Mass spectrometry-based proteomics in the life sciences: a review

Kambiz Gilany[1], Luc Moens[2], Sylvia Dewilde[2]

1 Avicenna Research Institute, Tehran, Iran
[2] Department of Biomedical Sciences, Laboratory of Protein Chemistry, University of Antwerp, Universiteitsplein 1, B-2610 Antwerp, Belgium

*Corresponding author: e-mail address: kambiz.gilany@hotmail.com

ABSTRACT

Proteomics concerns itself with the characterization and function of all cellular proteins, the ultimate determinants of cellular function. Mass spectrometry has emerged as the preferred method for in-depth characterization of the protein components of biological systems. Using mass spectrometry, key insights into the composition, regulation and function of molecular complexes and pathways have been gained. Now days, mass spectrometry-based proteomics has become an indispensable tool in the cellular and molecular life sciences. This review discusses current mass spectrometry-based proteomics technologies.

Keywords: Proteomics; Mass spectrometry

INTRODUCTION

1. Introduction to proteomics

A large number of DNA sequences from a wide range of organisms, including humans and mammals, have been produced by genomic projects. However, the genome determines only the proteins, which are produced by the cell and thus set the framework within which most of the highly complex intracellular processes takes place. In a long sequence of events, resulting in the syntheses of the proteins, the construction of the mRNA is merely the first step (Figure 1).

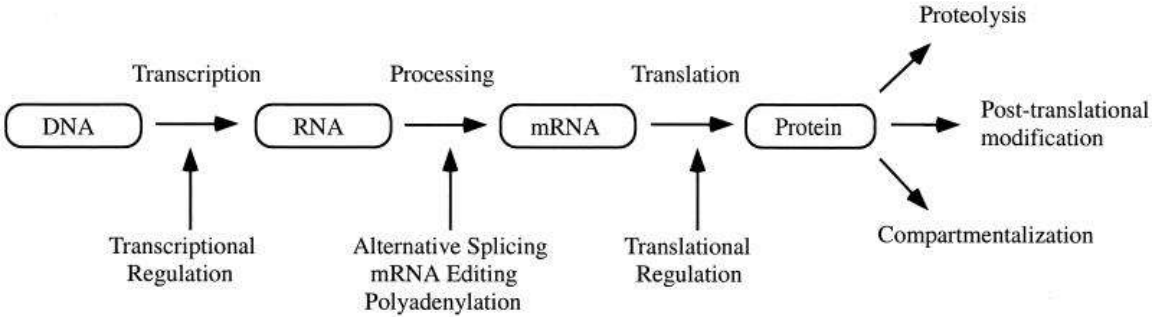

Figure 1. The mechanisms by which a single gene can give rise to multiple gene products [1]

Firstly, preRNA is exposed to posttranscriptional control in the form of polyadenylation, and mRNA editing, sometimes including alternative splicing [2]. At this step several different protein isoforms can be generated from a single gene. Secondly, the mRNAs are exposed to a regulation at the level of protein translation [3]. Those proteins, having been synthesized, are sometimes exposed to a posttranslational modification. It is believed that up to 200 different types of posttranslational protein modifications can be found [4]. The other mechanisms by which proteins are regulated, are proteolysis and compartmentalization [5, 6].

The average number of proteins produced per gene is estimated to be one to two in bacteria, three in yeast and more than 10 in humans (Table 1). Furthermore, a mRNA level has been quantified in many studies. A number of groups have produced data showing little or no correlation between the steady state protein levels and the mRNA abundance levels [7,8]. Thus, the complexity is found primarily at the protein level, not at that of the genome. The proteins, rather than the mRNA, influence the majority of the processes in the cell.

The proteins show a very broad range in concentrations, with a dynamic range in copy numbers of protein 10^5 in bacteria to 10^7-10^8 in human cells to at least 10^{12} in plasma [9,10].

Investigating the proteome of the cell gives a more representative picture of the cell at the molecular level. The proteome is defined as the time- and cell- specific protein complement of the genome and therefore encompasses all the proteins that are expressed in a cell at one time, including the isoforms and the posttranslational modifications [12]. The genome is static by nature, essentially identical in every cell of an organism, whereas the proteome is dynamic, constantly changing and responding to internal and external stimuli. The study of the proteome, called proteomics, includes the identification, characterisation and quantification of a complete set of proteins expressed by the entire genome in the lifetime of a given cell, tissue or organism. This encompasses isoforms and modifications, protein-protein interaction and the structure description of proteins and their complex [13].

Table 1. Complexity reflected in the proteomes. E. Coli, Escherichia Coli; S. Cerevisiae, Saccharomyces Cerevisiae; A. Thaliana, Arabidopsis Thaliana; D. Melanogaster , Drosophila Melanogaster [11].

Model organism	Size of genome (Mbp)	Number of gene	Number of protein per gene
E. Coli	4.6	4300	1-2×
S. Cerevisiae	12	6000	2-3×
A. Thaliana	125	25500	1-2×
D. Melanogaster	100	14000	5-10×
Homo Sapiens	3000	30000-36000	>10×

Today proteomics studies are divided into five central pillars: mass spectrometry-based proteomics, array based proteomics, structural proteomics, clinical proteomics and informatics. Protein array proteomics consists of a large number of protein interacting elements such as antibodies or peptides that are coated or immobilized on a solid support in a distribution-regulated manner. The arrayed molecules are then used to screen and assess patterns of interaction with samples containing distinct proteins or classes of proteins. In structural proteomics, different techniques, such as nuclear magnetic resonance and X-ray crystallography, are used to achieve a comprehensive coverage of individual protein structures or to analyse the structures of large protein complexes. Clinical proteomics covers a delineation of altered protein expressions, the development of a novel biomarker for diagnosis and/or early detection of diseases, and the identification of new targets for therapeutics. Informatics proteomics appears on several fronts. This field includes study designs, development of protein databases, and development of protein identification tools [14]. I will focus on mass spectrometry-based proteomics.

Mass spectrometry-based proteomics approaches are divided into two major parts: (i) one – or two-dimensional gel electrophoresis for protein separation (gel-based proteomics), followed by mass spectrometry and (ii) one- or two-dimensional liquid chromatography for protein or peptide separation (gel-free proteomics) combined with tandem mass spectrometry for protein identification.

Figure 2 gives the general workflow in mass spectrometry-based proteomics. In the following text, the fundamental principles of proteomics using mass spectrometry (MS) technology are discussed. Currently, MS has been overwhelmingly applied as the technological basis of proteomics analysis.

1.1. 2DE

In the classic proteomics, the first step is the separation and visualization of protein mixtures by a two-dimensional electrophoresis (2-DE), also called two-dimensional polyacrylamide gel electrophoresis (2-D PAGE). The 2-DE is

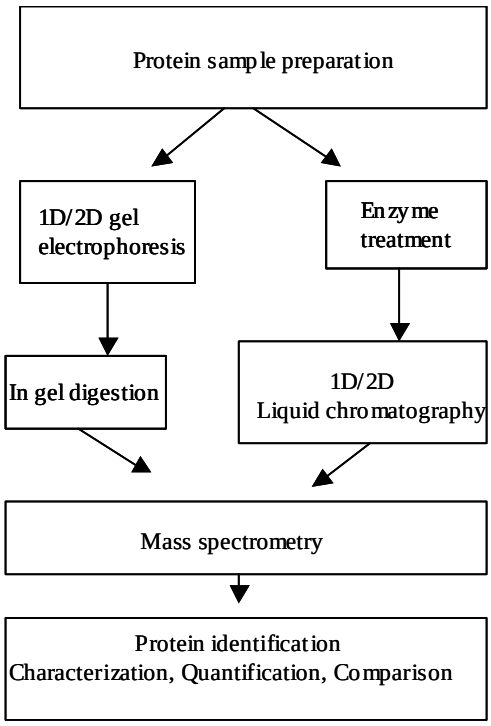

Figure 2. Workflow chart in mass spectrometry-based proteomics

not a new technique; it was first used in 1975 in the analysis of proteins.

The proteins are separated in 2-DE, based on their isoelectric points (first dimension) and their molecular weights (second dimension) (Figure 3). In the first dimension, which is called isoelectric focusing (IEF), protein separation is achieved in the context of a continuous pH gradient in which proteins migrate to the protein in the gel where they have no charge (isoelectric point). After the first dimension, a number of equilibration steps are performed to establish optimal physicochemical protein properties for the second dimension. During these equilibration steps, the proteins are treated with sodium dodecyl sulfate (SDS, for protein adsorption), a reducing agent such as dithiothreitol (DTT, to reduce disulfide bonds), and an alkylating agent such as iodacetamide (to stabilize newly formed –SH bonds). Subsequently, in the second dimension, electrophoresis is performed, usually in a polacrylamide gel (hence the term 2D-PAGE), which acts as a sieving device. All proteins move toward the positive electrode because they are covered by negatively charged SDS molecules. The amount of the SDS molecules is roughly proportional to the size of the protein, thereby allowing a separation, based on the size of the proteins [15[. The combination of IEF and SDS-PAGE in a two dimensional separation results in a resolution of 10,000 individual spots per gel [16]. In practice, depending on the sample and on the sensitivity of the detection technique, a maximum of 3000 spots can be resolved [17].

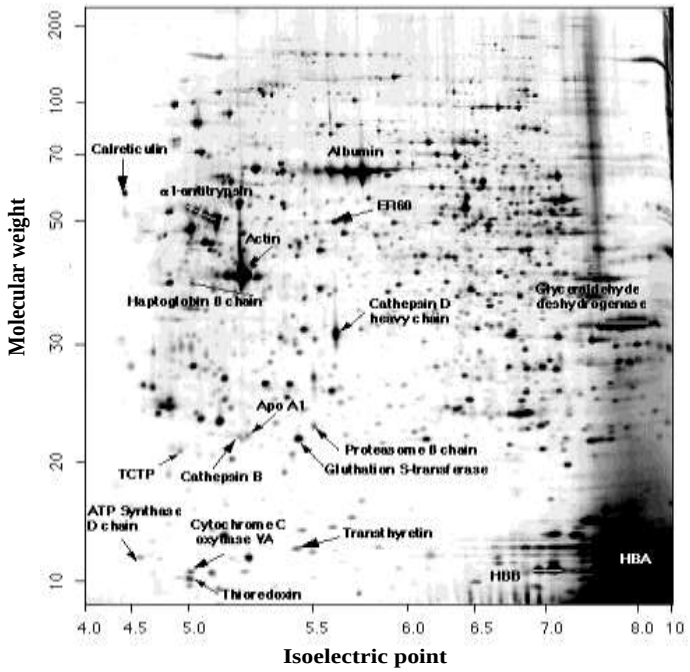

Figure 3. Two-dimensional electrophoresis (2-DE). Reference protein map of the normal human lympha (silver staining) accessible over the Internet [18].

After the 2-DE gel electrophoresis, the next step is the detection of proteins. There are several methods for the detection of proteins in 2-DE gels. Traditionally, it was accomplished by the use of a visible stain, whereas newer approaches use fluorescent dyes.

After the separation and visualization of the protein mixtures, images of gels are analyzed by software, for example Melanie or Z3. The gel images are analyzed in order to detect spot positions, calculate spot intensity and remove streaks on gels. Based on the algorithms, software packages used to analyse gel images can be divided into two groups: algorithms based on a direct comparison of images by distribution of intensity, Z3 [19], and algorithms based on spot characteristics, Melanie [20].

The 2-DE provides approximate value of mass and isoelectric points of a protein, but this information is not sufficient to identify proteins. To identify proteins of interest, gel pieces are excised; peptides are extracted after in-gel digestion, then analyzed by mass spectrometry. The two most commonly used techniques for protein characterization in the proteomics are peptide mass fingerprinting (PMF) and tandem MS of a proteolytic digest of a 2-DE spot (see 3.1).

The 2-DE is able to provide an unparalleled protein separation, but it has been shown that this technology has several limitations. These include an inability to detect membrane proteins, extreme basic and acidic proteins, proteins of a molecular weight less than 10 kDa and higher than 150 kDa and low abundance proteins [4,7,21,22]. Because of this underrepresentation of several classes of protein other methods have been developed.

1.2. Liquid chromatography techniques

An alternative technique to 2DE is the high performance liquid chromatography (HPLC/LC), a non-gel based protein/peptide separation method. LC is a long well-known method applied for the separation, identification and determination of protein/peptide components in complex mixtures. In the LC method proteins/peptides are separated based on their size (mass), pI (charge) or hydrophobicity - the three chemical characteristics that define any given protein/peptide [23]. Depending of the type of liquid and the type of stationary phase, many different modes of LC can be run [24). In this text we will describe only the most popular LC methods used in proteomics: the single-dimensional reversed-phase LC (RP-LC) and the two-dimensional strong cation exchanger-reversed phase LC (SCX-RP-LC) and the recently introduced combined fractional diagonal chromatography (COFRADIC) [21,25,26].

The separation method RP-LC is used in most cases for one-dimensional LC. The separation mechanism of RP-LC is a hydrophobic interaction between the column material coated with alkyl chains (e.g. C_4, C_8 or C_{18}) and the protein/peptide. The elution usually takes place under acidic conditions by a gradient water plus organic solvent, e.g. acetonitrile. The RP-LC has two advantages; firstly, the relative high peak capacity (maximum number of components that can be resolved) and secondly, the full MS compatibility, which allows the MS analysis of the column eluate without further treatment [27,28].

In spite of RP-LC having been proved to be an economic and effective way for protein and peptide identification, its use in proteomics is relatively restricted by the complexity of the samples. Samples in proteomic analyses often contain thousands of proteins. After proteolytic digestion, the hundreds of thousands peptides must be separated. This is beyond the analytical range of the RP-LC method because of its insufficient peak capacity. Thus, multidimensional separations are often necessary [29,30]. In multidimensional chromatography different separation mechanisms, the so-called orthogonal separation must be combined [31]. The most commonly used method in proteomics is the strong cation exchanger (SCX) chromatography (separation by positive charge) combined with the RP-LC [32]. This two-dimensional LC is better known as multidimensional protein identification technology (MudPIT) [33]. A benefit of the SCX-RP combination is that the salt ions in the SCX fractions, which would otherwise interfere with the MS analysis, are extracted from the peptide ions in the RP step [34]. The MudPIT can be performed off-line or on-line. The off-line is defined as LC, not

coupled to MS, in contrast to the on-line LC, which is connected to the MS. The advantage of the on-line method is that everything can be done automatically and sample loss will be avoided [35]. The advantage of the off-line over the on-line is an increased loading capacity, an improved chromatography, a greater flexibility and repeated sample analyses [36]. A major disadvantage of MudPIT still is the increased complexity. In a typical setup, the protein mixtures are digested prior to separation. A small calculation of the numbers of peptides shows that if a cell has between 23.000-40.000 proteins and a digest protein gives typical up to 40 peptides, the MudPIT has to be able to resolve 9×10^5-2×10^6 peptides [5].

Another peptide-based protein identification technique different from MudPIT is the combined fractional diagonal chromatography (COFRADIC), which uses RP-LC [37]. In COFRADIC, the proteins are first digested to peptides. A subset of peptides, which is highly representative of the parent protein originally present in the lysate, is then selected. COFRADIC thus reduces the complexity of the peptide mixture. Theoretically, any peptide carrying a group that can be specifically modified may in principle be selected. It is sensitive and is characterized by a broad protein coverage, including abundant and rare, large and small, acidic and basic and hydrophobic proteins. This concept has been applied to select methionine containing peptides and N-terminal peptides [38,39].

2. Introduction to MS

One of the most powerful modern analytical methods available to the laboratory analyst is the MS. Over the last decade, MS-based proteomics has rapidly become the preferred analytical method for the identification and characterization of proteins. MS makes it possible for the compounds to be identified by the production of ions and their subsequent separation and detection based on their mass-to-charge ratio, offering a very high level of specificity with sensitivity. A typical mass spectrometer consist of an ion source, which generates the ions, a mass analyser, where these ions are separated and a detector, which delivers the mass spectrum (Figure 4) [40].

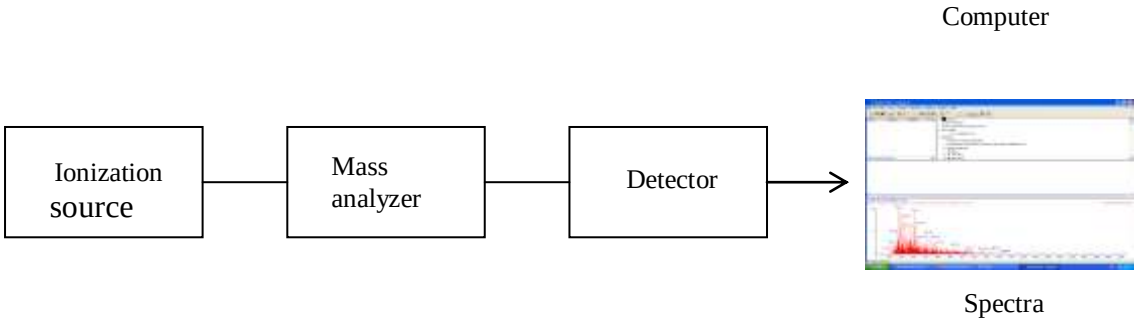

Figure 4. The principle components of a mass spectrometer

In the following text the fundamental principles of two of the most common types of mass spectrometers, MALDI-TOF and ESI-quadrupole MS, are described.

2.1. Ionization source

2.1.1. Matrix assisted laser desorption / ionization (MALDI)

MALDI is a soft ionization technique used in mass spectrometry [41]. Ionization of the molecules to be analyzed (analyte) by MALDI is obtained in two steps. In the first step, the analyte molecules are mixed with a matrix (a small, organic, and UV-absorbing molecule) and is pipetted on a sample probe (the target) and allowed to dry before analysis. In the second step, the resulting solid on the targets is irradiated by laser pulses, usually from nitrogen lasers with a wavelength of 337nm, in order to generate analyte ions (protonated molecules). The ionisation process in MALDI is still unknown [42], but it has been suggested that the irradiation induces an accumulation of a large amount of energy in the condensed phase, through an excitation of the matrix molecules.

This causes a desorption of the ions formed by a proton transfer from the photoexcited matrix to the analyte molecules (Figure 5) [43].
Process followed by a desolvation and then an introduction into the mass analyser. Figure adapted from [43].

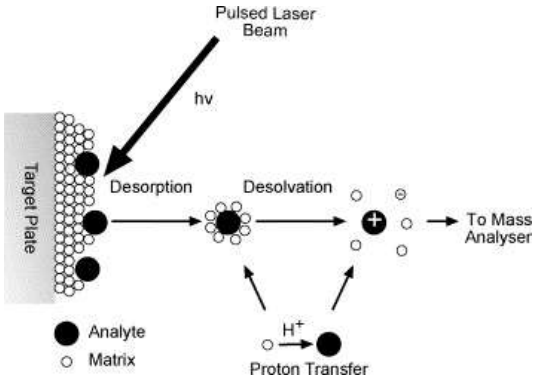

Figure 5. Ionization of analytes by MALDI. The co-crystal of matrix and sample is targeted by a laser, causing a desorption

The matrix has two other functions: it protects the analyte from the laser pulse by absorbing the induced laser energy and it changes the energy transfer from the laser to the analyte molecules [43]. In large scale proteomics, the α-cyano-4-hydroxycinnamic acid (HCCA) matrix is recommended. The choice of matrix, however, depends on the properties of the analyte molecule and the sample preparation [44]. MALDI rapidly became popular because the time-of-flight (see later) mass analyzer most commonly used with MALDI is robust, simple, sensitive and has a large mass range. Another reason for MALDI's popularity is that its mass spectra are simpler to interpret because the ions predominantly contain only a single charge (i.e. the mass of neutral analyte plus proton- $[M+H]^+$). Additionally, MALDI has been observed to be less resistant to interference from the non-peptidic components of a sample, such as salts and detergents that always accompany peptides into the mass spectrometer [43]

2.1.2. Electrospray ionization (ESI)

Even though MALDI is the most efficient method for ionizing peptides, ESI is the optimal method of ionization for the broadest range of polar biomolecules. The ESI process for forming gas-phase ions come from the work of Dole et al, but it was Fenn´s group at Yale University that coupled ESI with mass spectrometry [45,46]. Electrospray is a rather simple technique for the ionization of polar molecules. Figure 6 illustrates the electrospray process in the positive-ion mode and a photograph of the spray produced. Applying a high electrical field to a relatively small flow of liquid from a capillary tube makes an electrospray. The electrospray makes the liquid surface highly charged and a spray of charged liquid droplets emerge at the end of the capillary tube. The polarity of the charged droplets can be changed by the applied polarity on the capillary. Through the capillary tube to which a high voltage is applied, the sample solution flows at a low flow rate. Presuming positive potential, positive ions from the solution will accumulate at the surface of the tip. The positive ions are drawn out to establish a 'Taylor cone' (Figure 6, A and B). As the liquid is made to hold more electric charge, the Taylor cone makes a filament form that, when the surface tension exceed beyond the applied electrostatic force, makes positively charged droplets through a budding process. The charged droplets will evaporate in the area between the needle and the collector (the counter electrode). The electrical charge density at the surface of the droplets increases as the size of the droplet decreases. This increases the field around the droplet to the point where the electrostatic repulsion is greater than the surface tension (known as the Rayleight limit); it erupts and emits smaller particles. Continuous decrease of the droplet size will result in the creation of a droplet containing a single (or multi-charged) ion. This mechanism is better known as the charge residue model. A second mechanism (the ion-evaporation model) of the gas-phase ion production has been suggested in which the ion evaporation is believed to occur from small, highly charged droplets, with the ion formation taking place as a result of the repulsion between the charged ion and the other charges of the droplet. The relative usefulness of the two models remains the subject of discussion and research (Figure 7) [46,49] .

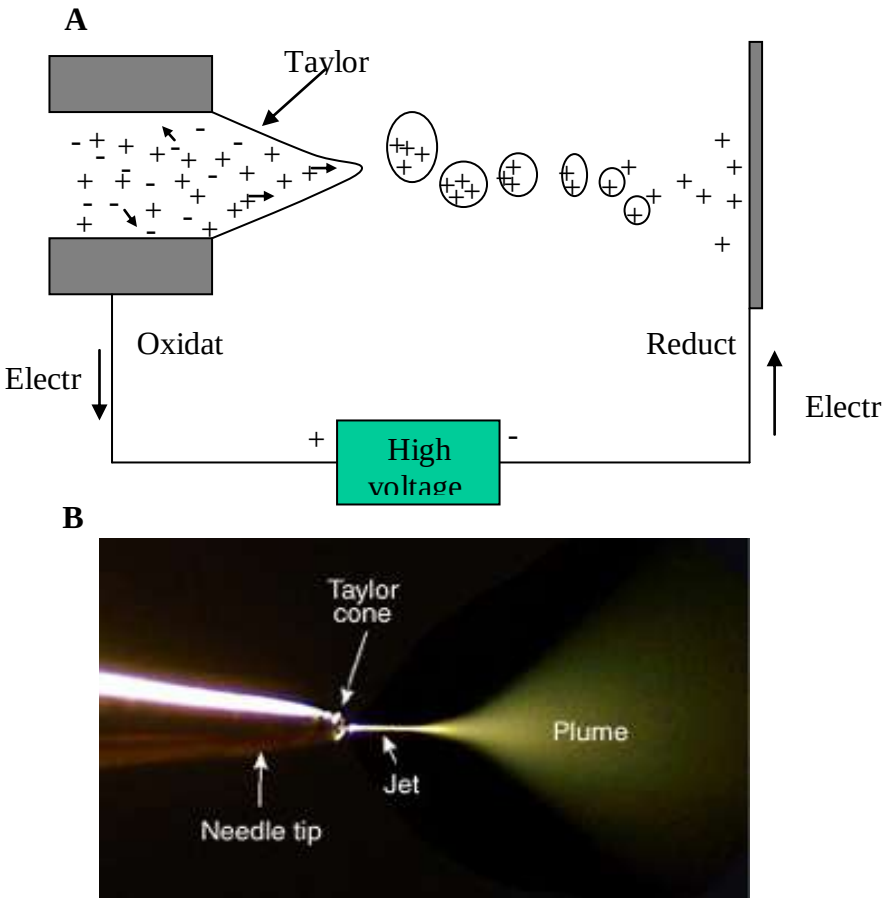

Figure 6. Principle of electro spray ionisation, **A)** schematic representation of electro spray in the positive mode and **B)** a photograph of electro spray [47,48]

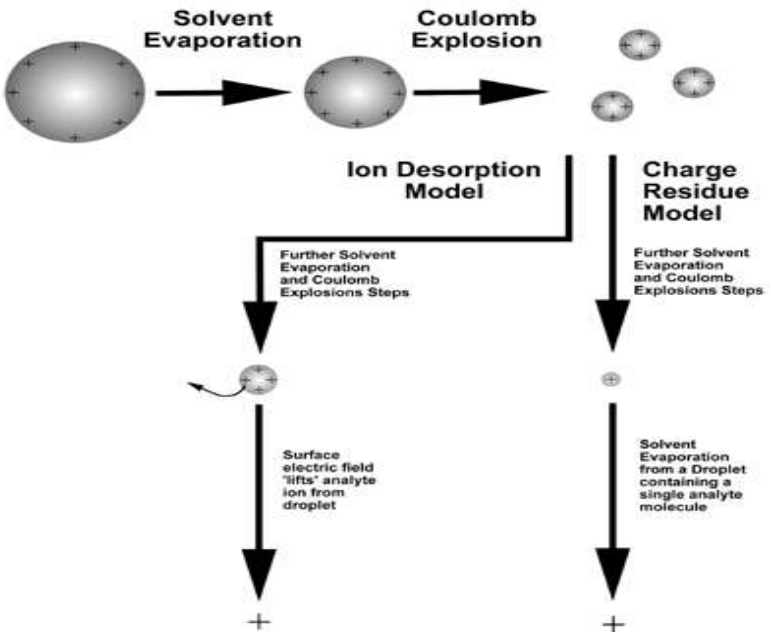

Figure 7. The mechanisms suggested behind electro spray ionization method. Adapted from [50]

The ESI technique is capable of making intact gas-phase ionic species by attaching more than one charge to each individual molecule. Peptides and proteins are thought to be protonated predominantly at the basic charge sites: the amino terminus, arginyl, histidyl and lysyl residues [45]. The resulting mass spectrum of a sample with multiple charges shows as a series of ions, representing the distribution of charge states (Figure 8).

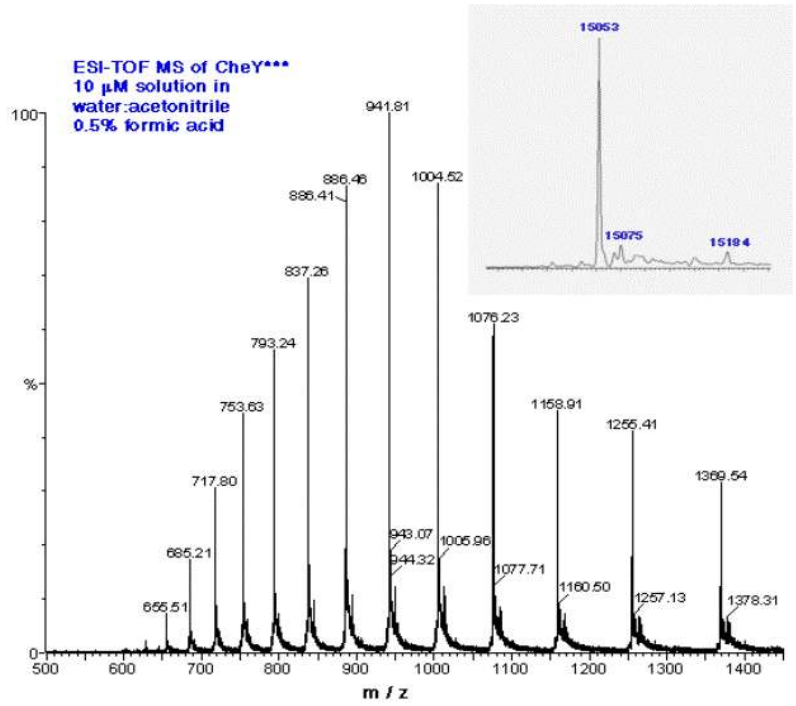

Figure 8. Multi-charge spectrum of CheY protein. Inset shows the molecular mass of the CheY interpret by a computer software (15052 Da) [51].

A protein ESI mass spectrum seems rather complex at first sight; peaks from a given protein, however, only differ by the degree of protonation where nearest neighbours differ by a single charge. Each peak shows a measure of the protein's mass, and so in each spectrum, multiple mass measurements are made, thereby enhancing the precision of the measurement.
For two successive peaks at position x_1 and x_2 on the m/z scale and corresponding to numbers of the same ion series, then:

$x_1 = (M+n)/n$

where M is the molecular mass and n is the number of charges which is equivalent to the number of added protons, and

$x_2 = (M+n+1)/(n+1)$

Solving these equation gives:

$n = (x_2-1)/(x_1-x_2)$

The molecular weight is obtained by taking n as the nearest integer value [52].
In reality the mass-spectometer data system performs such calculations and plots a molecular mass spectrum on a zero-charge-state x-axis (Figure 8).

2.2. Mass analyzer

2.2.1. The Quadrupole mass filter

The quadrupole consist of two pairs of metallic rods. The rods are perfectly parallel. It is controlled by a combination of direct current (DC) and radio frequency (RF) voltage. One set of rods is at a positive electrical potential and the other one at a negative potential. Ions flies in a continuous beam along the central axis between the poles and are filtered on the basis of their m/z ratios. The positive pair of rods is functioning as a high mass filter, the other pair is functioning as a low mass filter. Combining both sets of rods into a quadrupole arrangement, which overlaps the two mass filter regions, forms a stability area. This will allow ions with a certain m/z

ratio to pass through. Those ions, which have an unstable trajectory through the quadrupole mass analyzer will collide with the quadruple rods, never reaching the detector (Figure 9). The m/z ratio of ions that pass through the quadrupole is proportional to the voltage applied to the rods. In practice, the highest detectable m/z ratio is about 4000 Th. Scanning a quadrupole mass analyzer involves ramping the amplitude of the DC and RF voltage at a constant ratio. This will change the position of the stable area and allow different masses to pass through [43,46,49].

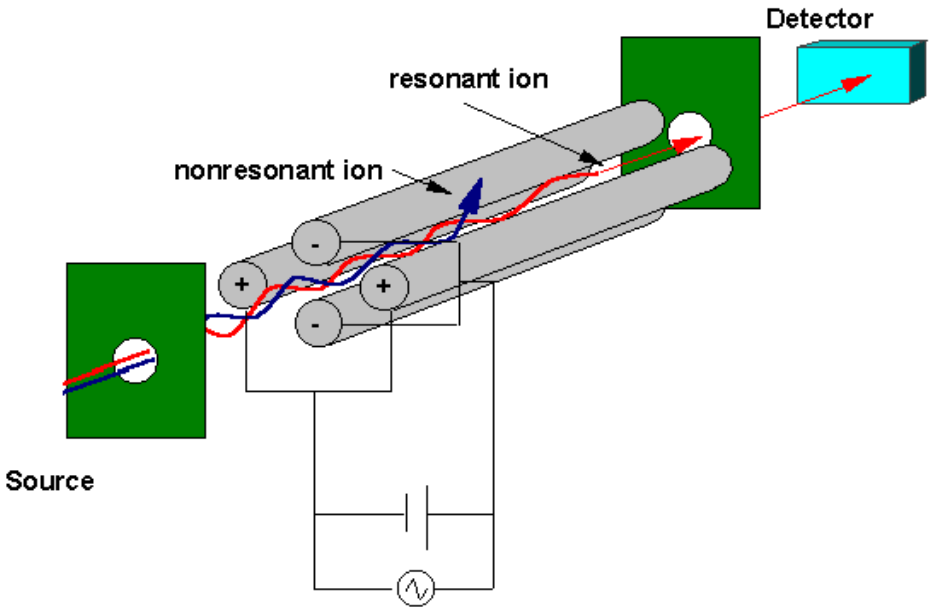

Figure 9. A schematic presentation of the quadrupole mass filter. Figure adapted from [53]

2.2.2. The time-of-flight (TOF) mass analyzer

The principle of the TOF mass analyzer is to measure the flight time of ions accelerated out of an ion source into a field free drift tube to a detector. The ions are accelerated by potential V_s and fly a distance *d* before reaching the detector. As an ion leaves the source with mass m and a total charge q = ze it has a kinetic energy:

$$Mv^2/2=qV_s=zeV_s=E_k$$

But velocity (v) = distance (d)/ time (t) and therefore the equation can be rewritten:

$$t^2 = m/z\ (d^2 / 2V_s e)$$

The equation shows that ions with different mass will take different amounts of time to travel the same distance. It also shows that the lower the mass of an ion, the faster it will reach the detector.

The most important drawback of the TOF analyzer is its poor mass resolution. Mass resolution is affected by slight variations in flight time, and factors that create a distribution in flight times among ions with the same m/z ratio will result in a poor mass resolution.

Two techniques have been introduced in TOF mass analyzers in order to improve the mass resolution. These are better known as delayed extraction and the use of an ion mirror or reflecton. In the former, a time delay between ion formation and extraction of ions from the source is introduced. Because of the delay, wide spatial and temporal distributions can be avoided [54]. In the latter technique, the reflecton creates a retarding field that deflects the ions, sending them back through the flight tube. The more energetic the ion, the deeper it penetrates the retarding field of the reflectron before being reflected. Thus a more energetic ion will travel a longer flight path and arrive at the detector at the same time as less energetic ions of the same mass [41].

2.3. Coupling Ion source to mass analyzers

MALDI is traditionally coupled to a TOF analyzer, which measures the mass of the intact peptides and proteins, while the ESI is mostly coupled to triple quadrupoles,

which allow a generation of fragment ion spectra from selected precursor ions. In the last few years, the MALDI ion source has been coupled to the TOF-TOF analyzer [55] in which two TOF sections are separated by a collision cell and Quadrupole-TOF (QTOF) [56] analysers in which a collision cell is placed between a quadrupole mass filter and a TOF analyzer. In addition, the QTOF is interchangeable with an ESI ion source [57].

3. Protein identification

The MS-based approaches are now the method of choice for most protein identification because of improvements in genomics and protein databases (Tabel 2) and the more powerful computational searching techniques. Protein identification using MS involves the use of one of two established methodologies: peptide mass mapping (or fingerprinting) and peptide sequencing. Both techniques make use of a proteolytic enzyme (typically trypsin) to specifically cleave the proteins into peptide fragments of a suitable length (mass) for the mass spectrometric analysis. The following text is a more detailed discussion of these methods.

Table 2. Databases used for proteomics searches.

Name	Content	Web address
Swiss-Prot	Annotated protein database	www.expasey.org/sprot/
MS protein sequence database (MSDB)	A comprehensive, nonidentical protein sequence database maintained by the proteomics department at the Hammersmith Campus of Imperial College London	ftp://ftp.ebi.ac.uk/pub/databases/Mass SpecDB/
dbEST	The division of GenBank that contains "single-pass" cDNA sequences, or Expressed Sequence Tags, from a number of organisms.	www.ncbi.nlm.nih.gov/dbEST
RefSeq	Comprehensive, integrated, nonredundant set of sequences including protein products, for major research organisms	www.ncbi.nlm.nih.gov/RefSeq
NCBInr	A comprehensive, non-identical protein database maintained by NCBI for use with their search tools BLAST and Entrez. The entries have been compiled from GenBank CDS translations, PIR, SWISS-PROT, PRF, and PDB.	ftp://ftp.ncbi.nih.gov/blast/db/FASTA/nr.gz
IPI	International Protein Index a non-redundant human proteome set constructed from UniProtKB/Swiss-Prot, UniProtKB/TrEMBL, Ensembl and RefSeq.	http://www.ebi.ac.uk/Databases/protein.html
TrEMBL	A computer generated protein database enriched with automated classification and annotation. Part of the UniProtKB.	http://www.ebi.ac.uk/Databases/protein.html

3.1. Peptide mass fingerprinting (PMF)

This first type of data applies a peptide mass fingerprint (PMF) or a map of the peptides derived from digestion of the protein (Figure 10). When a protein is treated with a site-specific protease like trypsin, this results in a peptide mixture. The molecular weights of this set of peptides are then measured and compared to the predicted masses of peptides from the theoretical digestion of proteins in the database, when cleaved with the same protease [58-62].When enough peptides of the measured mass and the theoretical one overlap, this generally indicates that the protein has been identified. This is the case even if a few peptide molecular weights may not match because of modifications.
Different mass search programs are available,

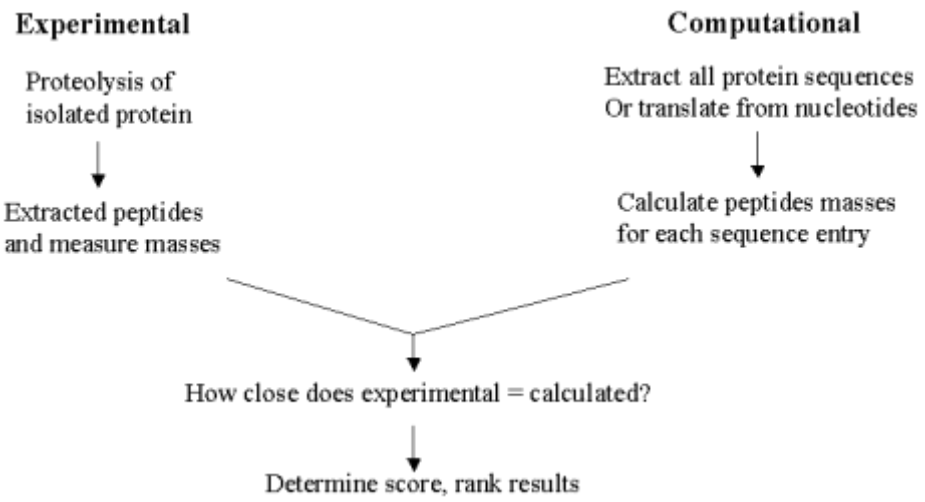

Figure 10. Chart of the peptide mass fingerprinting process. The figure is adapted from [63]

e.g. Mascot and ProFound [64]. Table 3 gives the names and URLs of a number of PMF tools.

The following information is necessary for a protein search: the protein mass range, the cleavage reagent, the peptide mass accuracy, the peptide masses, the modification during the gel process and the number of missed cleavages [65].

PMF is a suitable method for protein identification under two conditions. First, the method should only be applied when the sample to be analyzed contains a purified protein (typically, a 2-DE gel spot). The method is not reliable for mixtures containing more than two or three proteins. Second, the method is effective only when a species is well represented in the sequence databases. If these conditions are met, then the method can be an efficient manner to proceed [66-68]. However, the method can fail to identify the protein of interest if: 1) there are not enough peptide masses, 2) the peptide is posttranslational modified, 3) there are too many artifactual peptides in the spectrum, such as keratin peaks [69].

Table 3. Available PMF tools.

Software	**Query Website**
Aldente	www.expasy.org/tools/aldente
Mascot	www.matrixscience.com
MS-Fit	http://prospector.ucsf.edu/
PepFrag	http://prowl.rockefeller.edu/prowl/pepfrag.html
PepMAPPER	http://wolf.bms.umist.ac.uk/mapper/
Peptidesearch	http://www.narrador.embl-heidelberg.de/GroupPages/Homepage.html
ProFound	http://prowl.rockefeller.edu/prowl-cgi/profound.exe

3.2. Peptide sequencing

Protein identification by tandem mass spectrometry (MS/MS) is a key to most proteomics projects (Figure 11). The MS/MS spectra represent peptides that are produced by proteolysis of a protein prior to MS/MS analysis. The masses of the fragments are measured and reported as a raw spectrum, after isolation and fragmentation of one peptide type.

Peptide sequencing involves the production of fragment ion spectra by tandem MS. Peptide ions undergo fragmentation along the peptide backbone when they are introduced into the collision chamber of MS where they interact with the collision gas (usually nitrogen or argon). A nomenclature has been created to indicate what types of ions have been generated, since peptides can undergo multiple types of fragmentation. There are three different types of bonds that can be fragmented along

the amino acid backbone: the NH-CH, CH-CO and CO-NH. Each bond breakage gives rise to two species, one neutral and the other one charged, and only the charged species will be monitored by the mass spectrometer.

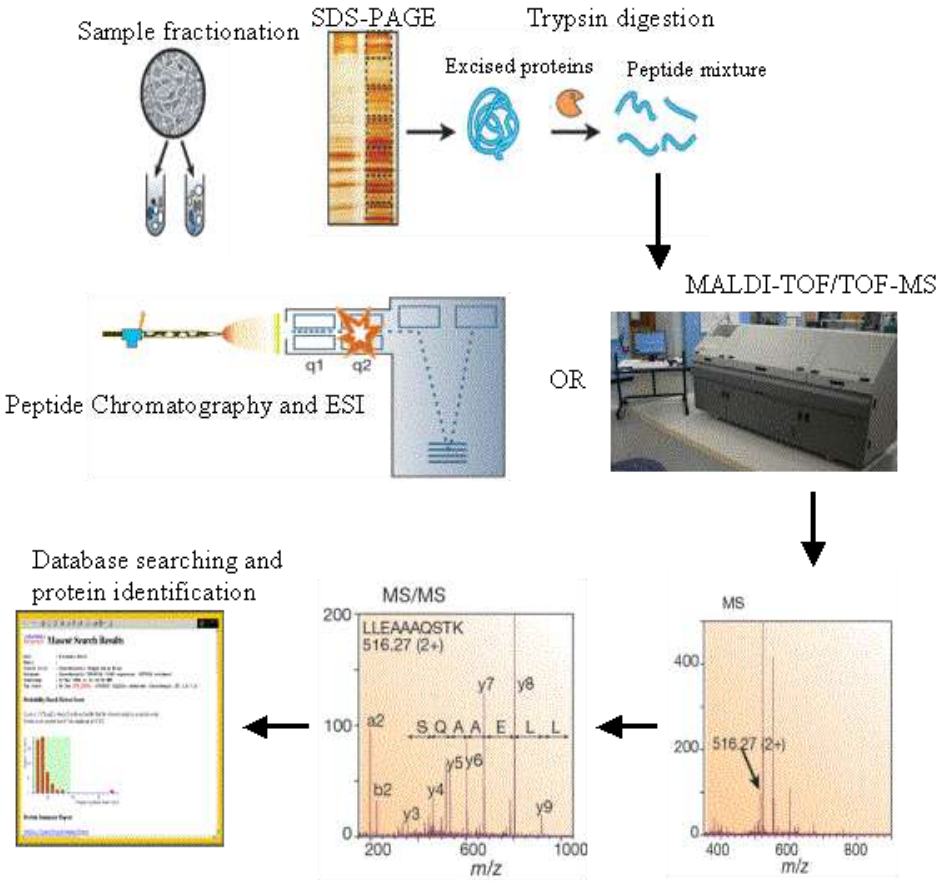

Figure 11. A typical MS/MS-based proteomics experiment [40]

The charge remains on either of the two fragments, depending on the chemistry and relative proton affinity of the two species. Hence there are six possible fragment ions for each amino acid residue and these are labeled as in Figure 12, with the a, b and c ions having the charge retained on the N-terminal fragment and the x, y and z ions having the charge retained on the C-terminal fragment. The most common cleavage sites are at the CO-NH bonds, which gives rise to the b and y ions. The mass difference between two adjacent b or y ions is indicative of a particular amino acid [70]. In addition to the fragmentation along the peptide backbone, cleavage occurs along amino acid side chains [71]. The types of ions detected in an MS/MS experiment vary with the peptide, the activation step, the instrument's observation time frame, and/or instrument discrimination factors [72,73].

MS/MS identification consists in correlating the sequence of peptides, present in a sample with their corresponding theoretical amino acid sequences, obtained from a protein or genomic database. When the amino acid sequence of a peptide is identified, it is used to search databases to find the protein from which it was derived. The MS/MS database matching, however, works optimal with peptides that match exactly the correct sequence in a database.

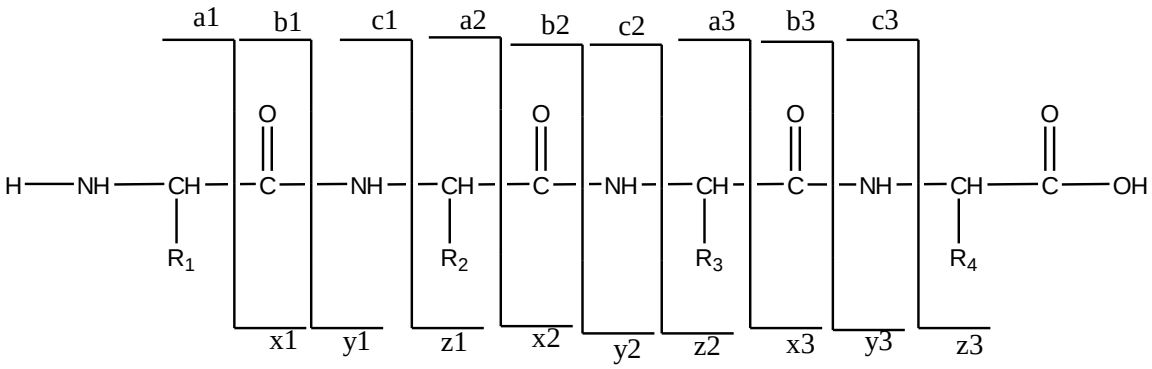

Figure 12. Fragmentation nomenclature of peptides [70].

There are several reasons why a spectrum might not match a sequence entry: if the peptide has undergone a non-specific proteolytic cleavage, if the peptide contains chemical in case of pre-, co- or post-translation modifications, in case of sequence polymorphisms, and finally if there are errors in the database sequences. In addition, other reasons may be a pitfall for identification algorithms: a spectrum may originate from a non-peptide contaminant or may be too noisy; it may originate from multiple peptides with the same m/z; the precursor mass may be incorrectly interpreted (e.g. because of incorrect precursor charge assignment); unusual fragmentation patterns may disturb the identification algorithms, for example due to the non-availability of a mobile proton. The major disadvantage of performing MS/MS is that the process is not easily automated. As a result, analysis and interpretation of the mass spectra are time-consuming processes.

4. Quantitative proteomics

One of the goals of proteomics research is to quantify proteins in the biological system. For practical purposes this field of proteomics is better known as expression proteomics. In this approach, the entire proteome or subproteome of samples from different biological states, e.g. normal versus disease, are compared [74]. There are different approaches for quantitative proteomics (Figure 13). In the following text the most developed quantification methods will be described.

4.1. 2-DE

The technique that is still the most widely used in quantitative analysis is the 2-DE, as explained in point 1.1. This technology has gained popularity for its unparalleled separation performance and its ability to provide relative protein quantification via visible staining reagents or fluorescent dyes. The most commonly used visible stain has traditionally been the silver stain or Coomasie blue. The silver staining is the more sensitive stain, detecting proteins in the 2-5 ng range, compared to the Coomassie staining, where the detected protein is over 40 ng [75]. However, the silver staining has decreased in popularity due to complications with background and reproducibility. The Coomasie blue staining is rather insensitive and a large number of proteins remain undetected using this method [76]. Because of these limitations, fluorescent dyes have been developed. This method relies on a fluorescence prelabelling of the protein mixtures before the 2-DE gel electrophoresis. The protein samples are labelled with up to three spectrally distinct, charged and mass-matched fluorescent dyes, known as CyDye DIGE fluors.

The labeled proteins then are mixed and then co-electrophoresed on the same 2-DE gel. The different protein extracts labeled with different CyDye DIGE fluors may then be visualized separately by exciting the different dyes at their specific excitation wavelengths. This is achieved by the use of an imager containing the appropriate laser wavelength for exciting the different dyes and a filter for collecting the light emitted. Each dye generates digital images of each individual sample. The advantages of this technique are its ability to overcome the reproducibility problem in the 2-DE and an improved sensitivity and accuracy. The sensitivity of the CyDye DIGE fluor minimal dyes are: Cy2 0.075ng, Cy3 0.025ng and Cy5 0.025ng.

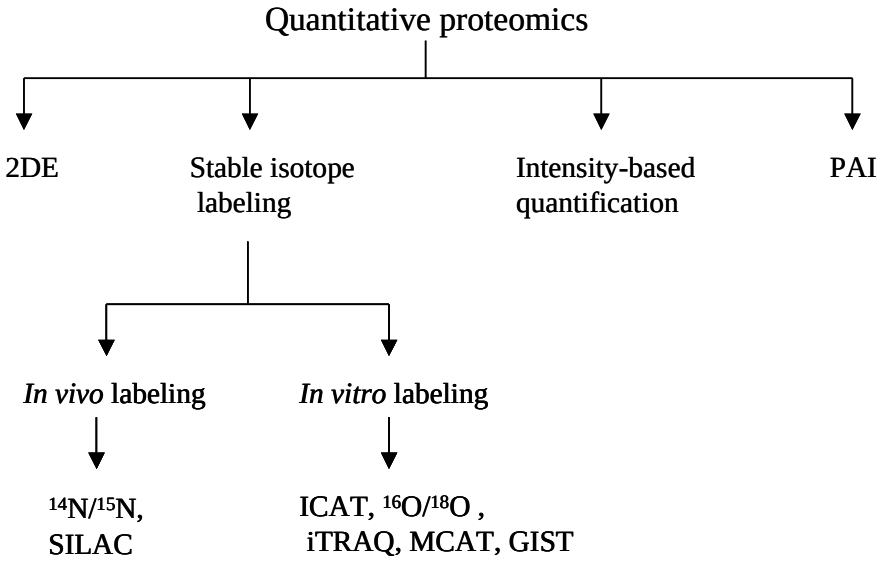

Figure 13. Strategies for quantitative proteomics profiling

The major drawback of this staining technique is that it is a property of GE Healthcare and requires expensive and specific equipment, such as a three-laser fluorescent scanner and dedicated software [77-79].

However, because of the limitations of the 2-DE, such as the existence of multiple gene products in one spot which make quantification difficult, over the past few years protein quantification approaches have rapidly evolved towards methods providing a higher throughput, a wide sample applicability and a smaller sample requirement [80].

4.2. Stable Isotope labeling

The alternative techniques that have been developed are mass spectrometry based on stable isotope labeling (Figure 13). Quantification in MS-based proteomics, using stable isotope dilution, makes use of the fact that the two analytes that differ only in stable isotope composition will be almost chemically identical, yet can be differentiated in a mass spectrometer, due to their mass difference. The ratio of the ion currents for the analyte pair will be equal to the abundance ratio for the two analytes. Therefore, a relative abundance of proteins in different cell or tissue states can be calculated, and an absolute abundance can be measured with the use of isotopically labeled standards [46,47,81,82].

The mass spectrometry based on the stable isotope labelling strategy may be divided into two classes (1) stable isotope labelling of proteins that is achieved metabolically *in vivo*; (2) stable isotope labelling that is achieved chemically or enzymatically *in vitro*.

4.2.1. *In vivo* labelling

Metabolic labelling is a method adapted for quantitative proteomics. Nutrients highly enriched with stable isotopes are provided to cells in culture, and isotope labels are incorporated into all cellular proteins simultaneously.

4.2.1.1. ^{15}N

The first approach developed for *in vivo* stable isotope labelling, utilizes media containing ^{15}N. In this procedure, cell cultures are grown in two separate media, one containing ^{15}N, and the other containing standard ^{14}N. The cells are then pooled together; proteins extracted, digested to peptides and quantified on ESI-LC/MS. Cells grown in media enriched with ^{15}N are used as an internal standard for all quantitative measurements. These internal standards are mixed with cells from different conditions early during the sample preparation so that any protein loss during cell lysis, digestion and measurement are accounted for by their respective ^{15}N-labelled proteins. The MS measurement readily differentiates between the resultant peptides from the two types of media. The ratios between the intensities or the areas of

the isotopically labelled peptide pairs directly reflect the difference in the amount of a given protein in two different cell pools [83-85].

4.2.1.2. Select isotopic labelling amino acid incorporation (SILAC)

Several researchers have described the use of selected stable isotope-incorporation for quantifying protein/peptide levels between two given cell lines. In this approach, the stable isotope labelling strategy termed SILAC (Stable Isotope Labelling by Amino Acids in Cell Culture), two groups of cells are grown in culture media where one contains the essential "light" amino acids and the other the essential "heavy" amino acid (e.g., L-arginine and ^{13}C-labelled L-arginine). By using essential amino acids, the cells are forced to use them and consequently incorporate them into all proteins as they are synthesized. Here, instead of isotopically labelling the entire pool of amino acids with ^{15}N, only the selected amino acids are labelled. The substitution of an isotopic amino acid in the growth media does not alter the cell growth, cell morphology, doubling time, or the ability to differentiate, and hence can be used as a powerful quantitative tool for proteomic analysis [86-88].

However, metabolic labelling has its limitations because it relies on proteomics samples derived from cell cultures. Other samples of interest, for example clinical tissue samples, are not amenable to this technique. Therefore other techniques are developed.

4.2.2. *In vitro* labelling

When metabolic labelling of proteins is not possible or not desirable, chemical-labelling techniques (*in vitro*) can be used as an alternative quantitative tool. *In vitro* stable isotope labelling can be applied universally for any type of sample, which is a major advantage of this technique over the *in vivo* metabolic labelling method described above. However, higher technical variations in the sample-processing step must be expected, compared to the *in vivo* metabolic labelling method. This variation appears because the two sample groups for the *in vitro* technique cannot be mixed until the isotope labelling has been accomplished, and the mixing usually takes place after the digestion of proteins. In the following text the most applied and recently developed technology will be described. It is worth mentioning that not all *in vitro* labelling techniques are well developed yet. This means that the advantages and disadvantages of some of the methods are not clear.

4.2.2.1. ICAT

Isotope coded affinity tags (ICAT) methodology is a well developed method for protein quantitation [89-91]. An example of the ICAT technology is shown in Figure 14.

The proteome extract from two different samples are covalently linked with one of two forms of the ICAT reagents, an isotopically light form in which the linker contains eight hydrogens or a heavy form in which the linker contains eight deuterium atoms. The ICAT reagent consists of three functional groups: an affinity tag (biotin) that is used to isolate ICAT-label peptides, a linker that incorporates stable isotopes and finally a reactive group, which reacts with cystein residues in the protein via a thiol group. After labelling, the samples are mixed. The combined sample is digested by a protease, usually trypsin. The ICAT-tagged peptides are selectively enriched/recovered by avidin-biotin affinity chromatography and then analysed by ESI-LC-MS/MS. Each cysteinyl peptide appears as a pair of signals, differing by the mass differential encoded in the mass tag. The relative ion intensities of peptides indicate precisely the ratio of abundance of the protein from which the peptide originates, and the MS/MS spectrum of either isotopic form of the peptide allows the protein to be identified. Similarly to other existing MS-based strategies for protein quantification, the ICAT approach has its advantages and disadvantages. The ICAT technology can be applied for the proteome extracted from bodily fluids, cells and tissues. The specificity of the ICAT reagents helps to reduce the complexity of the sample mixtures analysed during the LC phase. The drawbacks of this method include: the size of the ICAT chemical group (average MW=570.5 Da) may complicate database-searching algorithms, especially for short peptides (<7 amino acids). These

limitations would influence searches based on the LC/MS data. The method fails for Cys-free proteins or proteins that might contain posttranslationally modified Cys [93].

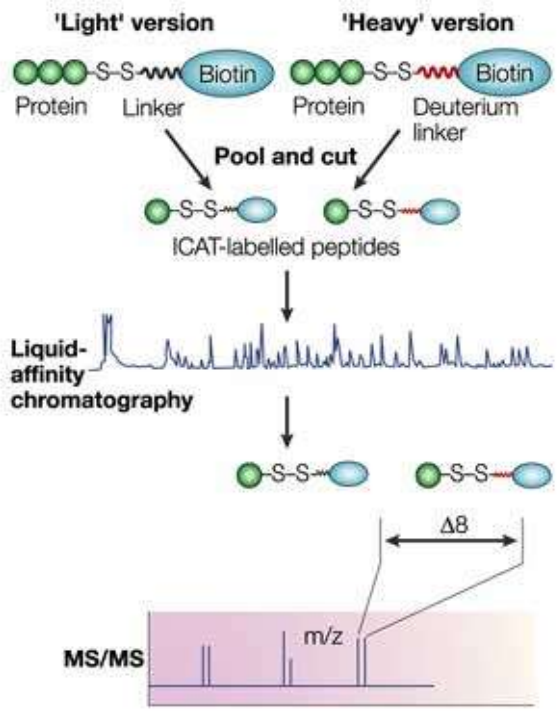

Figure 14. Two populations of proteins from different cellular states or growth conditions are isolated, and each population is tagged with a different isotope-coded affinity tag (ICAT) [92].

Recently, a second generation of ICAT reagents has become commercially available, making the technology widely available. The new version of the ICAT reagent allows a removal of an acid-cleavable bond of the biotin moiety, prior to MS and utilizes (13)C substitution for (12)C in the heavy-ICAT reagent, rather than (2)H (for (1)H) as in the original reagent [94,95].

4.2.2.2. ^{18}O labelling

A very straightforward technique for quantitative proteomics is to perform a tryptic digestion of one sample in $H_2{}^{16}O$ and the other in $H_2{}^{18}O$ [96-99]. During trypsin proteolysis, ^{16}O or ^{18}O isotopes can be incorporated into the C-termini of peptides (Figure 15). When the proteolysis of the control and experimental samples is carried out in $H_2{}^{16}O$ and $H_2{}^{18}O$ respectively, the peptides are differentially coded according to the sample origin [81]. The relative quantity of proteins is determined by the ratio of ion intensities of ^{16}O to ^{18}O-labelled peptides measured by mass spectrometry.

The advantages of the ^{18}O labelling include: the simplicity of the method and the separation of the labelling procedure from experimental conditions. Furthermore, unlike the metabolic labelling approach, the proteolytic ^{18}O-labelling technique works equally well with all samples. The method, however, requires separate proteolytical digestions of the paired proteome pools, which can lead to a decreased precision. In addition, because the mass difference between ^{16}O- and ^{18}O-labelled peptides are only 2 Da or 4 Da, the proteolytic labelling technique has a limited usefulness for larger peptides, where the isotope envelopes of the ^{16}O- and ^{18}O-labeled separated peptides overlap [25,100]. Another pitfall of the method is the possible incomplete incorporation of isotopic labels due to the use of enzyme [101].

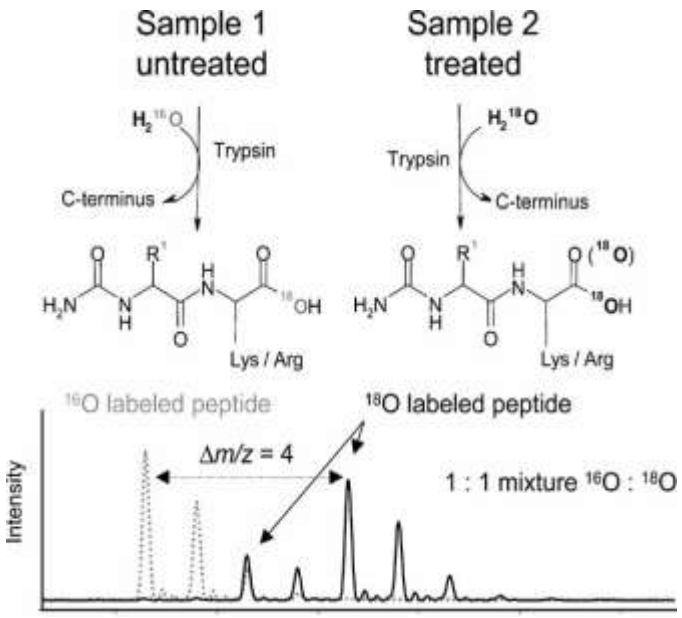

Figure 15. General scheme of the ^{18}O labelling procedure. Figure adapted from [102].

4.2.2.3. MCAT

Another technique, Mass-Coded Abundance Tagging (MCAT) can be used to compare the relative peptide abundances of two samples in complex mixtures [103]. As in the case of the ICAT, this approach is based on the principle of selective labelling of only a specific amino acid residue. In the MCAT, at a high pH, O-methylisourea is used after proteolysis of the sample by trypsin in order to selectively guanidinate the ε-amino group of the C-terminal lysine residues in the resulting peptides (Figure 16). This modification converts lysine into homoarginine, which is 42 atomic mass units heavier than lysine. The relative abundance of treated and untreated sister peptide species from the two samples can thus be estimated by measuring the ion signal intensities in a full scan LC/MS mode. The advantage of the MCAT is that it does not affect the peptide amino terminus or the other side group. Furthermore, guanidination increases the ionisation efficiency. However, the MCAT is only possible if glycine is the N-terminal residue [81,104].

4.2.2.4. Global internal standard technology (GIST)

Proteolytical cleavage of proteins generates a primary amine at the amino terminus of the peptides that, along with the primary amine on lysine residues, is easily acylated [105]. Coding by acylation is a global strategy often referred to as a global internal standard technology (GIST). The GIST protocol involves the proteolytic digestion of control and experimental samples, followed by an isotopic labelling of the resulting tryptic peptide by deuterated and nondeuterated versions of acylating agents such as ($^1H_3$)-N-acetoxysuccinimide and ($^2H_3$)-N-acetoxysuccinimide. After labelling, samples are mixed, and the peptide pair are relatively quantified by using the ion signal intensities from the LC/MS run [106].
The strength of the method lies in the fact that all peptides in the digest are universally labelled, independent of amino acid composition or post-translation modification (except from the N-terminal block amino terminus of proteins). The pitfall of GIST is that acylation reduces the charge on C-terminal lysine containing peptides, which reduces ionisation efficiency [105].

4.2.2.5. iTRAQ

In the iTRAQ technology, the proteome of four different samples are separately extracted and digested. The resulting peptides are modified by the different iTRAQ reagents, combined and analysed by LC/MS/MS.

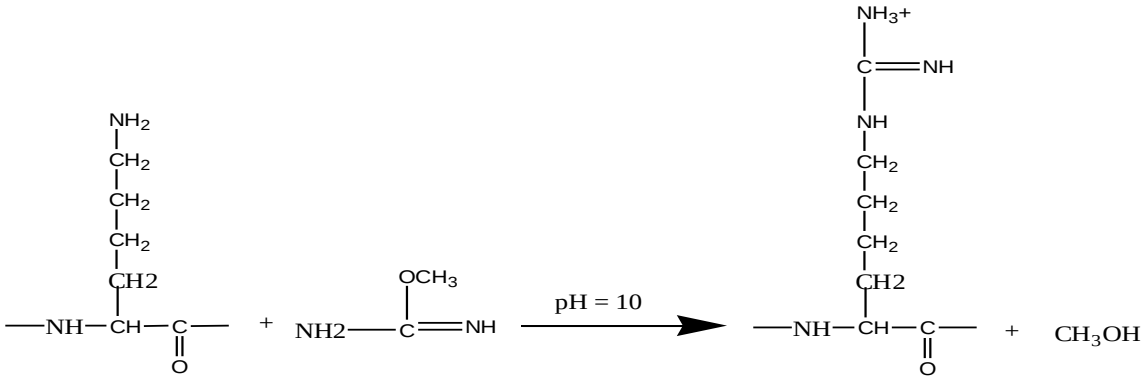

Figure 16. The guanidination reaction selectively modifies lysine

In the iTRAQ technology, peptides are labelled with isobaric reagents. These reagents contain a primary amine reactive group, which covalently binds to all peptides (N-termianl or ε-amino group of lysine) in a given sample digest, a reporter group of different mass (114,115,116 and 117 Daltons) and a balancer group (31-28 Daltons), which balances the mass of the reporter group. Due to the isobaric mass design of the iTRAQ reagents, differentially labelled peptides appear as single peaks in MS scans, thus reducing the probability of peak overlapping. The relative quantitation takes place in the MS/MS mode. When the iTRAQ tagged peptides are subjected to MS/MS analysis, the mass balancing group is released as a neutral fragment and the reporter group ions are generated. The peak areas of these ions represent the relative amount of a given peptide in the respective proteome extracted (Figure 17) [107,108]. The strength of the iTRAQ methodology is that the reporter group ions appear in the low-mass region, which is essentially free of other common ions. This reduces the error in the relative quantification [110]. An inherent drawback of the iTRAQ technology is that quantitative information is only obtained on those peptides that are subjected to the MS/MS analysis, e.g. the three most abundant peptides per MS survey scan obtained in the nano-HPLC-MS/MS runs [111]. Another disadvantage of the iTRAQ is that it suffers from the peptide overabundance problem and must be coupled with one or more dimensions of liquid chromatographic separation before the MS analysis, in order to limit the number of isobaric tagged peptides in the first MS dimension [112].

4.3. Intensity based quantification

Isotopic labelling of proteins is not always practical and has several disadvantages. For example, labelling with stable isotopes is expensive. Furthermore, there may not be enough different isotopes to allow for a simultaneous quantification of proteins from multiple samples [88]. As an alternative, several groups have presented methods of peptides and protein quantification without isotopic tags by comparing peptide signal intensities, measured in sequential ESI-MS analyses [113,114]. The signal intensity based quantification has the advantage that the observed linear correlation between peak areas of the measured peptides and their abundance [115]. The use of signal intensities for quantification is of particular interest in cases where the isotopic labelling is impractical or not feasible. It has for example been used for the quantification of protein present in multiple cellular fractions [116]. A major disadvantage of the peptide quantification by signal intensity is that it often includes experimental variation and signal to noise ratios which may affect the quantitative value and accuracy [88]. Additionally, no free user-friendly software is available.

4.4. PAI

A single LC-MS/MS analysis can easily generate a long list of identified proteins with the help of database searching, and additional information may be extracted, such as the probability score, the number of identified peptides per protein, LC retention times and so on.

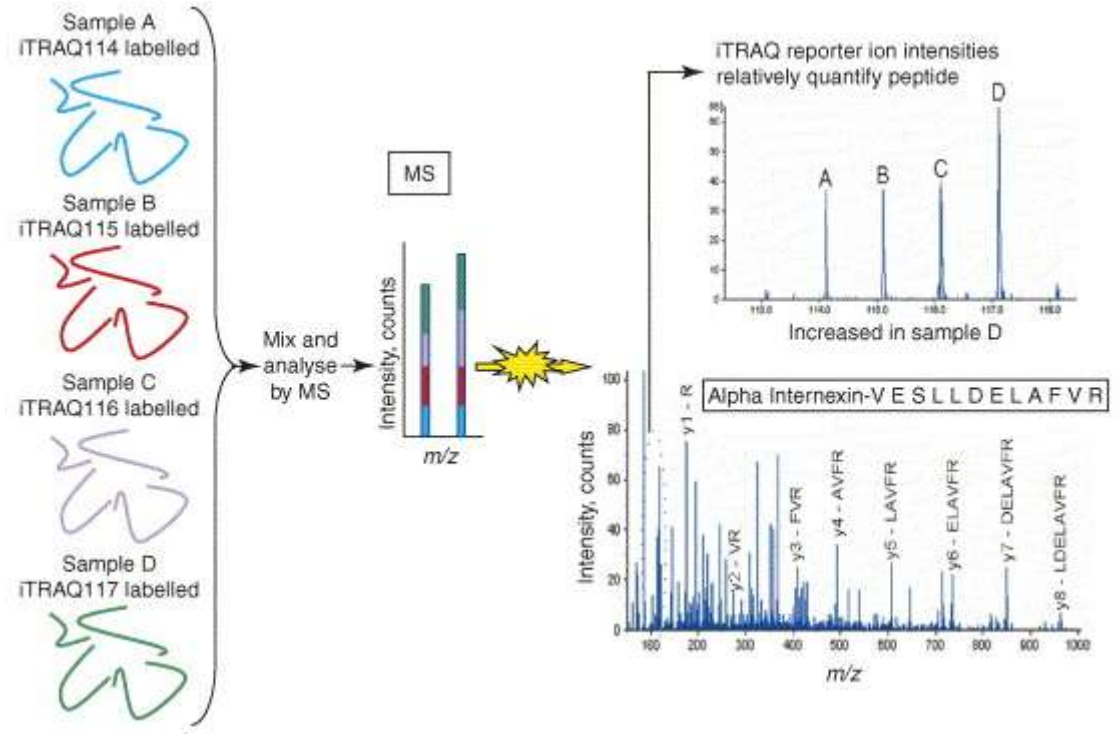

Figure 17. Strategy for relative quantification by the iTRAQ technology. Figure adapted from [109]

Some parameters, such as hit rank, score, and the number of peptides per protein, may be considered as indicators for the protein abundance in the analysed sample [117].

It has been observed that there is a linear correlation between the number of peptides sequenced per protein and the amount of protein present in the mixture. Because larger proteins can give rise to more peptides, a protein abundance index (PAI) is defined. In the PAI index, the theoretically observable peptides are used for normalization. The PAI index represents the number of identified peptides ($N_{observed}$) divided by the number of theoretically observable tryptic peptides ($N_{observable}$) [118]:

$PAI = (N_{observed}) / (N_{observable})$

More recently, a refined version of PAI has been developed [119]. The Exponentially Modified Protein Abundence Index (emPAI) shows an even better linear correlation between the number of peptides identified and protein abundance, defined as:

$emPAI = 10^{PAI} - 1$

The following example illustrates the application of the emPAI index: the protein pyruvate kinase is digested *in silico*, with one missed tryptic cleavage, and results in 76 peptides. Assuming that a nano-LC-ESI-MS/MS identify 35 unique peptides of the pyruvate kinase, the PAI and the emPAI values are:

PAI = 35/76~ 0.46 and emPAI= 1.88

Furthermore, the emPAI index can be used to express the molar fraction in a sample directly. The protein contents in molar fraction percentages are described as:

Protein content (mol %) = (emPAI /$\sum$ emPAI) × 100

$\sum$ emPAI is the summation of the emPAI values of all identified proteins of a sample.

emPAI can be used for relative quantification, especially in the cases where the isotope-based approaches can not be applied because the quantitative changes are too large for accurate measurements of ratios, or in the cases where a metabolic labeling is not possible, or in the cases where sensitivity constraint does not allow chemical labeling techniques [119].

In such cases, the emPAI values of the proteins in one sample may be compared to those in another sample, and the outliers

from the emPAI correlation between the two samples may be determined as increasing or decreasing proteins [120-122].

5. Studies of higher-order structures with MS

MS can be used to study the primary structures as well as the secondary, tertiary and quaternary structures of proteins [49,123,124]. In the following text we will describe the application of MS for disulphide bonds and protein folding.

5.1. Disulfide bond

Many proteins contain disulfide bonds between pairs of proximal cysteins (Figure 18).

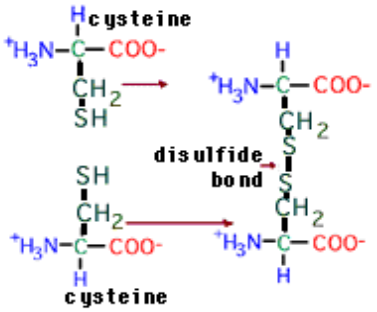

Figure 18. A disulfide bond formed by two cysteine residues

The disulfide bonds play a role in the stabilization of the tertiary and quaternary structure of proteins. For proteins it is important to characterize the disulfide-pairing pattern. Some proteins contain free cystein residues, not involved in disulfide bonds, and it is important to identify these residues. Although analytical methods have been developed to determine the thiol content in a protein, the determination of the locations of the disulfide bonds is more of a challenge. The total cysteine content can be determined by a complete reduction of the disulfide bonds with a suitable reagent (DTT or β-mercaptoethanol are commonly used), followed by alkylation with iodacetic acid or iodacetamide. If the procedure is repeated, but without prior reduction, only the free cysteins, not involved in disulfide bonds, are alkylated. It should be mentioned that some proteins may have free cysteines buried inside the tertiary structure, preventing them from being alkylated. The protein mass are then measured by mass spectrometry. The total mass increment of the intact protein after this procedure divided by the mass of the alkylating group yields the number of cysteins in the protein [125,126].

5.2. Protein folding

Proteins have evolved to carry out a wide range of specific functions, such as biomedical catalysis, transport, signalling and energy conversion. The higher order structure adopted by the polypeptide chains of a protein depends on its solvent environment. Proteins fold into a unique, highly ordered and compact structure under various physiological conditions of pH and temperature. This "native" confirmation is vital to the function of proteins. The native state of a protein corresponds to the confirmation with the lowest overall free energy [127]. Studying the way in which proteins fold is of fundamental importance for proteomics and structure genomics [13] and for de novo synthesis of proteins [128]. MS can provide information that is complementary to data obtained from more traditional techniques, such as nuclear magnetic resonance, optical spectroscopy, or calorimetry [129,130].

Electrospray ionisation mass spectrometry has often been used to study protein folding and unfolding [131]. As explained in 2.1.2, protein spectra recorded in the positive/negative ion mode typically shows a number of peaks that correspond to protein ions in different protonation states. Studies have shown that protein electrosprayed from solution conditions, that is in native confirmation tend to have a low net charge, whereas in the denaturing solution conditions (e.g., acidic, elevated temperatures, organic solvents) produce much higher charge peaks. Figure 19 shows an example of the ESI mass spectra recorded of cytochrome c protein under "native" and different pH solution conditions. As seen in the figure, protein under compact native state has a narrow charge distribution, centered around the $[M+^{8}H]^{8+}$ state, whereas in the unfolded condition (low pH) it has a broad charge distribution, centered around $[M+^{17}H]^{+17}$ state.

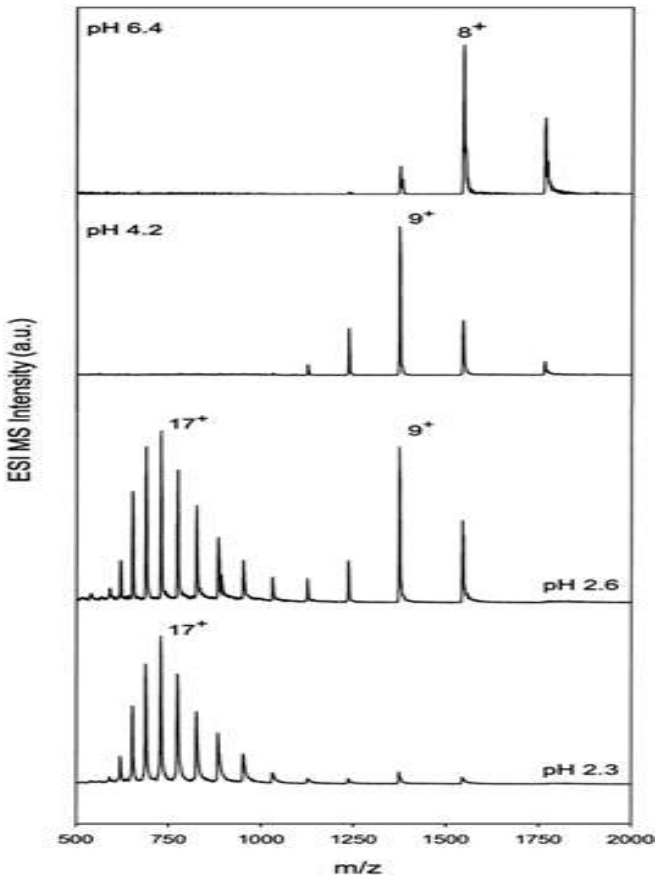

Figure 19. The various charge state distributions obtained by electrospray mass spectrometry from cytochrom c protein in water that contains 3% methanol and 0.5 mM ammonium acetate at A) pH 6.4, where the compact native state is fully populated, B) pH 4.2, C) pH 2.6, to D) pH 2.3, where the protein is completely unfolded, and hence can accommodate more positive charges. At intermediate pH values, the biomodal distribution demonstrates the co-existence of folded and unfolded state. Adapted from [131].

6. Concluding remarks

As it can be seen from the reviewed literature, any mass spectrometer having tandem mass analysis capabilities, regardless the ion source, mass analyser or separation technology to which it is interfaced, can be used for protein characterization. Specialized instruments with increased resolution and scan speed are providing and will continue to provide enhanced analytical capabilities, showing better performance in certain types of analyses. The fast developments of mass spectrometers in parallel to the advances in protein chemistry provide us with new ways for asking biological questions. It is likely that gel-free MS-based proteomics will occupy a central role in biological research, providing accurate measurements of changes in biological functions, which will be elegantly coded by researchers as changes in molecular mass.

REFERENCES

1. Graves P. R, Haystead T. A. Molecular Biologist's Guide to Proteomics. Microbiol.Mol.Biol.Rev 2002; 66: 39-63.
2. Newman A. the electrophoresis was performed with a Bio-Rad Mini Trans blotting system. Curr.Biol 1998; 8: 903-905.
3. Jansen M, de Moor C.H, Sussenbach J.S, van den Brande J.L. Translational control of gene expression. Pediatr.Res 1995; 37: 681-686.
4. Kearney P, Thibault P. J. Bioinformatics meets proteomics--bridging the gap. Bioinform. Comput.Biol 2003; 1: 183-200.
5. Rappsilber J, Mann M. What does it mean to identify a protein in proteomics? Trends Biochem.Sci 2002; 27: 74-78.
6. Colledge M, Scott J. D. AKAPs: From structure to function. Trends Cell Biol 1999; 9: 216-221.

7. Gygi S. P, Rochon Y, Franza B. R, Aebersold R. Evaluation of two-dimensional gel electrophoresis-based proteome. Mol.Cell Biol 1999; 19: 1720-1730.
8. Lian Z, Kluger Y, Greenbaum D. S, Tuck D, Gerstein M, Berliner N, Weissman S. M, Newburger P. E. Genomic and proteomic analysis of the myeloid. Blood 2002; 100: 3209-3220.
9. Sellers T. A, Yates J. R. Review of proteomics with applications to genetic epidemiology. Genet.Epidemiol 2003; 24: 83-98.
10. Anderson N. L, Anderson N. G. The Human Plasma Proteome: History, Character, and Diagnostic Prospects. Mol.Cell Proteomics 2002; 1: 845-867.
11. Kim S. I, Voshol H, van Oostrum J, Hastings T. G, Cascio M, Glucksman M. J. 2-DE proteomic profiling of Neuronal Stem Cells. Neurochem.Res 2004; 29: 1317-1331.
12. Rappsilber J, Mann M. Is mass spectrometry ready for proteome-wide protein expression analysis? Genome Biol 2002; 3, COMMENT2008.
13. Pandey A, Mann M. Proteomics to study genes and genomes. Nature 2000; 405: 837-846.
14. Tyers M, Mann M. From genomics to proteomics. Nature 2003; 422: 193-197.
15. Berkelman T, Stenstedt T. 2-D Electrophoresis - using immobilized pH gradients. Principles and methods.Amersham Biosciences 1998; 80: 6429-60.
16. Klose J, Kobalz U. Two-dimensional electrophoresis of proteins: an updated protocol and implications. Electrophoresis 1995; 16: 1034-1059.
17. Lilley K. S, Razzaq A, Dupree P. Two-dimensional gel electrophoresis: recent advances in sample .Curr.Opin.Chem.Biol 2002; 6: 46-50.
18. lympha. .SWISS-2DPAGE maps. 2008. http://expasy.org. Ref Type: Internet Communication
19. Smilansky Z. Automatic registration for images of two-dimen- sional protein gels. Electrophoresis 2001; 22: 1616-1626.
20. Appel R. D, Vargas J. R, Palagi P. M, Walther D, Hochstrasser D. F. Make2ddb: A simple package to set up a two-dimensional electrophoresis database for the World Wide Web. Electrophoresis 1997; 18: 2735-2748.
21. Mann M, Hendrickson R. C, Pandey A. Analysis of proteins and proteomes by mass spectrometry, Annu.Rev. Biochem 2001; 70: 437-473.
22. Kearney P, Thibault P. J. Bioinformatics meets proteomicsbridging the gap between mass Bioinform.Comput.Biol 2003; 1: 183-200.
23. Shen Y, Smith R. D. Proteomics based on high-efficency capillary sepearation. Electrophoresis 2002; 23: 3106-3124.
24. Meyer V.R. Practical High-Performance Liquid Chromatography, John Wiley & Sons Ltd 1994
25. Ferguson P. L, Smith R. D. Proteome analysis by mass spectrometry. Annu.Rev.Biophys.Biomol.Struct 2003; 32: 399-424.
26. Gevaert K, Van Damme P, Martens L, Vandekerckhove J. Diagonal reverse-phase chromatography applications in peptide-centric proteomics: ahead of catalogue-omics. Anal.Biochem 2005; 345: 18-29.
27. Shen Y, Zhao R, Berger S. J, Anderson G. A, Rodriguez N, Smith R. D. High-efficiency nanoscale liquid chromatography coupled on-line with mass spectrometry using nanoelectrospray ionization for proteomics. Anal.Chem 2002; 74: 4235-4249.
28. Ishihama Y. Proteomic LC-MS systems using nanoscale liquid chromatography with tandem mass spectrometry. J.Chromatogr.A 2005; 1067: 73-83.
29. Liu H, Lin D, Yates J. R. Multidimensional separations for protein/peptide analysis in the post-genomic era. Biotechniques 2002; 32: 898-902.
30. Wang H, Hanash S. Multi-dimensional liquid phase based separa- tions in proteomicsJ. Chromatogr.B Analyt. Technol.Biomed.Life Sci 2003; 787: 11-18.
31. Link A. Multidimensional peptide separations in proteomics J. Trends Biotechnol 2002; 20: 8-13.
32. Peng J. Gygi S. P. Proteomics: the move to mixtures J.Mass Spectrom 2001; 36: 1083-1091.
33. Washburn M. P, Wolters D, Yates J. R. Large-scale analysis of the yeast proteome by multidimensional. Nat.Biotechnol 2001; 19: 242-247.

34. Lin D, Tabb D. L, Yates J. R. Large-scale protein identification using mass spectrometry. Biochim Biophys.Acta 2003; 1646: 1-10.
35. Ducret A, Van O, I Eng J. K, Yates J. R, Aebersold R. High throughput protein characterization. Protein Sci 1998; 7: 706-719.
36. Peng J, Elias J. E, Thoreen C. C, Licklider L. J, Gygi S. P. Proteomics and peptide profiling. J.Proteome.Res 2003; 2: 43-50.
37. Gevaert K, Van Damme P, Ghesquiere B, Vandekerckhove J. Application of a Combined Weak Cation-Exchange/Crown Ether Column: First Demonstrations of a Versatile Tool for Proteome Subselection. Biochim.Biophys.Acta 2006; 1764: 1801-1810.
38. Gevaert K, Van Damme J, Goethals M, Thomas G. R, Hoorelbeke B, Demol H, Martens L, Puype M, Staes A, Vandekerckhove J. Research profile on Biomed Expert. Mol.Cell Proteomics 2002; 1: 896-903.
39. Gevaert K, Goethals M, Martens L, Van Damme J, Staes A, Thomas G. R, Vandekerckhove J. Exploring proteomes and analyzing protein processing by mass spectrometric. Nat.Biotechnol 2003; 21: 566-569.
40. Aebersold R, Mann M. Mass spectrometry-based proteomics. Nature 2003; 422: 198-207.
41. Hillenkamp F, Karas M, Beavis R. C, Chait B. T. Protein Epitope Mapping By Mass Spectrometry. Anal.Chem 1991; 63: 1193-1203.
42. Karas M, Gluckmann M, Schafer J. Ionization in matrix-assisted laser J.Mass Spectrom 2000; 35: 1-12.
43. de Hoffmann E, Stroobant V. Mass spectrometry Principles and applications, Second Ed 1999.
44. Kussmann M, Roepstorff P. Mass Spectrometry of Proteins and Peptides: Methods in Molecular Biology New Jersey. Methods Mol.Biol 2000; 146: 405-424.
45. Smith R.D, Loo J.A, Edmonds C.G, Barinaga C.J, Udseth H.R. A simple model for exploring conformation of highly charged. Anal.Chem 1990; 62: 882-899.
46. Lane C.S. phase proton transfer by photoionized matrix. Cell Mol.Life Sci 2005; 62: 848-869.
47. Willoughby W, Sheehan E, Mitrovich S. A global view of LC/MS, How to solve your most challenging analytical problems, Second Ed 2002.
48. Nano Electrospray. 2008. http://www.newobjective.com/electrospray/index.html
Ref Type: Internet Communication.
49. Jonsson A.P. Mass spectrometry for protein and peptide characterisation. Cell Mol.Life Sci 2001; 58: 868-884.
50. electrospray. 2008. http://www.jic.ac.uk/SERVICES/proteomics/electrospray.html
Ref Type: Internet Communication
51. Eksample of spectra. 2008. http://www.chem.ucsb.edu/~kalpe/bmse259/publie/lmg/Protein_ESI_MS.png
Ref Type: Internet Communication
52. Jensen O.N, Shevchenko A, Mann M. Protein anlysis by mass spectrometry. In protein structure- A Practical approach, Second Ed. Oxford: IRL press, Oxford University Press; 1997. 29-57.
53. Kicman A.T, Parkin M.C, Iles R.K. An introduction to mass spectrometry based ... Molecular and Cellular Endocrinology. Mol.Cell Endocrinol 2007; 260: 212-227.
54. Juhasz P, Roskey M.T, Smirnov I.P, Haff L.A, Vestal M.L, Martin S.A. Methidium intercalator inserted into synthetic oligonucleotides . Anal.Chem 1996; 68: 941-946.
55. Medzihradszky K.F, Campbell J.M, Baldwin M.A, Falick A.M, Juhasz P, Vestal M.L, Burlingame A. A proteomic aproach for the discovery of proteas substrate. Anal.Chem 2000; 72: 552-558.
56. Shevchenko A, Loboda A, Shevchenko A, Ens W, Standing K.G. The synovial proteome: analysis of fibroblast-like synoviocytes. Anal.Chem 2000; 72: 2132-2141.
57. Chernushevich I.V, Loboda A.V, Thomson B.A. Mammalian genomics. J.Mass Spectrom 2001; 36: 849-865.
58. Henzel W.J, Billeci T.M, Stults J.T, Wong S.C, Grimley C, Watanabe C. Identification of Proteins from Fluorescent SYPRO Orange Stained Polyacrylamide Gel. Proc.Natl.Acad.Sci.U.S.A 1993; 90: 5011-5015.
59. James P, Quadroni M, Carafoli E, Gonnet G. Protein Identification by Mass Profile Fingerprinting. Biochem. Biophys. Res. Commun 1993; 195: 58-64.

60. Mann M, Hojrup P, Roepstorff P. Identifying the proteome: software tools. Biol.Mass Spectrom 1993; 22: 338-345.
61. Pappin D.J, Hojrup P, Bleasby A.J. Rapid identification of proteins by peptide-mass fingerprinting. Curr.Biol 1993; 3: 327-332.
62. Yates J. R, Speicher S, Griffin P.R, Hunkapiller T. Challenges in the Overall Analysis of Microbial Proteomes. Anal.Biochem 1993; 214: 397-408.
63. Johnson R.S, Davis M.T, Taylor J.A, Patterson S.D. Purification and Characterization of a 44-kDa Recombinant Collagen I α 1 Fragment from Corn Grain. Methods 2005; 35: 223-236.
64. Shadforth I, Crowther D, Bessant C. Automation, parallelism, and robotics for proteomics. Proteomics 2005:5: 4082-4095.
65. Jungblut P, Thiede B. Protein identification from 2-DE gels by. MALDI mass spectrometry. Mass Spectrom.Rev 1997; 16: 145-162.
66. Thiede B, Hohenwarter W, Krah A, Mattow J, Schmid M, Schmidt F, Jungblut P. R. Peptide mass fingerprinting. Methods 2005; 35: 237-247.
67. Humphery S.I, Cordwell S.J, Blackstock W.P. Proteome research: Complementarity and limitations with respect to the RNA and DNA worlds. Electrophoresis 1997; 18: 1217-1242.
68. Liska A.J, Shevchenko A. Defining parameters for homology-tolerant database searching. Proteomics 2003; 3: 19-28.
69. Garbis S, Lubec G, Fountoulakis M. Limitations of current proteomics technologies. J.Chromatogr.A 2005; 1077: 1-18.
70. Peter James. Proteome research: Mass spectrometry, Spring Verlag 2001.
71. Paizs B, Suhai S. Fragmentation of protonated peptides. Mass Spectrom.Rev 2005; 24: 508-548.
72. Wysocki V.H, Resing K.A, Zhang Q, Cheng G. Mass Spectrometry of Peptides and Proteins. Methods 2005; 35: 211-222.
73. Mascot. 2007. http://www.matrixscience.com/help/fragmentation_help.html
Ref Type: Internet Communication
74. Hanash S. Disease proteomics. Nature 2003; 422: 226-232.
75. Wilkins M.R, Sanchez J.C, Gooley A.A, Appel R.D, Humphery S.I, Hochstrasser D. F, Williams K.L. Challenges in the Overall Analysis of Microbial Proteomes. Biotechnol.Genet.Eng Rev 1996; 13: 19-50.
76. Baggerman G, Vierstraete E, De Loof A, Schoofs L. Gel- based versus gel-free proteomics: a review. Combinatorial Chemistry & High Throughput Screening 2005; 8: 669-677.
77. Issaq H.J. Veenstra T.D. The role of electrophoresis in disease biomarker discovery. Electrophoresis 2007; 28: 1980-1988.
78. Freeman W.M, Hemby S.E. Proteomics for protein expression profiling in neuroscience. Neurochem.Res 2004; 29: 1065-1081.
79. Hoorn E.J, Hoffert J.D, Knepper M.A. The Application of DIGE-based proteomics to renal physiology. Nephron Physiol 2006; 104: 61-72.
80. Campostrini N, Areces L.B, Rappsilber J, Pietrogrande M.C, Dondi,F, Pastorino F, Ponzoni M, Righetti P.G. Thus, proteome analysis identified errors in sequences that were not predicted to correspond to an ORF. Proteomics 2005; 5: 2385-2395.
81. Julka S, Regnier F. 10.1021/pr0340734 15253416 Protein pathway and complex clustering of correlated mRNA and protein expression. J.Proteome.Res 2004; 3: 350-363.
82. Lill J. Protein Quantification by Isotope Dilution Mass Spectrometry of Proteolytic Fragments: Cleavage Rate and Accuracy. Mass Spectrom.Rev 2003; 22: 182-194.
83. Oda Y, Huang K, Cross F.R, Cowburn D, Chait B.T. Accurate quantitation of protein expression and site-specific. Proc.Natl.Acad.Sci.U.S.A 1999; 96: 6591-6596.
84. Conrads T.P, Alving K, Veenstra T.D, Belov M.E, Anderson G.A, Anderson D.J, Lipton M.S, Pasa-Tolic L, Udseth H.R, Chrisler W.B, Thrall B.D, Smith R.D. Microarrays and Mass Spectrometry - The Future of Proteomics. Anal.Chem 2001; 73: 2132-2139.
85. Masselon C, Pasa-Tolic L, Tolic N, Anderson G.A, Bogdanov B, Vilkov A.N, Shen Y, Zhao R, Qian W.J, Lipton M.S, Camp D.G, Smith R.D. SPS' Digest.Bioinformatics. Anal.Chem 2005; 77: 400-406.
86. Ong S.E, Blagoev B, Kratchmarova I, Kristensen D.B, Steen H, Pandey A, Mann M. Mol.Cell Proteomics 2002; 1: 376-386.

87. Zhu H, Pan S, Gu S, Bradbury E.M, Chen X. Mass spectrometry-based proteomics turns quantitative. Rapid Commun.Mass Spectrom 2002; 16: 2115-2123.
88. Ong S.E, Mann M. Mass spectrometry-based proteomics turns quantitative. Nat.Chem.Biol 2005; 1: 252-262.
89. Gygi S.P, Rist B, Gerber S.A, Turecek F, Gelb M.H, Aebersold R. Spectral techniques in proteomics. Nat.Biotechnol 1999; 17: 994-999.
90. Griffin T.J, Han D.K, Gygi S.P, Rist B, Lee H, Aebersold R, Parker K.C. Method for analysing amino acids, peptides and proteins. J.Am.Soc.Mass Spectrom 2001; 12: 1238-1246.
91. Han D.K, Eng J, Zhou H, Aebersold R. Advances in quantitative proteomics via stable isotope tagging and mass spectroscopy. Nat.Biotechnol 2001; 19: 946-951.
92. Huber L.A. Is proteomics heading in the wrong direction? Nat. Rev. Mol. Cell. Biol 2003; 4:74-80.
93. Hamdan M, Righetti P.G. Modern strategies for protein quantification in proteome analysis: advantages and limitations. Mass Spectrom.Rev 2002; 21: 287-302.
94. Li J, Steen H, Gygi S.P. Proteomic analysis of nipple aspirate fluid from women with early-stage breast cancer using isotope-coded affinity tags and tandem mass spectrometry reveals differential expression of vitamin D binding protein. Mol.Cell Proteomics 2003; 2: 1198-1204.
95. Yi E.C, Li X.J, Cooke K, Lee H, Raught B, Page A, Aneliunas V, Hieter P, Goodlett D.R, Aebersold R. Analytical Characteristics of Cleavable Isotope-Coded Affinity Tag-LC-Tandem Mass Spectrometry for Quantitative Proteomic Studies. Proteomics 2005; 5: 380-387.
96. Mirgorodskaya O.A, Kozmin Y.P, Titov M.I, Korner R, Sonksen C.P,Roepstorff P. Establishment and Optimization of Stable Isotope ^{18}O-labeling Strategy for Quantitative Proteomics Research. Rapid Commun.Mass Spectrom 2000; 14: 1226-1232.
97. Stewart I.I, Thomson T, Figeys D. 2 Experimental 2.1 Instrument and reagents the processing of mass spectrometry. Rapid Commun.Mass Spectrom 2001; 15: 2456-2465.
98. Yao X, Freas A, Ramirez J, Demirev P.A, Fenselau C. Anal.Chem 2001; 73: 2836-2842.
99. Patwardhan A.J, Strittmatter E.F, Camp D.G, Smith R.D, Pallavicini M.G. Bioinorganic chemistry. Proteomics 2006; 6: 2903-2915.
100. Miyagi M. Rao K.C. O-labeling strategies for quantitative proteomics. Mass Spectrom.Rev 2007; 26: 121-136.
101. Ivakhno S, Kornelyuk A. Quantitative proteomics and its applications for systems biology. Biochemistry (Mosc.) 2006; 71: 1060-1072.
102. Bantscheff M, Dumpelfeld B, Kuster B. Pathway Proteomics and Chemical Proteomics Team Up in Drug Discovery. Rapid Commun.Mass Spectrom 2004; 18: 869-876.
103. Cagney G, Emili A. De novo peptide sequencing and quantitative. Nat.Biotechnol 2002; 20: 163-170.
104. Steen H. Pandey A. Proteomics goes quantitative: measuring protein abundance. Trends Biotechnol 2002; 20: 361-364.
105. Regnier F.E, Riggs L, Zhang R, Xiong L, Liu P, Chakraborty A, Seeley E, Sioma C, Thompson R .A. Method for the identification and relative quantification of proteins based on the selective isolation of rrnk peptides for the simplification of complex mixtures of proteins. J.Mass Spectrom 2002; 37: 133-145.
106. Chakraborty A, Regnier F.E. FE Global internal standard technology for comparative proteomics. J.Chromatogr 2002; 949: 173-184.
107. Ross P.L, Huang Y.N, Marchese J.N, Williamson B, Parker K, Hattan S, Khainovski N, Pillai S, Dey S, Daniels S, Purkayastha S, Juhasz P, Martin S, Bartlet-Jones M, He F, Jacobson A, Pappin D.J. Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. Mol.Cell Proteomics 2004; 3: 1154-1169.
108. Boehm A.M, Putz S, Altenhofer D, Sickmann A, Falk M. Precise protein quantification based on peptide quantification using iTRAQ™. BMC.Bioinformatics 2007; 8: 214.
109. Unwin R.D, Evans C.A, Whetton A.D. Relative quantification in proteomics: new approaches for biochemistry. Trends Biochem.Sci 2006; 31: 473-484.

110. Guerrera I.C, Kleiner O. Application of Mass Spectrometry in Proteomics. Biosci.Rep 2005.; 25: 71-93.

111. Wiese S, Reidegeld K.A, Meyer H.E, Warscheid B. Exon level integration of proteomics and microarray data. Proteomics 2007; 7: 340-350.

112. Schneider L.V, Hall M.P. Advances in systems biology: Measurement, modeling and representation. Drug Discov.Today 2005; 10: 353-363.

113. Chelius D, Bondarenko P.V. Combined Statin and Niacin Therapy Remodels the High-Density Lipoprotein Proteome. J.Proteome.Res 2002; 1: 317-323.

114. Bondarenko P.V, Chelius D, Shaler T.A. Cathepsin D was below the detection limit in samples from the never-smokers. Anal.Chem 2002; 74: 4741-4749.

115. Tang K, Page J.S, Smith R.D. A Perspective on the Maillard Reaction and the Analysis of Protein Glycation by Mass Spectrometry: Probing the Pathogenesis of Chronic Disease. J.Am.Soc.Mass Spectrom 2004; 15: 1416-1423.

116. Andersen J.S, Wilkinson C.J, Mayor T, Mortensen P, Nigg E.A, Mann M. To elucidate further the nature of the mitotic cell cycle arrests observed. Nature 2003; 426: 570-574.

117. Corbin R.W, Paliy O, Yang F, Shabanowitz J, Platt M, Lyons C.E, Root K, McAuliffe J, Jordan M.I, Kustu S, Soupene E, Hunt D.F. Large-scale comparative genomic ranking of taxonomically restricted genes (TRGs) in bacterial and archaeal genomes. Proc.Natl.Acad.Sci.U.S.A 2003; 100: 9232-9237.

118. Rappsilber J, Ryder U, Lamond A.I, Mann M. Large-Scale Proteomic Analysis of the Human Spliceosome Genome Res. Genome Res 2002; 12: 1231-1245.

119. Ishihama Y, Oda Y, Tabata T, Sato T, Nagasu T, Rappsilber J, Mann M. A novel proteomic screen for peptide-protein interactions. Mol.Cell Proteomics 2005; 4: 1265-1272.

120. Graham R.L, Sharma M.K, Ternan,N.G, Weatherly D.B, Tarleton R.L, McMullan G. Mass Spectrometry-Based Proteomics and Peptidomics for Biomarker Discovery in Neurodegenerative Diseases. Genome Biol 2007; 8:110.

121. Yang Y, Thannhauser T.W, Li L, Zhang S. Targeting Therapy for Breast Carcinoma by ATP Synthase Inhibitor Aurovertin B. Electrophoresis 2007; 28: 2080-2094.

122. Mann K. whereas the most comprehensive list up to date. Proteomics 2007; 7: 3558-3568.

123. Goodlett D.R, Yi E.C. Quantitive MS for proteomics: Teaching a new Dog old tricks. Funct.Integr.Genomics 2002; 2: 138-153.

124. Aebersold R, Goodlett D.R. MS in proteomics. Chem.Rev. 2001;101: 269-295.

125. Nguyen D.N, Becker G.W, Riggin R.M. Protein mass spectrometry: applications to analytical biotechnology. J.Chromatogr 1995; 705: 21-45.

126. Gorman J.J, Wallis T.P, Pitt J.J. Identification and Characterization of Disulfide Bonds in Proteins and Peptides from Tandem MS Data by Use of the MassMatrix MS/MS Search Engine. Mass Spectrom.Rev 2002; 21: 183-216.

127. Konermann L, Simmons D.A. Structural Characterization of Short-Lived Protein Unfolding Intermediates by Laser-Induced Oxidative Labeling and Mass Spectrometry. Mass Spectrom.Rev 2003; 22: 1-26.

128. Schnepf R, Horth P, Bill E, Wieghardt K, Hildebrandt P, Haehnel W. Spectroscopic identification of different types of copper centers generated in synthetic four-helix bundle proteins. J.Am.Chem.Soc 2001; 123: 2186-2195.

129. Kaltashov, Stephen, Loo J.A Mass spectroscopt, instrumention,interpretation,. J.Am.Soc.Mass Spectrom 2005.

130. Winston R.L, Fitzgerald M.C. Visual Literacy in Mass Communications: A Proposal for Educators. Mass Spectrom.Rev 1990; 16: 165-179.

131. Kaltashov I.A. Eyles S.J. Studies of biomolecular conformations and conformational dynamics with mass spectrometry. Mass Spectrom.Rev 2002; 21: 37-71.