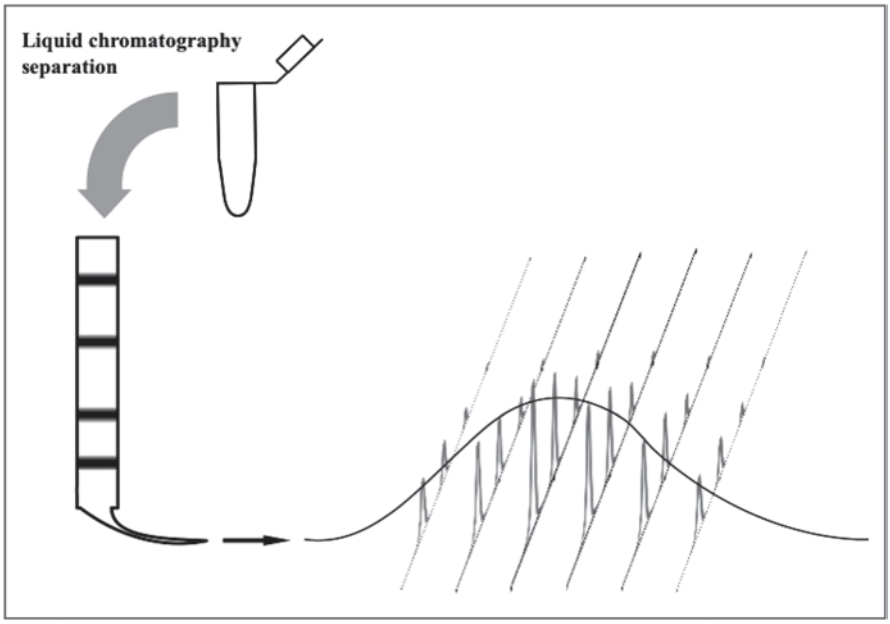

Fig. 2.5 In a sample separated by liquid chromatography the peptides or proteins are separated and during the elution the concentration of peptide/protein increase up to a maximum and then decrease. The right part of the figure illustrates the profile of elution of a peptide with the succession of mass spectra with changes of intensity along the elution

a critical review of quantitative proteomics by mass spectrometry). Another option is to introduce in each mass spectrum a known quantity of peptides or chemical and compare the expression of the peptide from the samples to this artificially introduced compound called spike. The spike is usually added in the solution for the liquid chromatography so it is found in all the spectra of the sample.

Label free quantification methods are relative quantifications or sometime called semi-quantitative methods. Indeed, with such method, no absolute quantities are determined. Once again, the complexity of the sample could interfere with the measurement. With the peak intensity evaluation, in a complex spectrum, overlaps could occur leading to misinterpretations of the results. For the spectral count technique the same drawbacks apply, the number of fragmentation spectra for a peptide may vary when the complexity of the sample changes. Label free quantification methods also have the disadvantage that several repeats of the measurement are required for a more statistical evaluation of the samples [19]. On the other hand, the advantage of label free methods is the low cost. There is no need to purchase expensive reagents in order to perform such experiments.

2.1.2 Stable Isotope Labeling

Using the powerful resolution of mass spectrometry, the enrichment in stable isotopes is a useful tool to quantify peptides in a mixture. The rational is that stable isotopes do not change the properties of a peptide, it simply change its mass. So utilizing any separation technique, either at the peptide level or at the protein level, a non-labeled peptide has the same characteristics compared to the stable isotope enriched one. Finally, for mass spectrometry analysis, both peptides are measured in the same mass spectrum. Then, their intensity could be compared allowing a relative quantification. Stable isotopes naturally exist and can be measured in the structure of peptides. However, they are so little in quantities that some are rarely seen. This is the reason why it is called enrichment. Mostly, techniques using stable isotopes are qualified of isotope labeling (See Table 2.2.).

The labeling of peptides or proteins with stable isotope may be realized at different stage of the sample preparation. For instance, starting from the original and complex sample a chemical labeling can be performed on the proteins. Thus, the separation method could be either with polyacrylamide gel or the proteins could be digested. The peptides with the labels then undergo a phase of separation. It is important to remember that in case of a method based on a chemical modification of the proteins, the protein properties are altered. A gel based separation of the proteins and especially 2D-PAGE, uses the physicochemical properties of the proteins for the separation. The labels change the properties of the proteins leading to different profiles of separation. Nevertheless, as the modification is the same in all analyzed samples, the comparison is still possible. The chemical modifications of proteins leads to an additional mass, shifting the molecular weight of the studied proteins. In addition, the modification targets amino acids participating to the isoelectric point of the protein and thus induces changes in the isoelectric point of the proteins. Considering that the modifications are performed on all samples, the only difference being the atomic composition of the label, the mass difference between stable isotope enriched and non-enriched protein corresponds to the stable isotopes that are not interfering in the separation. The advantage with stable isotope labeled proteins is that in a single gel all the proteins from different conditions are co-migrated. And in one spot or band the proteins contain different labels, then during the subsequent steps of the process, i.e., protein digest and peptide extraction, all the conditions undergo the same treatment, which finally leads to the mass measurement of all the labeled peptides at the same time. Finally, the mass difference between the isotopes

Table 2.2 List of the isotope frequently used for mass spectrometry quantification

Element	Natural/isotope	Mass isotopes	% Abundance isotope
Hydrogen	$^{1}H/^{2}H$	1.008/2.014	0.012
Carbon	$^{12}C/^{13}C$	12.000/13.003	1.07
Nitrogen	$^{14}N/^{15}N$	14.003/15.000	0.368
Oxygen	$^{16}O/^{18}O$	15.995/17.999	0.205

is visible in a single mass spectrum. The doublets, triplets or even *multiplets* of peaks massive are detectable either at the MS or at the MS/MS level depending on the type of modification added to the sample.

The first use of stable isotope in mass spectrometry was reported in 1971 [20]. The methodology presented in this study proposes an alternative method to an already existing measurement of bile production. Using quantification involving ^{2}H, a stable isotope of hydrogen, the research group presenting this study was able to replace utilization of ^{3}H, the radioactive equivalent, providing similar characteristics and reducing the hassles and hazard of radioactivity. The use of stable isotopes increased in the following years showing a growing interest for quantification by mass spectrometry. At present, stable isotopes are used in several techniques. In the era of proteomics, for example, it is possible to use stable isotopes to quantify protein expression.

In summary, there are three main stable isotope labeling methodology. First, the amino acids based labeling. For this method, mostly used in cell culture and called stable isotope labeling by amino acids in cell culture (SILAC), amino acids are replaced in the culture media by stable isotope labeled amino acids. This leads to the introduction of the stable isotope into the protein sequence. Then, during the mass spectrometry analysis the peptides containing these labeled amino acids have the same characteristics than the non-labeled ones but are heavier and comparable in the mass spectra. Second, the peptides are chemically modified, either the labeling is performed with stable isotope enriched biotin, allowing the purification of labeled peptide or the modification is performed with an inert label. These techniques are respectively called Isotope Coded Affinity Tag (ICAT) [21] and isobaric Tag for Relative and Absolute Quantification (iTRAQ) [22], tandem mass tag (TMT) [23] or isotope coded protein label (ICPL) [24]. ICAT technique targets and modifies Cysteine residues and then, as the chemical modification consists in the addition of a biotin, modified peptides could be specifically purified using avidin or streptavidin (see box 2.3.). Finally, during mass spectrometry analysis, the relative quantification is performed at the peptide mass spectrum level. iTRAQ, TMT and ICPL methods are based on the modification of amino group of peptides and/or proteins, i.e., lysine and arginine residues as well as the amino group at the N-terminal end of the peptide. While ICPL allows a quantification in the mass spectra, the quantification with iTRAQ and TMT occur at the MS/MS level. For those latter, the labeling reagent is made of two parts, first part the label, also called “tag”, includes the stable isotopes for the quantification, the second part, called the balance group, and links the peptide and the tag. The balance group, contains stable isotopes as well, which leads to the same mass for the whole molecules. For example, if the labeling reagent contains in total four stable isotopes, one label could be formed by no stable isotopes in the tag and four in the balance group, another label could then have four isotopes in the tag and none in the balance group. This latter correspond to the heavy tag. All the combination in between are possible, which means that several conditions could be studied in one spectrum. In summary, the mass spectrum only presents one peak with a normal isotopic distribution for each peptide, but at the fragmentation level, the spectrum presents additional peaks that correspond to

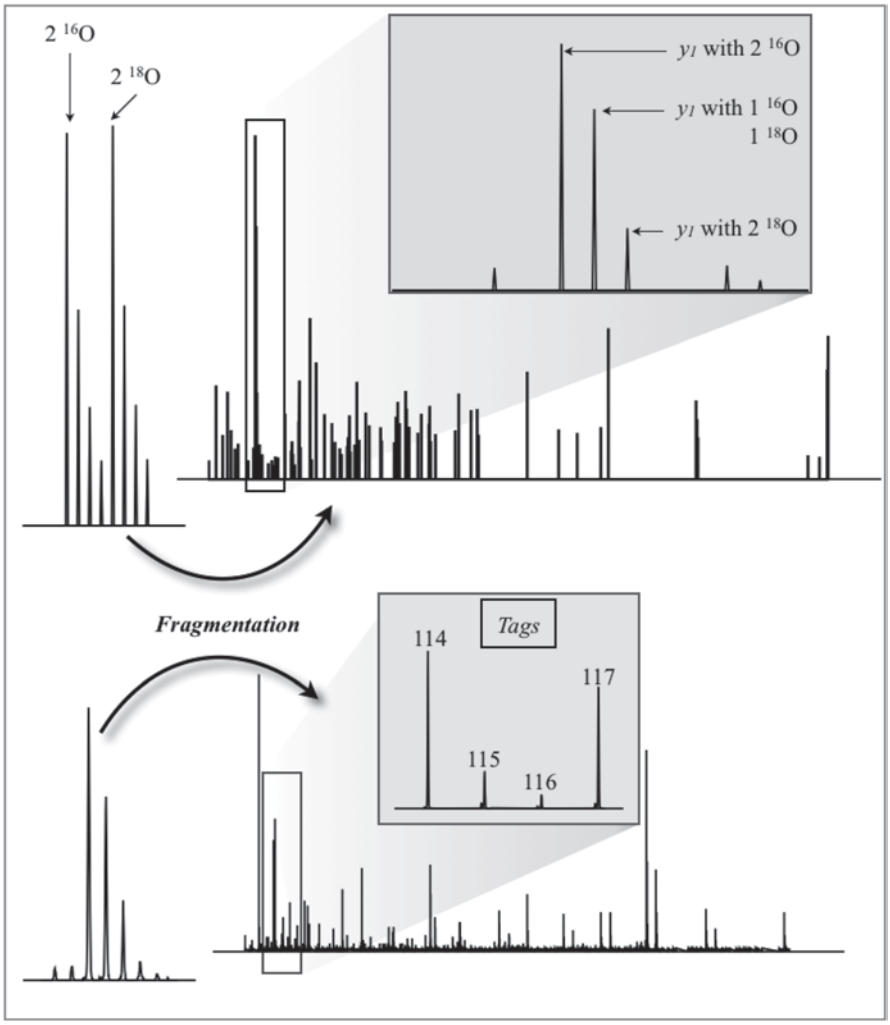

Fig. 2.6 The isotope labeling by $H_2{}^{18}O$ (*Top panel*) leads to the production of heavy peptides that could be partly detectable in the fragmentation spectra. However, generally the quantification is performed at the mass spectrum level. The iTRAQ technique (*bottom panel*), on the other hand, produces one peak in mass spectra and the quantification is possible using the tags released during the fragmentation

the tags with an intensity in correlation with the level of expression of the peptide in each condition (see Fig. 2.6). Two of the advantages of the tandem mass spectrometry-based quantification are, first it reduces the mass spectra complexity. Independently of the number of conditions studied each peptide is presented as one peak

in MS. Second, as the quantification is performed at the fragmentation level, the quantification is directly related to the peptide identity. The last level of labeling of the peptides form the samples takes place during the protein digestion. The cleavage of peptides, using enzymes, induces the breaking of covalent bonds and the addition of a molecule of water on the C-terminal carbon of the peptide. Then, by replacing H_2O, in the digestion media by $H_2{}^{18}O$, the water molecules added on the peptide will be heavier, increasing the mass of the peptides. Depending on the structure of the peptide, the addition of ^{18}O changes between one and two atoms (see Fig. 2.6). This must be taken into consideration during the data analysis. Few articles present recent developments for the quantification of proteins using ^{18}O [25, 26].
Another method of quantification using stable isotopes consists in introducing in the

Box 2.3. Interest of Biotin

The complex biotin-avidin or biotin-streptavidin is the strongest currently known with $Kd = 10^{-15}$. This interaction has been utilized in many methods such as protein and nucleic acid detection and purification.

Biotin is a small molecule that could be covalently liked to antibodies, nucleic acids, proteins, ligands, etc. Its size allow the unchanged interaction of the labeled molecule with its partner or ligand. One disadvantage is that biotin is naturally present is some tissues, which could cause unspecific interaction with (strept)avidin and interfere in the detection or the purification.

The interaction biotin-(strept)avidin requires harsh conditions to be removed. Then for example, the labeling with biotin, called biotination, could be performed with a spacer arm containing a disulfide bond in order to cleave the label with reducing condition in order to avoid drastic conditions.

biological sample, during or after digestion, a known amount of labeled peptides, this method is called spiking. In other words, a synthetic peptide, with a known amino acid sequence, labeled with one or several stable isotope labeled amino acids is added in the sample of interest. Knowing the concentration of the synthetic peptides it is then possible to evaluate the level of expression of the endogenous peptides, comparing the intensity of the endogenous peptides with the intensity of the synthetic peptides [27]. Such exogenous peptides allow as well the quantification of peptides of interest using a strategy called selected or multiple reaction monitoring (SRM and MRM respectively).

SRM or MRM are methods of quantification by mass spectrometry. These techniques employ generally triple quadrupole mass spectrometers. For this methodology, the first quadrupole selects one specific mass, the second quadrupole works as the dissociation chamber and the third quadrupole select one mass among the fragments. The first mass selected could correspond to different peptides, especially when those latter are multiply charged. This is the reason why the fragmentation selected in the third quadrupole, also called transition, must be specific to the peptide

of interest. More than one transition can be selected in the third quadrupole in order to increase the accuracy of the method. It is the combination of the selected precursor and the selected transitions that make the specificity of the technique. With such narrow range of compounds, the detector at the end of the analyzer, becomes an ion counter allowing the measurement of ion quantities. Box 2.4. indicates the necessary steps to perform the development of a MRM method.

Box 2.4 MRM Quantification

MRM is a method of quantification based on the *filtering* of the targeted molecules. It could be applied on peptides or non-proteinic molecules. The current description present few crucial steps for the quantification of peptides.

MRM experiments, in general, require a triple quadrupole mass spectrometer. In the first quadrupole (Q1) the target peptide is selected and accelerated toward the second quadrupole (Q2) that plays the role of collision chamber, for the collision induced dissociation (CID). Finally, in the third quadrupole (Q3) one or a few fragment ions, also called transition ions, are selected and measured. Few steps are crucial in the development of an MRM method for quantification and some are summarized below:

- The first step in the development could be an MS and MS/MS analysis of the sample to determine the potential targeted precursors.
- The selection of the precursor is important, the selected peptides must be specific to one protein (unique peptides) and must possess a high ionization rate. This step could be performed with the support of a software (see www.ms-utils.org)
- The transition ions must as well be specific to the peptide
- The integration of an internal standard allows a more accurate quantification
- The quantification requires the production of a calibration curve allowing the determination of the limit of detection (LOD) and the limit of quantification (LOQ), two essential values in quantification methods.

Mass spectrometry-based quantification method is a powerful technique that may allow the quantification of biomarkers as low as femtomol of compounds. However, it necessitates development that could be lengthy but needed to avoid some of the caveats of the procedure.

2.2 Amino Acids Analysis and Small Cell Component Analysis by Mass Spectrometry

The ability to measure peptides with mass spectrometry was a revolution for the technique but from the start mass spectrometry was used to measure small compounds, such as amino acids or drugs and even very small peptides. For this kind of

compounds, gas chromatography coupled with mass spectrometry was a powerful method. It allowed the mass measurement and the identification of the compounds using tandem mass spectrometry. Also SRM and MRM (see part 2.2.) can be applied for such compounds. In order to analyze smaller compounds, in body fluids for example, they must be measurable by mass spectrometry. However, some compounds in such tissue ionize easily and are the first to be observed, this may lead to the loss of the ion of interest. This phenomenon is called ion suppression. It is then essential to eliminate the protein and the salts that could be detrimental for the analysis. In most studies, the samples need a few clean up steps before the mass measurement. In such studies one field of research seek the identification and more specifically the quantification of metabolites in biological samples. This field, called metabolomics, looks for the down products of metabolic pathways. The outcome of this kind of studies are, among other, the determination of the degradation of a drug or the presentation of predominant pathways in cells or tissues. The main separation methods of this field are gas chromatography, liquid chromatography and capillary electrophoresis.

2.3 Lipid Analysis by Mass Spectrometry

In biology, lipids constitute another category of compounds with interest. Indeed, those compounds are involved in cell structure, signaling and energy storage. The structure of lipids is based on few families, the main ones in biology are: fatty acids, glycerolipids, sphingolipids and sterols. The fatty acids are the simplest structures and are constituted of two parts, (i) the lipophilic part, formed of a carbon chain with variable number of carbons. Those chains may be saturated (meaning no double bonds in the chain) or they may include double bonds, in which case they are called unsaturated. (ii) The second part, at the end of the carbon chain, fatty acids present an acid group, which is polar. Glycerolipids are a combination of a glycerol, to which are attached different substituents such as fatty acids, phosphate and functionalized phosphate. Sphingolipids have a sphingosine basis substituted with fatty acids and non-lipid constituents, such as Phosphocholine, phosphoserine or carbohydrates, to site the main ones in mammalians. Sterols are steroid alcohol including for example the most famous, cholesterol.

One of the difficulties of studying lipids is due to their characteristics. They are highly lipophilic, which makes them difficult to analyze. But their properties allow for example, the use of gas chromatography. The rational of the technique is to measure the ability of compounds to vaporize. In gas chromatography, the compounds interact with the stationary phase, liquid or polymer-based and are eluted by the flow of inert gas applied in the column. Lipids do not vaporize easily but after derivatization, producing for example methyl esters, they become volatile and can be separated and analyzed by gas chromatography. For gas chromatography coupled to mass spectrometry it is recommended to use derivative agents that contain nitrogen atom (e.g., pyrrolidides). However, gas chromatography is one method of

separation and lipid samples could be complex and require more than one separation method. Then, prior to gas chromatography lipids samples may be fractionated by reverse phase. With this latter, fatty acids are separated from lipids with more polar functions. On the other hand, lipids and more specifically unsaturated lipids undergo a differential elution using silver ions as stationary phase. This method can be applied with thin layer chromatography (TLC) and column chromatography and high performance liquid chromatography (HPLC). Such techniques add a new dimension for the separation and the simplification of lipid samples for subsequent GC-MS. Then, it is easy to understand that, depending on the type of sample and the information sought, it is possible to separate lipids using for example HPLC or Ag^+ chromatography, followed by gas chromatography to finally identify or characterize the lipids in a sample by mass spectrometry.

Mass spectrometry provides the mass value of the compound contained in each fraction of the sample, giving an indication of the lipids in the sample. In addition, tandem mass spectrometry, in particular collision induced dissociation (CID), produces fragments that are specific to individual molecular species. Therefore, in a complex samples containing lipids, all constituents are identified and characterized. For more details in regard to lipids chemistry and research the AOCS lipid library website [28] provides many information, protocols and advices to conduct a research study on lipid samples.

2.4 Mass Spectrometry of Nucleotides

DNA alterations is a phenomenon that can happen spontaneously in vivo, leading to modified bases or adducts. It also occurs when DNA is exposed to certain chemicals. Such changes can significantly affect the biology of the cells leading, for example, to tumors and carcinogenesis. Several studies showed a correlation between DNA adducts and cancers [29, 30, 31]. The relationship between DNA adducts and cell mutations depends on several factors such as the structure of the modification, the ability of DNA repair proteins to recognize the lesion and the position of the modification in the gene. A DNA lesion leads to mutations if the modification in the gene affects the structure and/or function of the resultant protein. This shows that the study of DNA changes is essential to determine the type and the position of DNA adducts. Taking the example of a tissue, such study could also reveal the amount or the concentration of modifications.

In a mass spectrometry-based analysis of adducts, prior to the mass measurement, the DNA contained in the samples can either be separated by gas chromatography or liquid chromatography. The first method used was gas chromatography, as it was the first technique to be coupled to mass spectrometry. But the last decades the field witnessed an increased use of liquid chromatography as the coupling to mass spectrometry is now possible, thanks to electrospray ionization (see Chap. 1). On the other hand, electrospray ionization is the predominant ionization method for liquid chromatography coupled to mass spectrometry analysis of DNA, but APCI

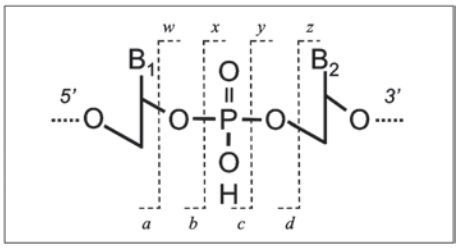

Fig. 2.7 The nomenclature for the nucleotide fragmentation is centered on the phosphate group. The fragments before the phosphate group are called *a* (*for the 5′ fragments*) and *w* (*for 3′ fragments*). The fragments between the oxygen on 5′ and the phosphor are called *b* (*for the 5′ fragments*) and *x* (*for 3′ fragments*). The fragments between the phosphor and the oxygen on the 3′ are called *c* (*for the 5′ fragments*) and *y* (*for 3′ fragments*). Finally, after the phosphate function the fragments are called *d* (*for the 5′ fragments*) and *z* (*for 3′ fragments*)

could be a more sensitive ionization technique [32]. The review by Tretyakova et al. [33] provides a thorough overview of the methods available to study DNA adducts and provides examples of studies performed the last years and describes as well the advances, in terms of research, that mass spectrometry had provided to the field of DNA adducts.

Among the advantages of mass spectrometry for DNA adduct analysis is first the determination of the mass of adducts. As any modification of the bases in a DNA sequence would necessarily be visible on the mass of the nucleotide as well as on fragments from the DNA. Second, it is possible to determine the position of adduct or adducts in DNA when sequencing the gene of interest. And third, methods have been develop in order to quantify the occurrence of adducts in a tissue. Mass spectrometry quantification of nucleotides, as for peptide quantification, uses stable isotopes such as ^{13}C and ^{15}N. The analysis of a DNA sample can be performed using different methods, the DNA sample is first hydrolyzed in order to obtain free nucleosides. Nucleosides are then either measured by mass spectrometry or separated by HPLC to acquire a first structural information on the adducts. The mass spectrometry measurement must be accurate, to be able with a single mass to determine a theoretical structure of the adduct. Finally, tandem mass spectrometry or MS^n allows the validation of the theoretical structure and identify specifically the adduct. Tandem mass spectrometry is also useful in terms of DNA sequencing. As mentioned earlier, the position of the modified nucleosides is important as well. It is thus essential to identify the position of the modification in order to determine where the mutation occurs and how it affects the related gene or protein. Among the first research group to sequence DNA Cerny et al. [34], in 1987 used fast atom bombardment (FAB) coupled to tandem mass spectrometry to sequence a small nucleotide constituted of six bases. Since then, techniques and instruments evolved leading to a predominant use of electrospray to study DNA adducts. In addition, although it is not the first application of mass spectrometry and other technologies are far more advanced to perform DNA sequencing, tandem mass spectrometry could

be used to sequence unmodified nucleotides as well. In order to determine the sequence of genes, as for the peptides fragmentation, nucleotides fragmentation has a nomenclature (see Fig. 2.7). However, in the review from Tretyakova et al. the authors claim that nucleotide analysis by mass spectrometry still suffers a lack of software dedicated to help in the processing of mass data produced by nucleotides studies.

References

1. Hua L, Low TY, Sze SK. Microwave-assisted specific chemical digestion for rapid protein identification. Proteomics. 2006;6(2):586–91.
2. Anderson L, Anderson NG. High resolution two-dimensional electrophoresis of human plasma proteins. Proc Natl Acad Sci U S A. 1977;74(12):5421–5.
3. Anderson NL, Anderson NG. The human plasma proteome: history, character, and diagnostic prospects. Mol Cell Proteomics. 2002;1(11):845–67.
4. Beck M, Schmidt A, Malmstroem J, Claassen M, Ori A, Szymborska A, Herzog F, Rinner O, Ellenberg J, Aebersold R. The quantitative proteome of a human cell line. Mol Syst Biol. 2011;8(7):549.
5. Pottiez G, Deracinois B, Duban-Deweer S, Cecchelli R, Fenart L, Karamanos Y, Flahaut C. A large-scale electrophoresis- and chromatography-based determination of gene expression profiles in bovine brain capillary endothelial cells after the re-induction of blood-brain barrier properties. Proteome Sci. 2010;15(8):57.
6. O'Farrell PH. High resolution two-dimensional electrophoresis of proteins. J Biol Chem. 1975;250(10):4007–21.
7. O'Farrell PH. High resolution two-dimensional electrophoresis of proteins. J Biol Chem. 1975;250(10):4007–21.
8. Roepstorff P, Fohlman J Proposal for a common nomenclature for sequence ions in mass spectra of peptides. Biomed Mass Spectrom. 1984;11(11):601.
9. Johnson RS, Martin SA, Biemann K, Stults JT, Watson JT. Novel fragmentation process of peptides by collision-induced decomposition in a tandem mass spectrometer: differentiation of leucine and isoleucine. Anal Chem. 1987;59(21):2621–5.
10. Zubarev RA, Kelleher NL, McLafferty FW. Electron capture dissociation of multiply charged protein cations. A nonergodic process. J Am Chem Soc. 1998;120(13):3265–66.
11. Syka JE, Coon JJ, Schroeder MJ, Shabanowitz J, Hunt DF. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. Proc Natl Acad Sci U S A. 2004;101(26):9528–33.
12. Cook SL, Collin OL, Jackson GP. Metastable atom-activated dissociation mass spectrometry: leucine/isoleucine differentiation and ring cleavage of proline residues. J Mass Spectrom. 2009;44(8):1211–23.
13. Karsch-Mizrachi I, Ouellette BF. The GenBank sequence database. Methods Biochem Anal. 2001;43:45–63.
14. Chelius D, Bondarenko PV. Quantitative profiling of proteins in complex mixtures using liquid chromatography and mass spectrometry. J Proteome Res. 2002;1(4):317–23.
15. Bondarenko PV, Chelius D, Shaler TA. Identification and relative quantitation of protein mixtures by enzymatic digestion followed by capillary reversed-phase liquid chromatography-tandem mass spectrometry. Anal Chem. 2002;74(18):4741–9.
16. Wang W, Zhou H, Lin H, Roy S, Shaler TA, Hill LR, Norton S, Kumar P, Anderle M, Becker CH. Quantification of proteins and metabolites by mass spectrometry without isotopic labeling or spiked standards. Anal Chem. 2003;75(18):4818–26.

17. Liu H, Sadygov RG, Yates JR. 3rd. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. Anal Chem. 2004;76(14):4193–201.
18. Bantscheff M, Schirle M, Sweetman G, Rick J, Kuster B. Quantitative mass spectrometry in proteomics: a critical review. Anal Bioanal Chem. 2007;389(4):1017–31.
19. Ryu S, Gallis B, Goo YA, Shaffer SA, Radulovic D, Goodlett DR. Comparison of a label-free quantitative proteomic method based on peptide ion current area to the isotope coded affinity tag method. Cancer Inform. 2008;6:243–55.
20. Klein PD, Haumann JR, Eisler WJ. Instrument design considerations and clinical applications of stable isotope analysis. Clin Chem. 1971;17(8):735–9.
21. Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nat Biotechnol. 1999;17(10):994–9.
22. Unwin RD, Pierce A, Watson RB, Sternberg DW, Whetton AD. Quantitative proteomic analysis using isobaric protein tags enables rapid comparison of changes in transcript and protein levels in transformed cells. Mol Cell Proteomics. 2005;4(7):924–35.
23. Dayon L, Hainard A, Licker V, Turck N, Kuhn K, Hochstrasser DF, Burkhard PR, Sanchez JC. Relative quantification of proteins in human cerebrospinal fluids by MS/MS using 6-plex isobaric tags. Anal Chem. 2008;80(8):2921–31.
24. Schmidt A, Kellermann J, Lottspeich F. A novel strategy for quantitative proteomics using isotope-coded protein labels. Proteomics. 2005;5(1):4–15.
25. Capelo JL, Carreira RJ, Fernandes L, Lodeiro C, Santos HM, Simal-Gandara J. Latest developments in sample treatment for 18O-isotopic labeling for proteomics mass spectrometry-based approaches: a critical review. Talanta. 2010;80(4):1476–8.
26. Bonzon-Kulichenko E, Pérez-Hernández D, Núñez E, Martínez-Acedo P, Navarro P, Trevisan-Herraz M, Ramos Mdel C, Sierra S, Martínez-Martínez S, Ruiz-Meana M, Miró-Casas E, García-Dorado D, Redondo JM, Burgos JS, Vázquez J. A robust method for quantitative high-throughput analysis of proteomes by 18O labeling. Mol Cell Proteomics. 2011;10(1):M110.003335.
27. Kirkpatrick DS, Gerber SA, Gygi SP. The absolute quantification strategy: a general procedure for the quantification of proteins and post-translational modifications. Methods. 2005;35(3):265–73.
28. lipidlibrary.aocs.org
29. Yang SF, Chang CW, Wei RJ, Shiue YL, Wang SN, Yeh YT. Involvement of DNA damage response pathways in hepatocellular carcinoma. Biomed Res Int. 2014;2014:153867. (Epub 2014 Apr 28).
30. Li LF, Chan RL, Lu L, Shen J, Zhang L, Wu WK, Wang L, Hu T, Li MX, Cho CH. Cigarette smoking and gastrointestinal diseases: The causal relationship and underlying molecular mechanisms (Review). Int J Mol Med. 2014;34(2):372–80.
31. Porru S, Pavanello S, Carta A, Arici C, Simeone C, Izzotti A, Mastrangelo G. Complex relationships between occupation, environment, DNA adducts, genetic polymorphisms and bladder cancer in a case-control study using a structural equation modeling. PLoS ONE. 2014;9(4):e94566.
32. Zhang F, Bartels MJ, Pottenger LH, Gollapudi BB, Schisler MR. Quantitation of lower levels of the DNA adduct of thymidylyl(3'-5')thymidine methyl phosphotriester by liquid chromatography/negative atmospheric pressure chemical ionization tandem mass spectrometry. Rapid Commun Mass Spectrom. 2007;21(6):1043–8.
33. Tretyakova N, Villalta PW, Kotapati S. Mass spectrometry of structurally modified DNA. Chem Rev. 2013;113(4):2395–436.
34. Cerny RL, Tomer KB, Gross ML, Grotjahn L. Fast atom bombardment combined with tandem mass spectrometry for determining structures of small oligonucleotides. Anal Biochem. 1987;165(1):175–82.

http://www.springer.com/978-3-319-13086-6

Mass Spectrometry: Developmental Approaches to Answer Biological Questions

Pottiez, G.
2015, VIII, 71 p. 12 illus., Softcover
ISBN: 978-3-319-13086-6