Contextualization

This thesis was developed in co-supervision between the Universidade Católica do Rio de Janeiro (PUC-Rio) and Université de Pau et des Pays de l'Adour (UPPA).

Between 2011 and 2012, part of the work was developed in the city of Pau in France, under the supervision of Professor Dirk Schaumlöffel in his laboratory, IPREM-LCABIE, at the aforementioned university. The studies with complexation and derivatization of peptides with model lanthanide metals using the DOTA chelating reagent, was carried out basically using techniques nano-HPLC (DIONEX) and ICP-MS (Agilent), both belonging to LCABIE. The identification of peptides cited was conducted by a MALDI TOF MS in a partner laboratory under the supervision of Professor Andreas Tholey, located in the city of Kiel in Germany.

After the period "sandwich", the work continued in Brazil, in LABSPECTRO and Bioanalytical laboratories under the supervision of Professor Tatiana D. Saint'Pierre and co-supervision of Dr. Rachel Ann Hauser-Davis. Before the focus of the study was the derivatization and labeling of biomolecules with lanthanide metals and their identification from mass spectrometry, here in Brazil the goal was no longer the study of pseudo-metal biomolecules. The continuity of the work done in France was not possible in Brazil because there was no access to the same techniques used in France. This was also the main reason for granting the scholarship, a novel approach to study for Brazil.

Back in Brazil, molecules that are naturally linked to metals, providing to use them as potential biomarkers of environmental contamination, in addition, the possible choice of certain metals likewise be used as biomarkers were explored. Moreover, in this second stage of doctoral research, real samples, acquired from producers of fish, was used for the purpose of exposure to metals and study of possible changes in bile and liver of the fish studied, especially related to metallothionein, biomolecules that are severely induced by the environmental contamination with metals. For the analysis of metalloproteins and metal fish

studied were used techniques such as UV-Vis spectrophotometry (Hamilton), SEC-HPLC (Shimadzu), ICP-MS (PerkinElmer), FT-IR (PerkinElmer), ESI-MS (Waters), MALDI-MS (Bruker), and SDS-PAGE (GE). With the exception of the molecular mass spectrometric techniques, ESI and MALDI, all analyzes were performed at PUC-Rio. These other techniques were carried out in partnership with the Universities of Campinas (UNICAMP) and of Espírito Santo (UFES), respectively.

1.
Introduction

1.1.
Proteomics and metallomics

The genomics field aims to determine the entire DNA sequence of a certain organism. This area has contributed greatly to the better understanding of life in the last few decades. A limitation of this field, however, is that DNA sequencing data do not provide enough information regarding protein expression in a specific tissue or cell.[1; 2]

In the post-genomic era, the emergence of proteomics,[3] described as the separation, identification, and characterization of proteins present in a biological sample, is directly related to the need to investigate the control of gene expression and its impact on cellular metabolism. In this regard, proteomic studies can generate important information, such as which proteins are expressed[4], their expression levels, when they are expressed, the presence of post-translational modifications, protein responses expressed by cells in different conditions or experimental treatments and molecular differences between strains of cells and gene interactions.

The term 'proteome' was first used by Wasinger *et al.* in 1995, [5] meaning the total protein complement expressed by a genome. The proteome of a cell or a body fluid is denominated as the population of complete set of proteins being at a certain time point and under determined conditions within biological matrices. In opposition to the genome, the proteome varies from cell to cell and is very dynamic, since proteins are constantly synthesized or degraded in the biological environment, and, thus the composition of the proteome is highly flexible, meaning that the presence of particular proteins, as well as the amount of a

specific protein that can be altered depending on the biological state of the organism, is also flexible.[6]

Proteome investigations represent a new way to understand metabolic processes and cell machinery. Proteins have many functions within the cell, including gene regulation, signal relaying within and between cells, and metabolic process regulation. The knowledge of the proteome composition, as well as of the in-depth characterization of isolated proteins and the identification of proteins with altered expression profiles has a great impact on a variety of applications in medical, pharmaceutical and environmental research, for example:

- Development of novel cell types (*e.g*. bacteria or eukaryotic cells) which can be used for the sustainable bioproduction of high quality chemicals;

- Identification of potential targets in cells for the development of novel drugs;

- Identification of novel biomarkers for the early detection of diseases and the development of strategies for their treatment;

- Identification of biomarkers for protection of environment.

Furthermore, the protein structure (amino acid sequence) encoded in the genome can be altered by the covalent attachment of other present chemical groups to diverse chemical functionalities in the amino acid sequence - posttranslational modifications, which further alter protein properties in the proteome, allowing for distinct regulation of biological processes within cells. Thus, the knowledge regarding the proteins present in a proteome (identification), the amounts of a particular protein within this mixture in different biological conditions, *e.g*. healthy and diseased states (relative or absolute quantification), and the knowledge on the presence of posttranslational modifications (characterization) is the basis for the understanding of the fundamental processes of life. However, the acquisition of this knowledge is interfered by divers factors: (i) the proteome is usually a very complex mixture of hundreds up to hundreds of thousands of proteins; (ii) protein expression is characterized by a very high dynamic range, ranging from only a few protein copies per cell to hundreds of thousands of copies; (iii) the analysis of regulatory proteins (which are frequently present in low abundance) alongside proteins present in high abundance is very challenging.

With the advancement of genomic and proteomic studies, the role of metals in the various functions of proteins and enzymes has been highlighted. To discover the behavior and functions of the complexes formed by metals with biomolecules, such as proteins, has been, increasingly, a challenge of interest for various fields of science. The study of set of biomolecules which include metalloenzymes and metalloproteins that bind or interact with metals in a cell, organism or tissue, is part of the relatively recent field named metallomics, [7; 8] and the set of metals present in a biological system, in their different ways, is defined as the metallome.[9;10] Metallomics is directly associated with genomics and proteomics, since the synthesis and metabolic functions of a wide range genes and proteins do not occur in the absence of metals,[7] and the identification and elucidation of the biochemical or physiological function of the metallome in a biological system are the research targets in this field. Thus, speciation analyses, the identification/quantification of one or more individual chemical species in a sample, for metallobiomolecules, can be referred to as metallomics, an analogy to the genomics and proteomics. In other words, metallomics can be considered as a subset of speciation analyses.[7; 8; 10;11]

1.2.
Protein and metalloprotein biomarkers applied to environmental biomonitoring studies

1.2.1.
Biomarkers

Biomarkers are biological changes that may be related to exposure to or the toxic effects of chemicals.[109] Biomarkers represent measures of biological changes that may indicate the presence of contaminants and provide a means of interpretation of environmental levels of pollutants, being of extreme importance in environmental biomonitoring studies.[110] This original definition was modified by Adams, specifically with aquatic organisms in mind. He included characteristics of organisms, populations, or communities that respond to changes in the environment in measurable ways.[111] Later, Depledge added behavioral

responses, latency and genetic diversity to this definition.[112] Based on these authors, three types of biomarkers were then proposed, in an attempt to classify organism responses as biomarkers of exposure, effect and susceptibility.[112] As more biomarkers have been characterized and identified, it has become apparent that this ternary definition has significant overlap, since some biomarkers can be used in each of these capacities. As a more recent review of the biomarker paradigm has been described, an effect resulting from stressor exposure may be defined as an early adaptive non-pathogenic event or as a more serious altered functional event, depending on the mechanism of action of the stressor, the toxicokinetics, Biomarkers of exposure and effect may, thus, often be combined into a single classification with susceptibility occurring along any stage.[113]

The main concept of the biomarker approach in the evaluation of adverse effects or stress is based on the hypothesis that the effects of these disturbances are typically expressed at lower levels of biological organization before they arise at higher levels, such as populations, communities or ecosystems.[111] These initial effects are observed primarily at the molecular level with the induction of cellular defense systems and may provide answers on organism adaptation after exposure to the contaminant. However, if these defense processes are flawed, harm can occur at higher levels, such as histological or tissue damage. If these processes are permanently affected or modified during vulnerable development periods, organism reproduction or even survival may be affected, leading to changes in the population and, possibly, at community organization.[114] Therefore, in biological processes, the effects at higher hierarchical levels are always preceded by earlier changes (Figure 1), providing signals recognized as biomarkers of effects on levels of posterior responses.[115]

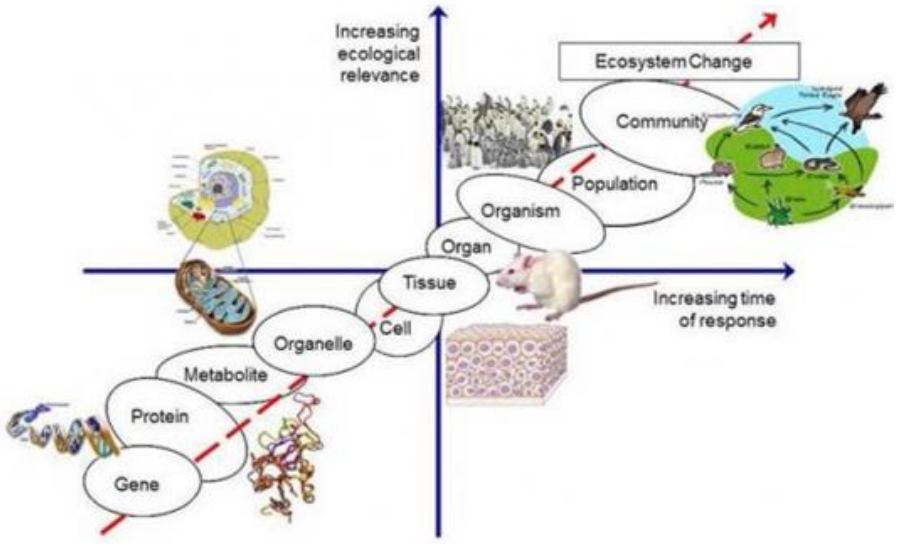

Figure 1. Scheme of the levels of responses of biological systems. From Bayne (1985).[115]

The study and the use of biomarkers often increases the possibility of identifying the cause of certain toxic effects and can also provide information regarding the bioavailability of pollutants and their potential environmental damage. However, these biomarkers can vary greatly in their specificity.[116] Such variations may reflect, for example, a biological response to external stimuli and / or contamination, resulting in different protein expressions or the redistribution of specific proteins within biological fluids or cells.[117; 118]

With the recent growth of genomics and proteomics-based new technologies the development and application of biomarkers has been the subject of intense research interest in this area. Proteomic techniques have allowed deeper investigations into environmental and biological issues, enabling the study of thousands of proteins. Until now, however, these techniques have benefited primarily well-characterized species such as humans, mice and yeast. However, in an environmental context, the species of interest is, many times, not well-characterized in proteomic or genomic aspects. Thus, new databases are of interest in proteomics, especially because these techniques are valuable in the identification of altered proteins after exposure to environmental pollutants, which may possibly come to be used as new and more comprehensive biomarkers,

since no prior knowledge regarding the biomarker relationship with pollutant toxic mechanisms is not necessary.[119]

1.2.2.
Biomolecules in fish used as biomarkers in an environmental context

Proteomics applied to environmental ecotoxicology studies and environmental monitoring has as one of its goals to discover and identify novel candidate biomarkers of response to environmental contaminants.[119; 120] These biomarkers may be a set of genes or proteins expressed simultaneously, which could potentially provide information in several diverse areas, such as: (i) in characterizing functions of genes and gene products with similar profiles or regulatory mechanisms in common,[121] (ii) which can be used to classify compounds with similar modes of action, analyzing their toxicological "fingerprints",[122] and (iii) indicating different stress levels by integrating general and specific markers in a single assay. Thus, the use of proteomic techniques is becoming increasingly useful in the discovery of such biomarkers in the fields of ecotoxicology.[119; 123] Such studies can contribute greatly to the understanding of cellular responses to altered environmental conditions. The development of molecular and genetic biomarkers of increasing sensitivity may still further develop by revealing exposures that are presently unknown or merely speculated.[117]

Proteomics, thus, has been a tool increasingly used in the study and discovery of biomarkers in ecotoxicological and environmental contexts and thereby has contributed to the understanding of organism responses to changes in the environment.[123; 124; 125] This type of analysis allows the isolation of sets of proteins within the proteome that are specifically modified by different stressors, either biological, physical or chemical. These sets of responses to specific pollutants that modify protein expression are used as protein biomarkers and are termed "protein expression signatures" (PES),[126] and they not only allow the discovery of new protein biomarkers, but also provide insight into the hidden mechanisms of contaminant toxicity.

There is currently a growing number of studies applying a comparative proteomic 2-DE-based approach to marine pollution biomonitoring using fish as a

model organisms in both laboratory and field experiments.[127] This has led to the establishment of "PES" following exposure to a variety of environmental contaminants including polychlorinated biphenyls (PCBs),[128] polyaromatic hydrocarbons (PAHs),[127] crude oil and metals.[129] Several research groups have pioneered the proteomic approach in environmental toxicology, but this field is still at a relatively early stage of its development.

Various biomolecules have been used as biomarkers of environmental exposure.[119] The antioxidant metalloenzymes superoxide dismutases, which bind to copper, iron, zinc, manganese and nickel are fairly well expressed in situations of oxidative stress in various organisms, including marine organisms [130]. Biotransformation enzymes, metallothioneins (MTs),[131;132;133] acetylcholinesterase (AChE) activity,[136;137] cytochrome P450 activities (responsible for the biotransformation of several xenobiotic compounds),[134] and morphological, haematological, histological and immunological parameters [139; 140] are also often used as measures of changes in organisms.

An interesting, biomolecule metallothionein (MT), has been extensively used in marine organisms as a biomarker of metal contamination. MT are proteins that bind to metals through cysteine residues present in high amount (approximately 30 % of MT amino acids) containing thiol groups, have molecular weight between 6 and 7 kDa, often appear in the form of dimers and are thermally stable.[135; 136] These molecules are capable of binding metal ions forming metal -thiol areas [137] and have an important role in cell regulation of metabolically important metals such as Cu and Zn. They also bind to non-essential elements such as Ag, Cd and Hg , thus presenting an additional role in reducing the toxicity of these metals in contaminated environments.[144; 145; 146; 147] The detoxification mechanism by these small molecules occurs through transcriptional activation of their genes by metals, leading to an increased synthesis of these proteins and their subsequent binding to free metals.[138]

Studies using MT for environmental monitory are abundant regarding the aquatic environment and are usually conducted by the analyses of liver tissue [139; 140], although studies using muscle,[141; 142] kidney [143; 144] and gills also exist.[145; 146] Liver measurements, however, are still the most employed, since this is the main detoxifying organ of the body and is a validated organ regarding exposure to environmental contaminants.[147;148;149;150] An alternative way to evaluate

contaminant effects on the proteomic or metalloproteomic of fish in environmental monitoring studies has been proposed, by using fish bile.[18; 151]

1.2.3.
Proteomics in fish biomarker identification

In the last decade, proteomic technologies have been increasingly used in fish biology research. These organisms are good models for this type of study, because they live in different environments and must adapt to environmental parameters and different stresses, which can usually be easily reproduced under controlled laboratory conditions, if necessary.[152] Fish are recognized as bioindicators of environmental change, including environmental contamination by different compounds.[164; 165; 166] Moreover, they are also important links between the environment, possible contaminants and human populations through aquaculture and consumption.[164; 165] In view of this, studies of exposure to environmental contaminants and their subsequent responses are of great relevance.

Proteomics have been applied primarily to investigate the physiology, developmental biology and impact of contaminants in these organisms, with global analyses of cellular signaling routes involved in physiological mechanisms related to growth, reproduction, disease and stress (including stress linked to environmental contamination). Studies as a way to quickly identify new proteins and homologs of known proteins that play important role in other animal groups have also been also conducted. However, the lack of genetic information for most fish species has been a big problem for a more general application of the currently proteomic technologies available.[153]

Several studies focusing on proteomics in fish have been published, such as the study by Martin et al. regarding the effects of lack of food on the protein profiles of trout livers compared with control animals, in which differentially expressed proteins were observed.[154] More recently, Papakostas performed a proteomic study investigating the differences in the proteome of the growth phase of salmonids, investigating development as a fundamental aspect of the biology of this species and showed that this can be affected by environmental

parameters.[155] Other studies using proteomic techniques aimed at the exposure of organisms to contaminants in the laboratory, providing information about possible reactions of certain contaminants and suggesting new biomarkers.[156; 157]

In this context, an interesting study examined the exposure of rainbow trout to sublethal doses of cadmium, carbon tetrachloride and pyrene and β-naphthoflavone for some days, and verified that the redox condition and metabolism of oxidative species was affected in the fish, in addition to increased expression of antioxidant and a large number of proteins involved in oxidative stress, the latter, particularly by β-naphthoflavone.[156] Another study with rainbow trout exposed to sublethal doses of zinc found observed the induction of the expression of the beta-chain of C3-1 protein, which plays an important role in immune response and imunoregulatory functions,[158] which may indicate that the induction of this protein by the metal in question has a stimulating effect on the immune system of this species.

The approach to the analysis of differential proteomic expression in complex field situations has also been studied. This approach can pinpoint which genes, proteins or metabolites are more interesting to study in the laboratory, based on the answers found in the real environment.[157] Field studies in proteomics are still scarce, however some studies have been published outlining the differences in gene expression of sampled fish from contaminated and uncontaminated estuaries in eastern England and metabolic differences and protein in the liver of sole with and without tumors.[159] In this study, 56 proteins were up-regulated and 20 down-regulated in tumor tissues, and 12 of these proteins exhibit the potential to act as biomarkers relative to neoplastic lesions. Some recognized environmental contaminants such as cyanotoxins, specifically microcystins, have also been studied in the context of fish proteomics. One study identified 17 differentially expressed proteins in the livers of fish exposed to these toxins for just two hours, including proteins involved in cellular structures, signal transduction within cells, regulation of enzymes and oxidative stress, such as methyltransferases and B - tubulin, among others.[160] These altered proteins corroborate studies on the toxic effects of these contaminants and their modes of action, e.g., causing cell disruption.

These studies dealt with non-model organisms, making a connection between traditional and emerging proteomic studies within the field of

ecotoxicology. Using sampled environmental organisms, attempting to discover biomarkers in environmental sentinels based on "omics" techniques, these types of research have the potential to be applied as rapid screening methods for disease classification and may in the end provide a mechanistic understanding of the effects of environmental stressors on the health of exposed organisms. Observing these different examples, it is clear that protein modifications can useful markers in environmental exposure to contaminants, and opportunities to study protein modifications are increasingly becoming important in ecotoxicological studies and environmental monitoring programs.[161]

1.2.4.
Bile as a bioindicator matrix

This biological matrix is a validated bioindicator regarding environmental contaminants, since it also excretes exogenous substances from blood and liver that were not excreted by the kidneys, such as metals and several organic compounds.[18; 19; 151; 162]

It is known that some environmental contaminants lead to morphological changes in various organs, especially the liver. Based on this, Morozov noted that a change in bile is also possible to be found.[163] Based on the aforementioned approaches, it is important to analyze this matrix as a potential bioindicator of environmental contamination, both in relation to the presence of metals and in the proteomic characterization of aspects that may possibly identify proteins of interest in an environmental context.

Bile or bile juice is a fluid made by the liver and stored in the gallbladder, and operates mainly on fat digestion (through the action of pancreatic lipase, an enzyme produced by the pancreas), in certain microorganisms to avoid decomposition of some foods and on the absorption of nutrients in the diet to pass through the intestine, besides also acting as a means of excretion of endogenous and exogenous substances.[164] Environmental studies have shown that many chemicals foreign to the body are excreted from the liver into fish bile and then expelled from the body, such as the hydrocarbons and their metabolites.[165; 166; 167; 168] Bile is also a route for drug elimination. The factors

influencing such elimination include chemical structure, molecular size and polarity, as well as the characteristics of the liver, such as specific transport active sites in the membranes of liver cells.[169; 170]

Studies also indicate that bile is also an excretion route for metals.[182; 183; 184] In laboratory studies with fish rainbow trout (*Oncorhynchus mykiss*) Cu, Hg and Pb were excreted in bile,[171] indicating that bil is the preferred route of excretion for Cu has, according to Grossel *et al.*[172] In the case of other elements such as zinc, excretion occurs equally either through the intestine or bile.[173] Ballatori shows bile as a route for disposal and transport of mercury, also demonstrating a close relationship of this metal to GSH present in bile.[174]

Studies with human bile on the other hand have identified several differentially expressed proteins in healthy patients and patients with tumors of the biliary tract such as cholangiocarcinoma, cholestasis (reduction or interruption of the flow of bile) and biliary stenosis.[175] Interestingly, studies regarding bile as a potential biomarker for contaminants are still very little explored. With the exception of the human species, there are few or almost no reports about bile in a pathologic context as a disease biomarker, or on identification of differentially expressed proteins in different environmental situations.

1.3.
Analytical techniques used in proteomic and metallomic studies

Analytical chemistry methods have an immense heaviness on approximately all fields of biomedical, biochemical and biotechnological research and are, therefore, indispensable for both the generation of fundamental knowledge as well as for the opening of novel applications in various fields of life sciences. Despite significant methodological and technical improvements achieved in the past, there is still a tremendous need for improvement of established methods as well as for the introduction of alternative approaches. In particular, protein analytics, both the analytics of single proteins (classical protein analytics) and the investigation of complex protein mixtures as found in living cells (proteomics) is still a great challenge for bioanalytics.[12] Several analytical techniques have been extensively applied in this area, especially in protein

quantification. Before applying analytical techniques to this problem, however, adequate sample preparation must be conducted.

1.3.1.
Sample preparation for proteomic and metallomic analyses

With the increase in the technological advances in proteomic techniques, sample preparation procedures have become a fundamental and critical step in order to obtain relevant data.[13] Due to the great variety of protein present in biological samples, the optimum sample preparation procedure for a given sample must be determined empirically. Ideally, the process should result in the complete denaturation, reduction and dissolution of the proteins in the sample. Some additional steps can also be used to further better sample quality. Considerations of this kind are crucial to obtain relevant results from proteomic experiments, and some specialists propose that the proteomic field has become limited due to lack of significant advances in sample preparation techniques.[14] Kozuka-Hata suggests an experimental flow (figure 2) for advance in proteomic studies by mass spectrometry from two standard methodologies used as sample preparation steps.[15]

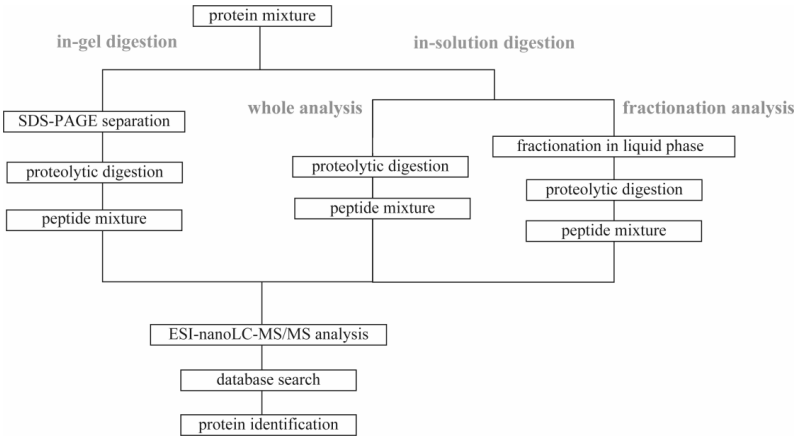

Figure 2. Scheme of experimental workflow for mass spectrometry-based proteomics. Adapted from Kozuka-Hata (2013).[15]

Basically, two major strategies for converting proteins extracted from biological material to peptides suitable for mass spectrometry (MS)-based proteome analysis are available. The first involves protein solubilization with detergents, their separation by sodium dodecyl sulfate (SDS) polyacrylamide gel electrophoresis and the digestion of the gel-trapped proteins ('in-gel' digestion).[13; 16] The second method uses no detergents, comprising protein extraction with strong chaotropic reagents (i.e, urea and thiourea), protein precipitation and their digestion under denaturing conditions ('in-solution' digestion). This second approach can be followed by two-dimensional peptide separation, for example, in the multidimensional protein identification technology strategy.[17]

The great heterogeneity of proteins and interfering contaminants makes the global protein extraction, dissolution and simultaneous release of all proteins a major challenge for any sample source. The incorporation of membrane proteins and the formation of complexes with other proteins or nucleic acids complicate the process significantly. Minimizing these effects by optimizing the extraction protocol is a challenge that usually has to be repeated for each new sample source and analytical purposes. Several studies regarding sample "clean-up" for biological matrices have been reported, that applied ultra-centrifugation, adding commercial solutions for lipid removal, filtration for salt removal, protein precipitation and fractionation by adding different solvents such as acetone and dichloromethane for the removal of proteins present in such high quantitates that they may be considered an interferent,, such as albumin.[18; 19]

Basically, sample preparation for proteomic studies is composed of three steps: (i) contaminant removal or inactivation. This can be performed by using protease inhibitors, by conducting salt removal by dialysis, gel filtration or precipitation and by conducting the removal of nucleic acids, which increase the viscosity of the sample and can cause spots, besides obstructing gel pores. Recently, Wisniewski *et al*.[13] conducted a novel approach in detergent removal from membrane proteins by using a filter assisted sample preparation technique, where the use of strong detergents is used to "clean" the proteome; (ii) enrichment of the peptides of interest (modified or not), and, (iii) sample pre-fractioning through chromatography and/or electrophoresis.[20; 21; 22] The sample pH must be carefully monitored, heat must be avoided (especially in samples containing urea) and care to avoid proteolytic degradation is paramount (using

proteinase inhibitors or working on ice).[23] Protein dissolution is also critical, since possible interferences by unwanted protein precipitation may occur.[24]

1.3.2.
Analytical methods used in protein quantification and characterization

Protein identification and quantification can be achieved either in the level of proteolytic peptides (shotgun or bottom-up approach) or in the level of intact proteins (top-down approach). In bottom-up approaches, protein identification relies on peptide fragment fingerprinting, in other words, the analysis of the MS/MS spectra of the peptides. In top-down approaches, after the separation of intact proteins, they can be identified either after proteolytic digestion by peptide mass fingerprinting analysis or by a combination of the latter with MS/MS data.[25; 26] Stimulated by the establishment of novel experimental and (bio)informatics approaches, such as the development of soft ionization techniques for molecular mass spectrometry (MS) (i.e. electrospray ionization (ESI) and matrix assisted laser desorption/ionization (MALDI) for the analysis of biopolymers, peptide sequencing by tandem mass spectrometry (MS/MS), the development of database driven identification of proteins, differential two dimensional gel electrophoresis (DIGE) and multidimensional chromatographic separations of complex peptide and protein mixtures),[27] significant progress has now been achieved in the field of protein identification as well as in the qualitative analysis of posttranslational modifications. The acquisition of this type of data is also possible by using hyphenated (coupled) techniques. These techniques combine a high resolution separation technique with sensitive element- or molecule-specific detection.[28]

Despite the progress made in different fields of protein and proteome analytics in the past, there is still the need for further improvement of these techniques. Both for classical protein analytics and for proteomics, protein quantification and the quantification of the degree of possible posttranslational modifications [27] is still a challenge: quantification, both absolute and relative (differential), is still characterized by low accuracies; furthermore, in most methods, the number of samples analyzed by multiplexing is still restricted. Thus, the development of novel approaches for the separation of proteins on level of intact proteins, the development of novel separation schemes for separation in shotgun proteomics and novel instrumental improvements on the side of mass spectrometry bear significant potential for further increase of confidence of identification and for the sensitivity, which is necessary for the coverage of complex proteomes. In fact, integrated elemental and molecular MS approaches could eventually prove to be the most useful in the field of proteomics.[29]

Up to now, most of the available quantitative proteomics research is based on relative approaches, this means comparing protein expression between two (or more) organism states.[30] Relative quantification of proteins based on imaging techniques (e.g. after protein labelling with dyes, fluorophores or radioactive substances) have good general performance, but require high resolution protein separations (typically using 2D gels) and do not give information regarding the identity of the measured protein. On the other hand, MS techniques may overcome such limitations, and so MS derived techniques have been increasingly used for this purpose. The major steps of these analyses involve sample preparation, the application of well-suited separation methods, mass spectrometric identification, characterization and quantification followed by data analysis and interpretation with bioinformatics tools.[13; 20; 31]

Mass spectrometry (MS) has experienced consistent instrumental improvements. Due to these important developments and their associated applications the current role and potential of MS is huge for proteomics-based investigations.[32] In fact, MS today is not just an established tool for structural proteomics research, but it is more than ever at the the first stage of functional proteomics.[33] So far, however, many efforts launched to proteomics have focused on developing methodologies that promote the efficient identification (and perhaps structural characterization) of a large number of proteins, while quantification studies are, so far, comparatively few.[32]

For the differential quantification of protein expression in two or more biological samples, labelling with stable isotopes can be used.[34; 35] The labelled peptide (or proteins) in most cases share equal or, at least, similar physicochemical properties, allowing for a simultaneous separation of differentially labelled proteomes. The peptides originating from the different biological states can be distinguished by their different masses in MS analysis; their ratio, and hence the ratio of the proteins in the biological matrix to be investigated, can be read out by comparison of the corresponding signal intensities of the light and heavy isotope signals in the MS spectra. In other approaches, *e.g*. iTRAQ,[36] relative quantification can be performed by using signals of reporter ions derived upon release from the labelling reagent in MS/MS analysis. This method also shows the potential for a medium multiplexing capacity (an assay that simultaneously measures multiple analytes - dozens or more - in a single run). The introduction of (isotope) labels in the peptides or proteins can be achieved either by chemical modification,[37] by means of enzymatic methods,[38] or by methods allowing for an *in vivo* labelling, as SILAC (Stable Isotope Labeling by/with amino acids in cell culture)[39] or proteome metabolic labelling.[40] The latter approaches are restricted to organisms allowing for supply with labelled precursors; thus the application for human samples is restricted; furthermore, these methods provide only minor capacities for multiplexing. First approaches for label free quantifications (*e.g*. by spectral counting, by Liu *et al*.)[41] have been introduced very recently.

Regarding metallomic studies, there is a significant difference in the number of discoveries in this field when compared to strictly metallomic studies. The late advent of this field is due to several singular questions that should be considered when analyzing metalloproteins. These include the absence of any polymerase chain reaction similar to PCR in the genomics field, the occurrence of post-translational changes and, finally, the low concentration of the trace-elements present in biological tissues (usually < 50 $\mu g\ g^{-1}$) and the high complexity of these matrices.[42] These factors mean that the analyses of metals bonded to biomolecules is very difficult and challenging. Because of this, metallomic analyses require sophisticated multidimensional analytical approaches. The continuous development of techniques that mix atomic spectrometry and biochemical or proteomic techniques, such as gel electrophoresis, capillary chromatography and multidimensional nanoflow, and

the development of strategies for the complementary application of elements and specific techniques for detecting molecules, have led to new possibilities in this field of research.[43] For example, mass spectrometry with approached such as coupling with inductively coupled plasma (ICP-MS), electrospray (ESI-MS) and laser assisted ionization and desorption (MALDI-MS) are now routinely used, alongside 1 or 2D protein separation by gel electrophoresis, in order to quickly and accurately identify the metalloid component of individual proteins. By applying these techniques, large volumes of data can be collected.[44]

Element mass spectrometry with inductively coupled plasma ionization (ICP-MS) is a very interesting alternative for absolute peptide and protein quantification in metallomics. This technique is highly sensitive, allows for multielement and multi-isotope detection, has a large dynamic range, and the ionization process is almost compound- and matrix independent.

The chemical derivatization of several types of biomolecules with different functional groups, isotopes, metals, or radionuclides has been a widely used in bioanalysis, allowing for sensitive and specific detection of the analytes.[45; 46; 47] Within this field an emerging new trend is the labelling of biomolecules, such as peptides and proteins which do not naturally contain an ICP-MS-detectable element, with heteroatom tags in order to make them visible and quantifiable by ICP-MS.[48;49] In 1992, the derivatization of sulfhydryl groups with p-chloromercuribenzoate was described. With the view to quantify cysteine residues in metallothionein via the ^{202}Hg signal in HPLC-ICP-MS.[50] For ultra-sensitive detection, the most suitable elements are metals, such as lanthanides, with high ionization efficiency, low background and no ICP-MS interferences, as recently described by Ahrends and co-workers.[51] Since a direct covalent bond of a lanthanide to a protein is not possible, metal complexes with strong chelators, such as DOTA derivatives or DTPA, can be alternatively used as tags binding to functional groups such as amino or sulfhydryl groups in peptides and proteins.[52; 53; 54] Lanthanide-DOTA tags have been recently applied to quantify entire proteins in flow injection (FIA) ICP-MS,[51] and first results on tagged peptide and protein analyses with LC-ICP-MS are available.[55] Most importantly, they showed complete tagging of all sulfhydryl groups even in complex protein mixtures when the reagent was applied in an excess of 20 times. Amino groups in peptides can be also derivatized with DTPA, in a reaction with its bicyclic anhydride (DTPAA).[52; 56] Other recent preliminary study used lanthanide-DTPA tags and

ICP-MS for absolute peptide quantification, which requires a highly specific, efficient and ideally complete, reaction.[57]

1.3.2.1.

Total metal determination by ICP-MS and protein-bound metal determination by hyphenated systems coupled to ICP-MS detection

Inductively coupled plasma mass spectrometry (ICP-MS) was developed in the 1980s. This technique combine the easy sample introduction and rapid sample analysis of the ICP with the accurate and low limits of detection of a quadrupole mass spectrometer, thus obtaining an isotopic and elemental analyzer in a single instrument. The resulting instrument is capable of performing multielemental analysis, often at the ng L^{-1} level.[58]

The ICP-MS technique can accept mainly liquid samples, but solid samples can be introduced into the ICP through any specific accessory. Aqueous samples are usually introduced pneumatically by means of a peristaltic pump that takes the sample to a nebulizer operated with argon at high speed, forming a fine mist. The aerosol then passes into a mist chamber where larger droplets are removed through a drain.[59] Typically, only 2% of the original solution is aspirated through the spray chamber.[60] This process is necessary to produce droplets that are small enough to be vaporized in the plasma torch.

Once the sample solution passes through the nebulizer and is partially desolvated, the aerosol in movement towards the torch blends with more argon gas and reaches the plasma. The initial plasma is formed by ionization of argon, then occurs the transfer energy to this gas, partially ionized, through a coupling coil, which is used for transmission of radio frequency, in order to heat it.[59] The hot plasma removes any solvent and causes the atomization of the sample, followed by excitation and ionization.

The atomization/ionization occurs at atmospheric pressure; therefore, the interface between the ICP and the MS becomes crucial to create a vacuum environment for the MS system. The ions pass through a small hole in a vacuum system, where a supersonic jet is formed and the ions are then passed to the MS

system at high speeds, expanding in the vacuum system.[59] The vacuum is needed so that the ions are free to move without collision with the air molecules. In the first stage of MS, the ions are removed from the plasma by an extraction pump system. An ion beam is produced and focused and then taken to the mass separator and subsequently to the detector.[61]

The use of ICP-MS has increased significantly within the area of proteomics. One of advantages of the ICP is its capability to generate, as an ion source, monoatomic positive ions from most elements, besides being multielemental. Thus the elements can be identified based on their atomic mass spectra without the isotope peaks that complicate the mass spectra of the parent ions of biopolymers. Another advantage is the independence of the signal intensity of the molecular environment of each element. This allows the possibility of using inorganic elemental standards for the protein analysis.[62]

As the ICP-MS detector information is elemental, only the heteroatoms are in fact measured, while the interest may lie in the molecules (e.g. peptides). Thus, for proteomic and metallomic analyses, the elemental detector should be coupled to a previous separation (e.g. HPLC, capillary electrophoresis (CE) or 2D gel electrophoresis) of the element-containing biomolecule. Such instrumental approaches, mainly HPLC-ICP-MS, have been widely developed in the last decade within the particular field of trace element speciation in biological materials.[63; 64] In this context, several elements can be identified in the same chromatographic peak. In HPLC-ICP-MS approaches all peptides relevant to quantification must be separated before entering the MS, because all structural information is lost in the plasma, unlike molecular MS. Consequently, separation of the biomolecules and information on their stoichiometry obtained by parallel molecular MS is mandatory to accurately quantify them via the element signal.

The increasing use of ICP-MS for biochemical applications has been inseparably linked with progress in three major areas. This included (i) advances in non-metal detection, (ii) increase in detection limits of ICP mass spectrometers allowing the coupling of capillary and planar electrophoretic techniques, and (iii) easier access to stable metal isotopes.[11]

The current state of the art of these techniques for the bio-inorganic speciation analysis is shown in Figure 3.

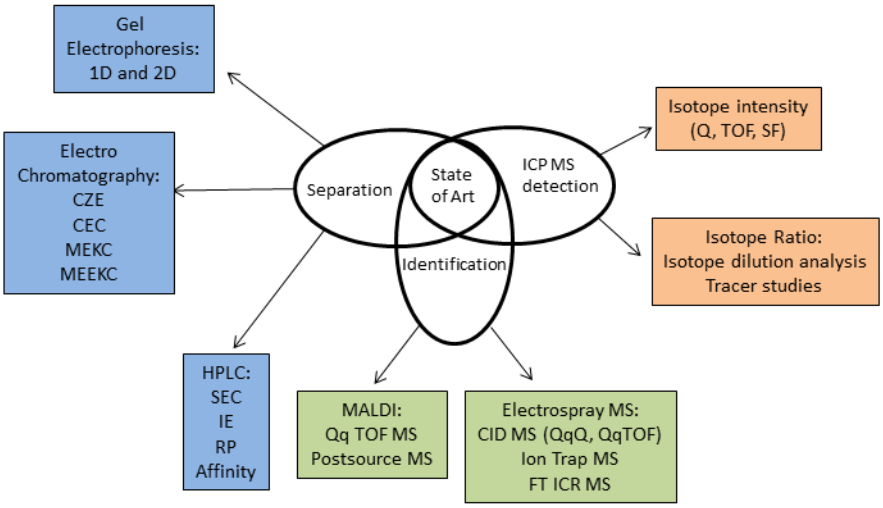

Figure 3. Scheme of hyphenated systems using ICP-MS detection.

For most elements, ICP-MS is a routine detection technique in chromatography and has been discussed elsewhere.[28; 62; 65] However, some biologically important elements, such as sulfur and phosphorus, are not efficiently ionized in the ICP. They also suffer from several polyatomic interferences. The problem of the low ionization efficiency also happens with halogens, especially fluorine. Also, some vital transition elements, such as iron or vanadium, show interferences by polyatomic ions [66]. Polyatomic interferences can be removed by using a high resolution mass spectrometer or a collision/ reaction cell.[67] The latter offers two approaches to the resolution of interferences: (i) by collision/ reaction of the interfering ions with gases such as H_2, He or Xe, or (ii) by reaction of the ion of interest with oxygen.[68]

1.3.2.2. Ultraviolet-visible molecular absorption and its applications in proteomics

Ultraviolet-visible molecular absorption spectroscopy is routinely used as an auxiliary technique for the characterization and identification of the proteome and metallome. This is a simple and quick method, indicated for the

determination of naturally colored analytes, or those that acquire color adjusted by the use of derivatization, absorbing light at certain wavelengths which can then be analyzed by a spectrophotometer.[69]

Molecular absorption spectrophotometry involves the energy absorbed in the ultraviolet-visible (UV-Vis) region by promoting the transition of valence electrons. This technique is directly related to the Lambert-Beer law, an empirical relationship that relates the optical absorption of light to the properties of the solution which it traverses. This law says that there is an exponential relationship between the light transmission through a substance and the concentration of said substance, as well as between the transmission and the length of the body that the light passes through. The concentration (c) of the substance can be deduced from the amount of light transmitted, through the equation $A = \varepsilon l c$, where A is the absorbance, ε is the molar absorptivity of the species and *l* is the optical path traveled by the light. Since protein molecules are able to absorb energy, the quantification of protein concentrations in solutions can be performed using this colorimetric technique.[70]

Some colorimetric methods commonly used in protein quantification are:

- The Biuret method (sensitivity - 0.5 to 10 mg mL^{-1})

The reagent of biuret is a compound formed by heating urea at 180 °C. When it is placed in the presence of a copper sulfate solution in an alkaline medium a blue compound is formed. The color is due to the formation of a complex between cupric ions and four adjacent nitrogen atoms. This type of reaction also occurs with peptides containing at least two peptide bonds and with proteins in general. Substances containing two carbonyl groups attached directly or via a nitrogen atom also appear blue in the copper sulfate alkaline solution. The colored product of the reaction has maximum absorption at 540 nm.[71; 72]

- Folin-Lowry method (sensitivity - 0.1 to 0.3 mg mL^{-1})

In alkaline conditions a divalent copper ion is capable of forming a complex with peptide bonds and is reduced to a monovalent ion (the biuret reaction). The monovalent copper ion alongside some the lateral chains of some protein amino acids (tyrosine, tryptophan, cysteine, histidine and asparagine) lead to the reduction of acid components present in the Folin reagent, amplifying

the color first obtained by the biuret reaction. In this method, the maximum absorption occurs at 650-750 nm.[73]

- BCA method (sensitivity - 0.1 to 0.5 mg mL^{-1})

This method is also known as bicinchoninic acid method. This procedure is applied for analysis in microplates. Unlike the Folin-Lowry colorimetric reagent, bicinchoninic acid is more stable under alkaline conditions.[74] BCA follows the principle of the Folin-Lowry assay, reacting with the complexes between copper ions and the peptides to produce a purple color which absorbs strongly at 562 nm. One advantage is that, as the BCA reagent is more stable under alkaline conditions, it can be added to the copper solution to allow a one-step procedure, making it faster than the Folin-Lowry technique.[75; 76]

- Bradford method (sensitivity – 0.06 to 0.3 mg mL^{-1})

When at acidic pH, the anionic dye Coomassie Blue forms complexes with proteins which contain basic and/or aromatic amino acids. The interaction between the protein and the dye causes the change in the wavelength of maximum absorption of the dye (465 nm - free dye) to 595 nm (protein complexed dye).[77]

- Ellman's reaction for the determination of free thiols (sensitivity – 50 to 1000 $\mu mol\ L^{-1}$)

It works well for small peptides and proteins synthesized using standard solid phase synthetic methods. Peptides from these syntheses are usually in their reduced form, and are usually stable to oxidation in acidic solutions. Free thiol can be determined in solutions collected from chromatographic separations or from reconstituted lyophilized samples. This protocol has been used for peptides with a single Cys residue present and lacking tryptophan. The technique should be feasible for multiple Cys residues.[236]

There is, therefore, no universal quantification method. It is recommended to set the method to be used taking into account (i) the sensitivity required (dependent on protein concentration), (ii) the presence/absence of certain amino acids, because of the operating mechanism, i.e specificity, of each method, (iii) the nature and concentration of non-protein substances (potential interferents)

(iv) protein solubility under the method conditions (v) ease and reproducibility; (vi) speed and cost.

The choice for using spectrophotometric quantification is considered more adequate than other available techniques when the aim is to screen several environmental samples at a time, since it is, for example, quicker than voltammetric analyses and is also simpler and less costly. Because of this, this technique has been used for several decades as a simple quantification tool for MT analyses in ecotoxicological and environmental monitoring studies.[183; 184]

1.3.2.3. One- and two-dimensional gel electrophoresis

There is still no technique able to separate all the proteins of an organism. A widely used technique with efficient resolution is one- and two-dimensional gel electrophoresis. Gel electrophoresis has several advantages, such as being cheap, and, especially robust against impurities, that interfere, mainly, in protein digestion. In-solution digestion is easily automated and minimizes sample processing, but the proteome may not be completely solubilized and digestion may be impeded by interfering substances, making gel electrophoresis an interesting alternative.[17]

The polyacrylamide gel, introduced in 1959, is the most commonly used matrix in electrophoresis to separate the proteins of a biological sample. The gel is a matrix consisting of acrylamide cross-linked with N, N-methyl bisacrylamide. Gel porosity can be chosen, to better adapt to the sample to be separated. The usual percentages are 5%, 7.5%, 10%, 12.5% and 15%. The higher acrylamide concentration, the smaller the pores of the resulting gel. Ammonium persulfate is used to generate free radicals and tetramethylethylenediamine (TEMED) is the catalyst that assists in the transfer of the electron of the free radical. The use of SDS in the gel system (SDS-PAGE) is so that the separation depends only on the molecular weight of the protein, rather than its form and native charge. SDS-PAGE was first described by Laemmli in 1970.[78] SDS is an anionic detergent which interacts with the peptide chains of the protein, denaturing them and

forming negatively charged SDS-protein complexes (Figure 4). By applying an electrical current, all proteins, now with the same charge, migrate toward the positive electrode and are separated only by differences in their molar masses. The smallest proteins migrate more quickly, while larger proteins find it more difficult to pass through the gel mesh and, thus, move more slowly. SDS-PAGE electrophoresis is a robust technique that has a wide range of application.[79] It is nowadays mainly used routinely to (i) estimate protein size, (ii) establish protein purity, (iii) quantify proteins, (iv) compare protein content of different samples; (v) analyze the size and number of protein subunits.[20; 24]

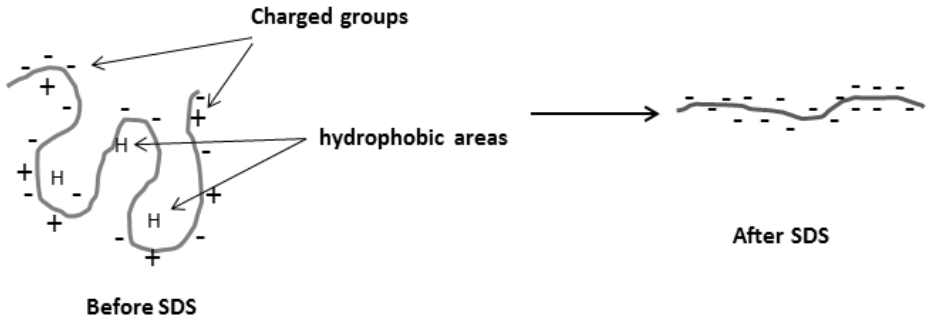

Figure 4. Scheme of protein modification with SDS in gel electrophoresis.

One-dimensional (1D) SDS gel electrophoresis in denaturing conditions (1D SDS-PAGE) consists in the migration of the solubilized proteins in a discontinuous electrophoresis system consisting of two polyacrylamide gels with different porosities and buffered at different pH, see Figure 5. The extracted proteins are solubilized, heated to 100 °C in a buffered Tris- HCl solution, pH 6.8, containing excess SDS and a reducing agent of the thiol group generally β-mercaptoethanol or DTT. SDS charged the proteins negatively and denatures them, aided by warming around the polypeptide backbone. The reducing agent reduces the disulfide bridges of the proteins that will thus unroll into the individual polypeptide subunits. Gels are bonded and polymerized in a vertical system - sandwich - between two glass plates. The sample is placed in the first gel (stacking gel) in a buffered Tris-HCl pH 6.8 solution, with wider crosslinking. Under these conditions, the presence of tricine in the buffer solution will allow protein concentration in the stacking gel which will then be separated according

to their molecular weights in the separation gel, buffered at pH 8.8 with much smaller cross-linking. pH 8.8 leads to an increase in the speed of migration of tricine ions, which will then migrate faster than the proteins, which are then not entrapped in a thin layer and are now separated according to their size by migration towards the positive electrode and, at the end of the run, they may be compared to a standard solution of known molecular weight.

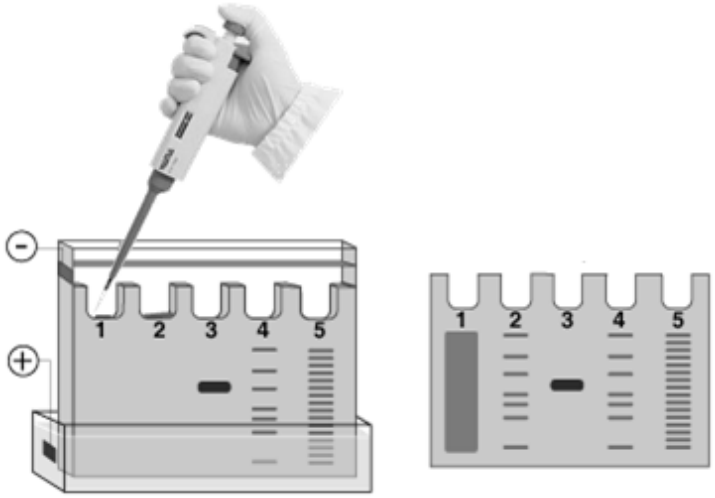

Figure 5. Example schematic of protein bands separated by 1D electrophoresis, where each trace in numerical columns represent protein bands of decreasing molecular weight.

Two-dimensional gel electrophoresis in denaturing conditions (2D SDS-PAGE) was introduced in the 1970s and perfected by O'Farrell.[80] Today this technique allows the resolution of up to 5000 proteins and continues to be the more direct way to map the proteome of an organism.[81; 82] In this technique, proteins are separated by two steps, through two of their physicochemical properties, isoelectric point (pI) and molecular weight (MW) (Figure 6). To separate proteins by pI, they must first pass through isoelectric focusing (IEF). A pH gradient, in which the charged proteins move to their pI, where no charge is present, is formed. Prior to the development of commercially available immobilized pH gradients backed on plastic strips, two-dimensional electrophoresis was performed using a discontinuous system, which did not allow for high reproducibility of IEF gels. With the development of these immobilized gradients, electrophoresis became extremely reproducible, and is still the most used to perform protein separation. After IEF, the strips are equilibrated in a solution DTT-containing to break the disulfide bridges (SS bonds) present in

proteins, with SDS facilitating access to their inner parts, due to the elimination of protein tertiary structure, and iodoacetamide to prevent reoxidation of thiol groups (alkylation). After IEF, the proteins are separated on SDS-PAGE gels in the same manner as 1D SDS-PAGE.

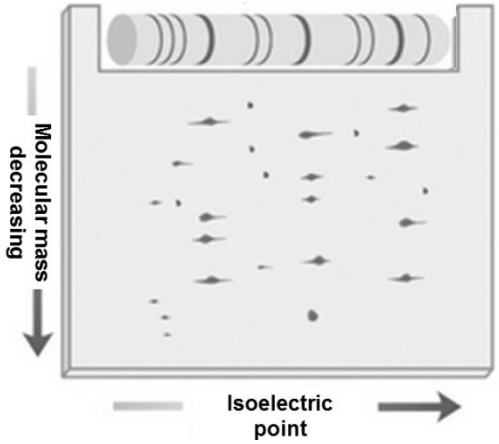

Figure 6. Example schematic of protein spots separated by two-dimensional gel electrophoresis, after the isoelectric focusing.

The final result, after gel staining, is a protein profile in which each spot corresponds to a set of few proteins, facilitating subsequent analysis by mass spectrometry. However, before MS analyses, the different proteins in the gel must be revealed, by staining with different methods. The choice of dye or fluorescent agent will determine the sensitivity of the technique. The most commonly used staining methods are reported in Table 1 with their respective detection limits and corresponding linearity ranges.

- Comassie blue: although less sensitive than other methods, this method is perfectly adequate for comparative studies, due to its linear response.

- Silver staining: this method is 100 times more sensitive than Coomassie blue staining, however, due to a large number of steps and reagents, may become more costly. Under certain concentrations the response becomes non-linear and incompatible with mass spectrometric analysis by MALDI and ESI, as reported by Gevaert *et al*.[83]

- Fluorescence staining: currently a very sensitive staining, with a wider range of linearity, however, uses a fluorescent scanner, which is extremely expensive.

Other types of staining have also been developed in order to increase sensitivity and compatibility with mass spectrometry but have not been widely applied.[83; 84]

Table 1. Main protein staining methods after one- or two-dimensional gel electrophoresis.[84]

Staining method	Limits of detection (ng)	Linearity range (order of magnitude)
Coomassie blue	50 - 100	1 - 1,3
$AgNO_3$	1	2
Sypro® Orange, Ruby (Fluorescence)	1 - 8	>3
Deep purple ® (Fluorescence)	1	4

The analysis method of gel staining evolved thanks to the combined progress of computers and imaging analyses. The first stage of analysis is gel scanning, i.e., the transformation of the image in an experimental numerical information usable by the computer. The scanned image is then handled by specific software. The relevant information is then obtained from the differential and comparative analysis between proteins present in the gels. After choosing which protein bands or spots are of interest, they are excised, digested (usually by trypsin) and analyzed by MS.

1.3.2.4.

Protein identification by mass spectrometry after separation by gel electrophoresis

In biology, the study of proteins involves the determination of their primary amino acid sequence. Edman degradation was, for a long time, chosen for the determination of the amino acid sequence using the automated sequencer marketed in 1967.[85]

Since the early 90s this method was overtaken by mass spectrometry, which currently is the most sensitive method for biomolecule characterization. Molecular mass spectrometry is now a standard technique for the identification and characterization of biomolecules. This growth was based on the development of two new ionization techniques in 1989, by Toichi Tanaka and John Bennett Fenn, respectively Matrix Assisted Laser Desorption - MALDI and electrospray Ionization - ESI. These techniques allow the transformation of biomolecules into ionized gas for the detection of their mass by mass spectrometry.

1.3.2.5.
MALDI mass spectrometry

Michael Karas and Franz Hillenkamp were responsible for the name and initial development of this technique in the late 1980s.[86] In the early 2000s, Tanaka pioneered the use of the technique for protein analyses,[87] and was one of the winners of the Nobel Prize in chemistry for developing desorption and ionization spectrometric methods for biological molecule analysis.

This technique involves radiation of a pulsed laser beam usually in the UV in a crystalline deposit containing an organic matrix and the sample (Figure 7). The analyte molecules are co-crystallized with the sample in a chemically inert matrix on stainless steel. A MALDI matrix is a small molecule capable of strongly absorbing the UV laser. The most commonly used matrices are α-cyano-4-hydroxycinnamic acid (α-CHCA) for the analysis of peptides and proteins, sinapinic acid, for the analysis of whole proteins, and 2,5-dihydrobenzoic acid (DHB), which is a polyvalent matrix. In order to facilitate ionization of the analyte molecules, the array is used in a molar excess of about 500x. This excess matrix is needed since the crystals are the matrix that that absorb the UV laser light, with the sample being only, metaphorically, a built-in "impurity" in the crystal matrix. When the laser pulse impacts with the matrix in vacuum, the matrix molecules release their excess energy absorbed by sublimation and concomitant

molecular fragmentation, however, if properly prepared, the sample is brought into the gas phase by a sublimation plume without fragmentation.[88] The process of ionization is not fully elucidated; the mechanism is complex and involves physical and chemical processes.

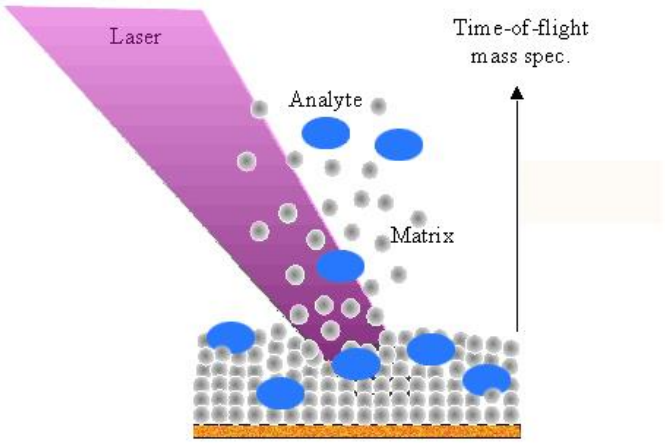

Figure 7. MALDI-MS scheme. Adapted from Fitzgerald (1995).[89]

The MALDI technique was first used with Time-of-Flight (ToF - Time of Flight) mass spectrometers analyzers. The ions are formed in the ionization source after a laser pulse, and are then accelerated by an electric potential. The ions then traverse the flight tube, which has a region free of electric field and arrive at the detector at characteristic times, according to their m/z. This period is roughly proportional to the square root of the mass-charge ratio (m/z) ion, and can be used to calculate the mass of the ion. Therefore, lighter ions have a shorter ToF than heavier ions (Figure 8). The fast analysis time (less than a millisecond) enables quick and simple analysis of different samples.[89]

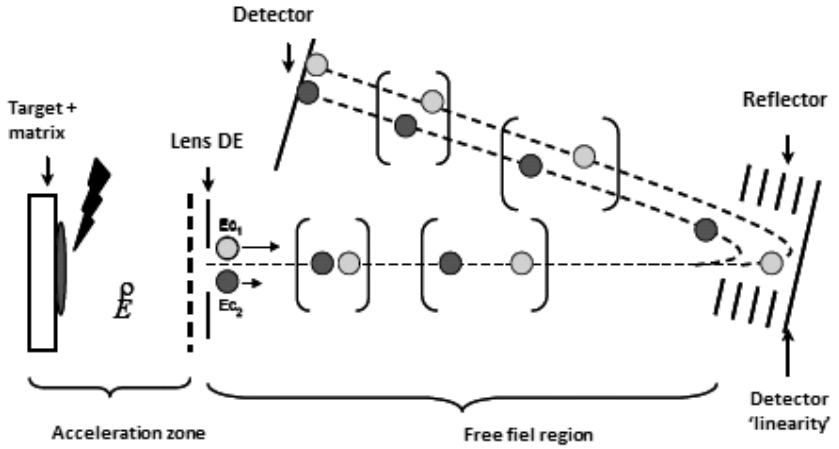

Figure 8. Schematic representation of the operating principle of a MALDI-TOF. Adapted from Horn (2004).[90]

The MALDI-TOF technique allows for the analysis in a wide mass range (unlimited in theory, however, what occurs in practice is up to 500 kDa) with good precision measurement (10 - 20 ppm), good resolution, good sensitivity, and a relatively simple and fast implementation, tolerant to salts and detergents commonly used in biology.[90]

1.3.2.6.
ESI MS

In a traditional approach the peptides are produced from a tryptic digestion, after gel electrophoresis, or directly from aqueous sample solutions. In cases where MALDI TOF MS analyses are not feasible, high performance liquid nanochromatography (nanoHPLC) coupled to an MS electrospray is an alternative analysis for the determination of peptides and identification of proteins from which they are derived.

The phenomenon of electrospray ionization was described by Dole in 1968,[91] but it was only in the 80s that this ionization technique was coupled to a quadrupole analyzer for the purpose of protein analysis.[92; 93] Because of its particularity in forming multicharged ions, ESI-MS allows the analysis of more heavy and complex molecules with a considerably "thin" (0.01%) mass

measurement accuracy. With this, the formation of ions at atmospheric pressure in the electrospray source allowed this technique to be coupled with liquid chromatography (LC-MS).[92] The general principle of mass spectrometry ionization electrospray is displayed in Figure 9 and detailed in the next paragraph.

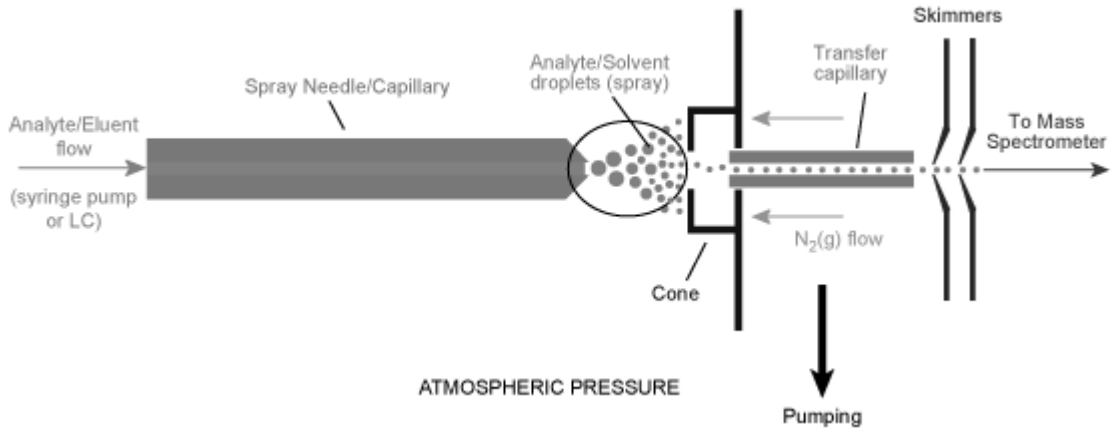

Figure 9. Ion formation in ESI. Adapted from Cole (1997).[94]

Electrospray ionization involves the formation of an electrostatic spray, from which charged droplets are generated and ions are released. The implementation of an electrospray source is relatively simple compared with other sources for mass spectrometry.[94] A source of high voltage (1000-7000 V) in contact with the solution containing electrolytes is required. Typically, the solution is pumped through a capillary (i.d. 50 to 100 mmol L^{-1}) with a flow rate of less than 10 L min^{-1}. In the case of less than 1 µL min^{-1} flow, the process is called nanoelectrospray.

All the source region is at atmospheric pressure. When a positive potential, for example, is applied to the solution, positive ions tend to move away to a less positive region, i.e., toward the counter electrode. Thus, the drop being formed at the tip of the capillary is enriched in positive ions. This type of charge separation is called electrophoretic process. As the charge density increases in the droplet, the electric field between the capillary and the electrode increases, causing the deformation of the drop. The drop takes the form of a cone which is called Taylor's cone.[95]

This cone-shaped drop remains "stuck" to the capillary until the charge density on the droplet surface and the increase in the repulsion between ions overcome the surface tension of the liquid, with the consequent release of small drops with high charge density. The frequency of this process depends on the magnitude of the electric field, the surface tension of the solvent and the conductivity of the solution. These small droplets formed are subjected to the same process several times until they form completely desolvated ions, and finally are transformed into gaseous form.[96] Once formed, the ion reaches the region of the capillary or sampling cone with a very small hole (of the order of a few micrometers).

From there, the system switches to the low pressure region. At the entrance, there is a set of lenses that lead ions to the mass analyzer according to their mass/charge ratio. Several types of analyzers compatible with electrospray ionization mode are marketed and the choice of mass analyzer used will be made according to the desired resolution, mass accuracy, sensitivity and ability to perform tandem mass spectrometry, MS/MS. In proteomic analyses, the following mass analyzers are commonly used:

- ion trap (IT);[97]
- orbitrap;[236]
- time of flight (TOF);[98]
- quadrupole (Q);[92]
- Fourrier transform ion cyclotron resonance (FT-ICR).[99]

These analyzers are used alone or in combination to create hybrid analyzers such as the Q-TOF type, benefiting the strongest points of each of the chosen analyzers.

1.3.2.7

Approaches to protein identification after MS analyses - databases and bioinformatics

The set of research strategies currently available for the automatic identification of proteins with mass spectrometry data was recently compiled by Hernandez *et al.*[100] Protein identification is performed using the values of the molecular weights of protein peptides obtained by mass spectrometry. In the case of MALDI, we use the monocharged signals comparing the experimentally obtained values with expected values stored in the database. All types of database sequences can theoretically be used, such as protein sequence banks, ESTs (Expressed Sequence Tag) banks and genomic libraries.[101] The proteomic analyses mostly use protein banks, which are usually "corrected" by selection and classification performed by specialists.

These databases are numerous, but the most known and used are the National Center for Biotechnology Information - NCBI and UniProt consortium banks. The NCBI database is bulky and sometimes redundant, but with the advantage of being relatively complete. It is accessible at http://www.ncbi.nlm.nih.gov/gquery/gquery.fcgi. The UniProt Consortium centralizes SwissProt, TrEMBL and PIR-P banks. These banks are accessible at http://www.expasy.uniprot.org/ site. The PIR-PSD (Protein Information Resource Protein Sequence Database) database is organized according to the classification by proteins families and super-families. This library contains functional annotations and structural and genetic bibliographic references. The SwissProt data bank is a library of less redundant protein sequences and is quite accessible. A special precaution is required to indicate the experimental nature or bioinformatics, functional information, as well as the confidence level given to this information.[102] When it comes to identifying proteins, the quality of the proteome present in the library varies depending on the organism. Thus, it is important to consult information related to multiple databases before using the data, to know the number of entries in the target organism, how these inputs were generated (by simple automatic or manual annotation prediction) and what is the status of the genome sequencing.[103] The knowledge of the genome of an organism is of great importance to enable the accurate identification of proteins by peptide

standards and, therefore, organisms that are not very well studied often do not have their proteins identified.

1.3.2.8.
Alternative methods for protein characterization

1.3.2.8.1.
Fourier-Transform infrared (FT-IR) vibrational spectroscopy in qualitative protein characterization.

Fourier-Transform infrared (FT-IR) vibrational spectroscopy is a measurement of the wavelength and intensity of infra-red radiation absorption of a sample. This is one of the oldest and most well-established experimental techniques for the analysis of the secondary structure of polypeptides and proteins. It is a non-destructive technique that requires little practice in sample preparation, and can be used under a wide variety of conditions.[104]

The vibrational spectrum of a molecule is determined by its three-dimensional structure and vibrational force field. An analysis of this spectrum in different regions, such as infra-red, can, therefore, provide information on the molecule structure and on the intramolecular and intermolecular interactions. A greater probing analysis of this spectrum is able to provide increasingly detailed information about the studied molecule.

While the three-dimensional structure and force field are determined solely by the vibrational frequencies of the molecule, the structure is, in general, not obtained directly from the spectrum. However, the atomic displacements in many of the vibrational modes of a large molecule are transferable between molecules. Therefore, in early peptides and proteins studies, efforts were mainly directed towards the identification of such characteristic frequencies and determination of their relationship to the molecule structure.[105]

Early protein studies by infrared spectroscopy were only qualitative. The position and number of bands were not accurately verified, and only

approximations regarding changes in the relative intensities of shoulders or amide I band shifts could be made.

Overlapping components in spectral profiles are common for complex samples such as proteins. The complexity of information contained in infrared spectra of proteins increases the difficulty of analysis, even with the increasing number of new methodologies for quantitative analysis of IR bands.[106]

The pioneering work in quantitative protein analysis by IR was performed by Susi and colleagues.[107] The work reported by Ruegg[108] included the analysis of the 1750-1300 cm^{-1} region of seven proteins, four of them with known X-ray structures. The information necessary to begin the quantification process was the number of tracks, the shape of the peaks, the peak frequencies and their allowed frequency range, and landmarks outside the absorption region where the total absorbance was initially assumed to be zero. Since the number of component bands and their "real" positions were not known, the number of bands and peaks at the positions of ribonuclease were based on previous studies on proteins and polypeptides with known conformations. The spectra were then resolved into Gaussian components with the aid of a computer program. Thus, even very complex spectra can be deconvoluted statistically into simples spectra for further analyses.