We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

6,900
Open access books available

185,000
International authors and editors

200M
Downloads

154
Countries delivered to

Our authors are among the

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Tandem Mass Spectrometry of Peptides

Renata Soares[1], Elisabete Pires[1], André M. Almeida[1,2],
Romana Santos[1,3], Ricardo Gomes[1], Kamila Koči[1],
Catarina Ferraz Franco[1] and Ana Varela Coelho[1]

[1]*ITQB/Universidade Nova de Lisboa*
[2]*Instituto de Investigação Cientifica Tropical& Centro*
Interdisciplinar de Investigação em Sanidade Animal
[3]*Unidade de Investigação em Ciências Orais e Biomédicas,*
Faculdade de Medicina Dentária/Universidade de Lisboa
Portugal

1. Introduction

Tandem MS is considered as a mass spectrum of a mass spectrum raised to the power *n*, where *n* is the number of MS/MS spectra, obtained per molecular ion. When a peptide is analyzed by mass spectrometry, the obtained MS spectrum provides the peptide mass. When tandem MS (or MS/MS) is performed, such peptide is fragmented into daughter ions, which provide information regarding the amino acid sequence of the peptide. Due to this particularity, peptide tandem MS can be used in several applications involving protein characterization and identification, such as the study of proteomes (Franco *et al.*, 2011a,b) and differential proteomics (Puerto *et al.*, 2011) of different organs, tissues, cells or biological fluids. In our laboratory we have also been using this approach for the characterization of proteins with specific function in adhesives (Santos *et al.*, 2009), the identification of protein adducts as potential biomarkers of toxicity (Antunes *et al.*, 2010), and protein glycation (Gomes *et al.*, 2008), it has also been useful in the identification of peptides with immunomodulation properties (Koči *et al.*, 2010). Two types of information obtained by mass spectrometry experiments are used for the study of proteins and peptides. They can be identified or characterized based on their peptide map and primary structure. The advantage of using tandem MS, is that it provides further data, and hence, confirm the assigned peptide identification, thus reducing the chance of obtaining wrongly assigned peptide/protein identifications. Additionally, the information obtained allows localizing of post-translational or chemical modifications at the amino acid residue level.

2. Peptide fragmentation

2.1 Proteins into amino acids

Nowadays, peptide fragmentation is the most commonly used MS information for protein identification in proteomic studies. In this type of studies, there are two approaches that can be used: a gel-based approach (Franco *et al.*, 2001a) or a gel-free approach (Washburn *et al.*, 2001), where the former involves an initial separation of a complex mixture of proteins in,

for example, SDS- or two-dimensional-PAGE. Instead, in the latter, the separation of the complex mixture of previously digested proteins is accomplished using one or two different types of chromatographic separations. In this case, before the separation step, the proteins need to be cleaved into smaller compounds, namely peptides. This is usually accomplished through the use of specific proteases, most commonly trypsin. Different peptides are generated during protein digestion with different proteases, for example trypsin cleaves after lysine or arginine, and hence tryptic peptides will always end in one of these residues. This digestion is necessary as, for most mass spectrometers, proteins are too big to be analysed by tandem MS. This type of approach, where protein digestion is performed before mass spectrometry analysis, is called bottom-up proteomics. Alternatively, intact proteins can be directly analysed in particular mass spectrometers, for example FTICR and Orbitrap mass spectrometers, for protein identification and characterization without the need of proteolytic digestion. In this strategy, named top-down proteomics, proteins are introduced into the mass spectrometer, its mass measured and directly fragmented in the equipment (Reid & McLuckey, 2002). This procedure has several applications related to protein analysis, such as the characterization of post-translational modifications, protein confirmations, protein-ligand and protein-protein complexes, among others (see section 4.2 of this chapter). For more detailed information, we suggest the review by Cui and colleagues (2011).

Sample handling for tandem mass spectrometry analysis should always be done with special care due to the fact that contaminations and induced modifications during sample processing can occur. A common contamination that can occur when performing, in particular, protein digestions is with keratin, which can hinder the identification of proteins. The most common keratin contaminations occur from human hair and hands from the operator , indicating the crucial use of clean lab coat, cap and nitrile gloves. The use of an electrostatic eliminator can also be used to reduce this type of contamination, as demonstrated by Xu *et al.* (2011).

Standard protein digestion procedures are time consuming and involve multisteps. Automatization has been introduced, although it is only cost effective and efficient for a large number of samples. Alternatively simplified protocols have been developed. In a study by Ren and co-workers (2009) a quick digestion protocol was implemented in order to minimize the digestion-induced modifications on proteins, such as asparagine deamidation and N-terminal glutamine cyclization, found to be directly proportional to incubation time in reduction or alkylation and depending on digestion buffer composition. The authors tested this modified protocol on immunoglobulin gamma, which allowed reducing the total experimental procedure to a few hours, beside a protein coverage of 98.6% for IgG.

Particular attention must be driven to the amount of salts and detergents that protein digests have prior to MS or MS/MS analysis due to the fact that these can interfere with the ionization of peptides and can create strong interference on the mass spectrum signal. This can be by-passed by using a HPLC system prior to the MS analysis due to the fact that salts and detergents do not bind to the typical reverse phase columns used for peptide separation eluting in the first steps of the chromatographic run. If a HPLC system is not used, then alternative simpler methods should be used, such as micro columns packed with reverse –phase resins or graphite (Larsen *et al.*, 2007) and, when using MALDI-TOF-TOF, this desalting step can be done directly onto the MALDI target plate (Jia *et al.*, 2007).

2.2 The fragmentation process

The MS fragmentation process occurs in the mass spectrometer mass analyzer or in a collision cell through the action of collision energy on gas phase ions generated in the mass spectrometer ion source. Several parameters influence this fragmentation process, including amino acid composition, size of the peptide, excitation method, time scale of the instrument used, ion charge state, etc (Paiz & Suhai, 2005). Presently, there are several fragmentation processes available in commercial mass spectrometers, namely collision induced dissociation (CID), electron capture dissociation (ECD), electron transfer dissociation (ETD), etc.

The CID fragmentation method can be of two types: low- or high-energy, where the former uses up to 100 eV and the latter from hundreds eV up to several keV (Wells & McLuckey, 2005). The low-energy CID can be found in quadrupoles whereas the high-energy is used in, for example, tof-tof instruments.

The nomenclature of the daughter ions generated by CID was first established by Roepstorff & Fohlmann (1984) and later reviewed by Biemann (1988), where the *b*-series ions extend from the N-terminal and the *y*-series ions extend from the C-terminal. (see Fig. 1). The calculation of the mass difference between consecutive daughter ions belonging to the same ion series (for example *b* or *y* ion-series), allows the determination of peptide's primary sequence.
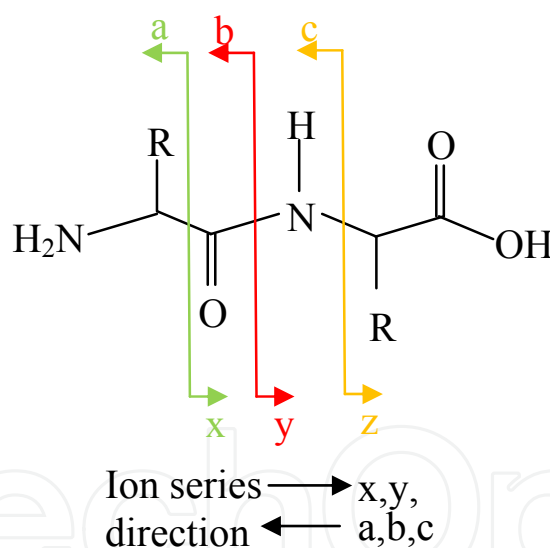

Fig. 1. Schematic diagram of daughter ion nomenclature, adapted from Roepstorff & Fohlmann (1984). A positively charged peptide (in black) is fragmented and the daughter ions are shown (*a, b, c, x, y, z*).

CID fragmentation of peptides uses an inert gas, such as helium, nitrogen or argon, to hit the peptides resulting in the excitation of the molecular ion leading to the polypeptide chain breaking, and to a lower extent to that of the amino acid side chains, generating daughter ions. The low-energy CID allows the rearrangement of the peptide after the loss of a fragment (Yague *et al.*, 2003) and it allows for multiple collisions (up to 100) in a time scale for dissociation of up to milliseconds whereas, the high-energy CID does not permit such process as it is too fast, breaking the peptide. Advantages of the high-energy CID are the

differentiation of isobaric amino acids, such as leucine and isoleucine, as well as more reproducible fragmentation patterns, however only a few collisions can occur (usually no more than ten) due to the fast time scale for dissociation. The information obtained from both types of CID is different; the low-energy CID will generate predominantly *b*- and *y*-series ions, whereas with the high-energy CID *a*-, *x*- and immonium ions are also observed (Fig. 1).

There are several models of how peptides fragment using CID, as reviewed by Paizs & Suhai (2005), namely the *mobile proton* and the *pathways in competition* models. According with the *mobile proton model* a peptide can acquire a positive charge at several sites, named the protonation sites, and these include the terminal amino group, amide oxygens and nitrogens and side chain groups creating several isomers. So, as soon as the peptide becomes excited, a proton is added and this proton will move from protonation site to protonation site before fragmentation (Paiz & Suhai, 2005). The *mobile proton model* can be mainly used to interpret MSMS spectra in a qualitative manner. With the *pathways in competition model*, peptide fragmentation is seen as a competition between charge-remote and –directed in peptide fragmentation pathways, where generation of different peptides follows probability rules based on energetic and kinetic characterization of their major fragmentation pathways. For the charge–remote peptides selective cleavage can occur at the asparagines containing peptides and at oxidized methionines, leading to fragment ions containing information regarding the peptide's amino acid sequence, whereas for the charge-directed peptides fragment ions correspond to the loss of water, ammonia or other neutral losses (Paiz & Suhai, 2005).

A disadvantage of CID is that the side-chain of peptides can be lost. With the other types of fragmentation processes (ECD and ETD) this does not occur, which is particularly important for the identification and characterization of some labile peptide modifications (Zubarev, 2004). While with CID the CO-N bonds are broken along the peptide backbone, with ECD and ETD the N-Cα bonds are broken creating *c*-ions from the N-terminal and *z*-ions from the C-terminal (Fig. 1). Both processes allow extensive peptide backbone fragmentation, while preserving labile side chains (Bakhtiar & Guan, 2006). In ECD there is a reaction

$$[M + nH]^{n+} + e^- \rightarrow \left[[M + nH]^{(n-1)+} \right]^* \rightarrow fragments$$
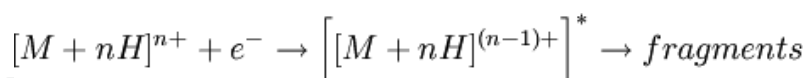
where electrons attach to protonated peptides hence creating peptide cations with an additional electron. After this, the peptide undergoes a rearrangement leading to dissociation (Mikesh *et al.*, 2006). In most cases the information obtained with ECD is complementary to the information obtained with CID (Zubarev, 2004). ECD is mainly used in FTICR mass spectrometers, however it has been developed a similar type of collision energy, ETD, for ion-trap quadrupoles. The ETD advantage is the ability to analyze larger, non-tryptic peptides, allowing the detection of multiple PTMs. ETD dissociates peptides in the same bonds as ECD creating the same ion series. However, ETD does not use free electrons but employs radical anions where A is the anion.

$$[M + nH]^{n+} + A^- \rightarrow \left[[M + nH]^{(n-1)+} \right]^* + A \rightarrow fragments$$
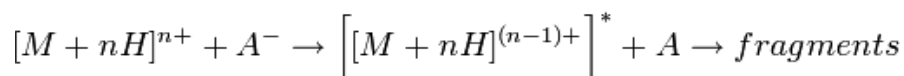
This radical anion transfers an electron to the protonated peptide leading to its fragmentation (Mikesh *et al.*, 2006). ETD cleaves randomly along the peptide backbone while

side chains and modifications, such as phosphorylation, are left intact. The technique only works well for higher charge state ions (z>2), however comparing to CID, ETD is advantageous for the fragmentation of longer peptides or even entire proteins, making this technique important for top-down proteomics.

ECD and ETD produce a complete or almost complete fragmentation spectrum of peptides leading to more information regarding the peptide sequence (Mikesh *et al.*, 2006). However, these types of fragmentation are usually associated with expensive mass spectrometers. Despite the fact that CID is the routine method for fragmentation of peptides, ETD has been described as a preferred method for peptides carrying labile PTMs.

A new method for fragmentation of peptides has been developed by Budnik and co-workers (2001) named electron detachment dissociation (EDD).

$$[M - nH]^{n-} + e^-_{fast} \rightarrow [M - nH]^{(n-1)-\cdot} + 2e^-$$
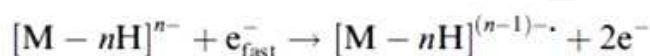
Like ECD and ETD, in EDD, fragmentation of peptides occurs in the N-Cα and provides information regarding the primary structure of peptides. This method is particularly useful for the analysis of acidic proteins (Ganisl *et al.*, 2011) as they ionize better using electrospray and are detected in negative ion mode. A drawback of EDD is the common loss of small molecules (such as CO_2) from the amino acid side-chains.

3. Analysis of tandem mass spectrometry data

After the mass spectrometer analysis a file is created containing a list of masses observed for the peptides, which could be used as precursor ions. To each peptide selected for fragmentation there is, in this experimental file, associated peptide fragments and all this data can be analysed in order to obtain information regarding the amino acid sequence of the peptide and further used for protein identification or characterization (Cottrell, 2011).

There are basically two ways to analyse this type of data: manually or submitting the data to search engines where they are compared to selected protein sequence databases. The manual analysis is done by looking at the mass difference between peaks of a tandem mass spectrum and determining if this mass difference corresponds to the mass of a particular amino acid (see section 3.2 for *de novo* spectrum analysis). However, this is quite a difficult and time-consuming task with several cons, as both ion series are observed in the same spectrum and each series might not be completed, as there can be missing fragments, which produce gaps in the analyzed amino acid sequence. Additionally other fragments, mainly arising from neutral losses and ions from other ion series are usually detected. As a first option fragment patterns interpretation using automatic search engines is usually tried. Although when homologous proteins are not available in protein sequence databases, this approach is not successful. Alternatively, *de novo* interpretation softwares can be used and the obtained results should be carefully checked, usually involving a manual case by case inspection. This situation arises frequently for proteins isolated from organisms with non-sequenced genomes (section 4.1) or if the protein is strongly modified (section 4.3).

3.1 Search engines for protein identification

There are several softwares available for this purpose, based on mathematical algorithms that can help on the interpretation of this type of data. There are some free-ware softwares,

MSKNFLVK
FNCLCVNSLMCVLLK
LSHSHSCVK
WAWSSAIDPWSCCK

Selected theoretical peptides for
database search

MSKNFLVKFNCLCVNSLMCVLLKSVYL
SNVLALAQSPAASWNFDSAICKWAWS
SAIDPWSCCKSPAADLJADPAWIRNVG
LKALALAAERLSHSHSCVKHASAAIYSP
ASSFVFRERTASILRFTHKKNWRDLLV
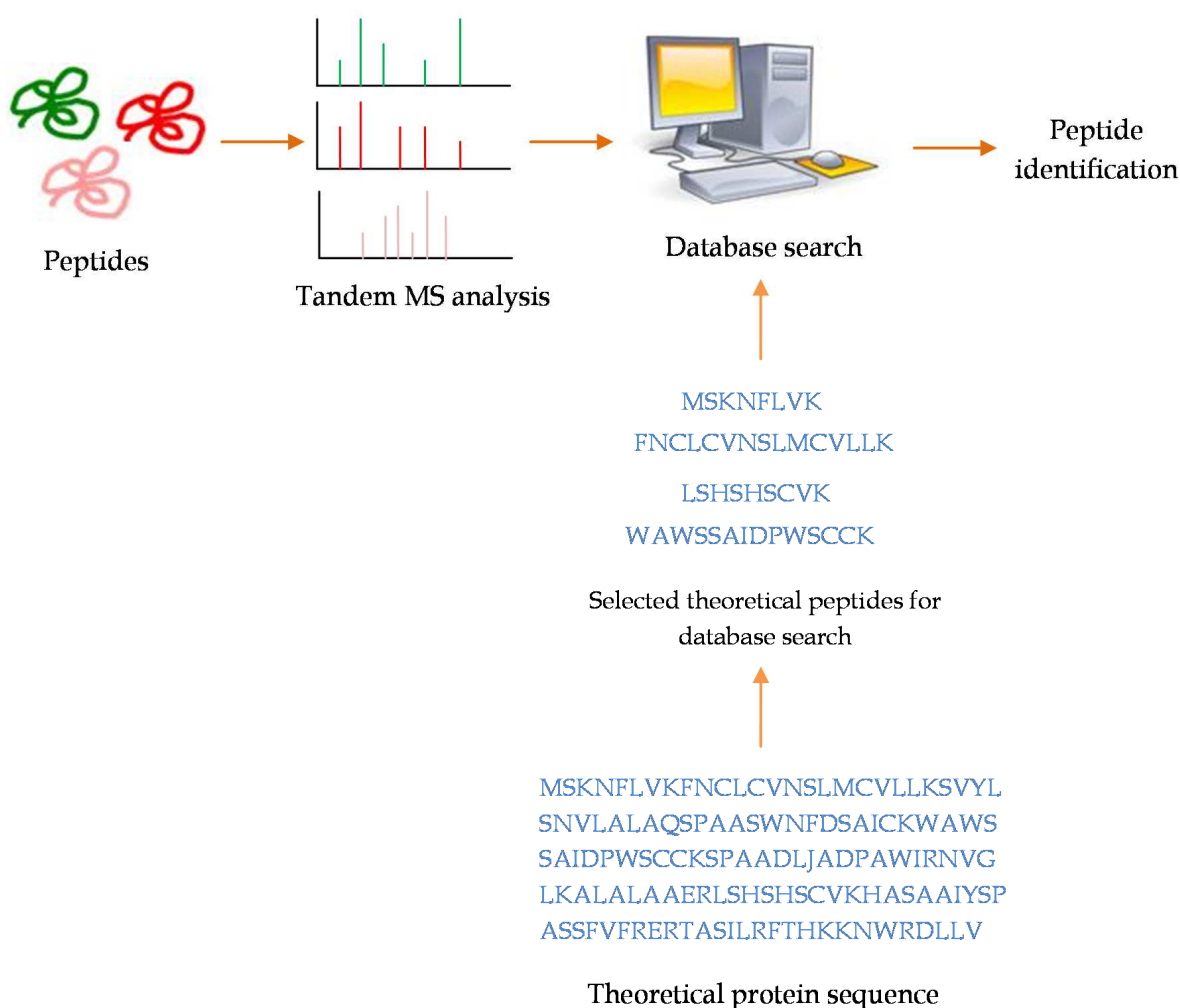
Theoretical protein sequence

Fig. 2. Schematic diagram of the common workflow of database search for peptide identification.

for example MASCOT (Perkins *et al.*, 1999) from Matrix Science (Mowse algorithm), Crux (Park *et al.*, 2008) from University of Washington (Sequest algorithm), and commercially available softwares, for example Protein Pilot (Shilov, *et al.*, 2007) from AB Sciex (which uses Mowse and Paragon algorithms), Bioworks from Thermo Electron (Sequest algorithm) or Peaks from Bioinformatic Solutions. But how do these softwares perform the analysis of the peptide mass spectrometry data? Basically, they make a theoretical digestion, or *in silico* digestion, of all proteins present in a database and generate theoretical fragments of these peptides. After, it is done a comparison between the virtual and experimental data obtained in the mass spectrometer, attributing a score to the peptide or protein identified, where the highest the score the higher match is achieved and, consequently, more confident is the identification generated (Fig. 2). The search process is governed by specified parameters. In a recent review by Eng and co-workers (2011), it is presented a description of these parameters and criteria common in available search engines, as well as their impact on the identification results. Briefly, the mass tolerance for peptides and their fragments should be defined (which varies depending on the resolution and mass accuracy of the mass spectrometer used), the modifications introduced during sample treatment (for example, the

alkylation with iodoacetamide) as well as known modifications of the protein under study, the protease used for digestion and the maximum number of miss cleavages allowed, as well as the database and taxonomy restriction to apply. The major limitation of database searching is that only the peptides that are present in the database can be identified in the search and sometimes by having one different amino acid, the peptide may not be found. The most popular search algorithms include Sequest (Eng *et al.*, 1994), Mowse and X!Tandem.

A recent methodology for peptide identification from tandem MS data, consensus-based method, implies the use of several search engines, hence different algorithms for peptide identification, merging results files and rescoring of identified peptides using platforms such as Scaffold (from Proteome Software Inc.) (Dagda *et al.*, 2010). With this type of strategy it is possible to increase the accuracy, sensitivity and specificity comparing to the use of individual search engines due to different mathematical and algorithmic strategies considered. An advantage of this approach is that it minimizes false positive identifications by using different types of search engines, the disadvantage of one is overcome by the other. The usual search engines used together for this type of strategy are Mowse, Sequest and X!tandem. The use of a fourth search algorithm does not appear to improve the results to a further extend (Dagda *et al.*, 2010).

3.2 *De novo* analysis

De novo sequencing of peptides by MS/MS is the process used for the determination of peptide primary structure not based on the information available in databases. It is ideal for the identification of proteins from organisms without sequenced genome or for the characterization of novel proteins, their isoforms, biological or induced post-translational modifications or peptides with non-proteic amino acids. An interesting review on the subject was published in early 2010 (Seidler *et al.*, 2010). Here we will present further details on manual sequencing and available automatic algorithms, as well as some examples on the characterization of peptides containing modified amino acids.

De novo sequencing analysis of MS/MS fragmentation data can be done manually or using specific softwares. As previously pointed in this section, manual analysis of peptide tandem MS spectra raises usually some difficulties, due to their complexity. Typically a sequence of steps is followed during this exercise, although some have to be performed interactively: a) assignment of major ion series, *b-* and *y-* or *c-* and *z-*; b) calculation of mass differences between peaks attributed to the same ion-series, in order to determine the amino acid residues present in the sequence; c) checking the presence of characteristic patterns for the type of fragmentation process used, namely relative intensity of fragment peaks along the *m/z* range, detection of neutral losses, immonium ions. In order to simplify the fragmentation pattern, several peptide derivatization procedures have been developed that allow the reduction of the peptide fragments to only one ion series, since the second loses its ionization capacities (An *et al.*, 2010, Franck *et al.*, 2010, Hennrich *et al.*, 2010, Miyashita *et al.*, 2011, Nakajima *et al.*, 2011). This strategy presents some drawbacks, namely due to interference from the excess reagent and corresponding side products and low yield of derivatization products. In our laboratory we have successfully used this approach for the identification and N-terminal characterization of a cutinase purified from *Colletotrichum kahawae*, a causal agent of the coffee berry disease (Zhenjia *et al.*, 2007). Although having four tryptic peptides common to *C. gloeosporioides* cutinase, the 21-kDa protein from *C. kahawae*
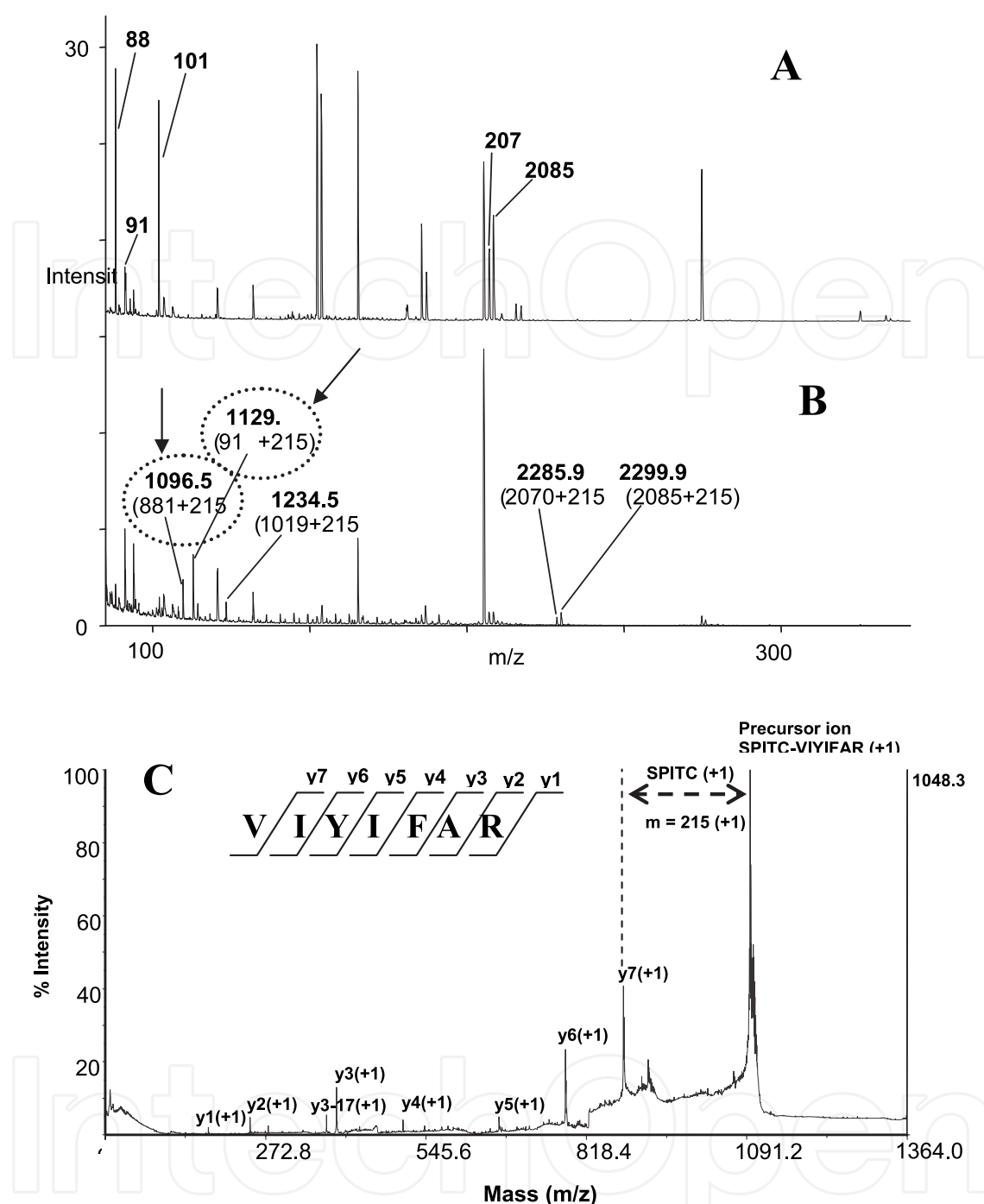
Fig. 3. Mass spectra from *C. kahawae* after trypsin cleavage. A) MALDI-TOF MS spectra of the tryptic peptides before derivatization with SPITC (peptides that were derivatized are labelled). B) MALDI-TOF MS spectra of the tryptic peptides after derivatization with SPITC. Peptides derivatized with SPITC (increase of 215 Da in *m/z* value) were subjected to PSD analysis to obtain sequence information. Arrowheads indicate the peptides that were subjected to further sequencing. C) Peptide sequence from peptide *m/z* 1.096. The resulting *y* ions from *m/z* 1.096 fragmentation were identified in the database as the sequence VIYIFAR (MASCOT, MS/MS ions search). This peptide was found in cutinase precursor form *C. capsici*. The score of 63 for this identification is higher than the minimum of 38, which indicates identity or extensive homology (*p*<0.05). Figure from Zhenjia *et al*, 2007.

could not be identified with a significant score. To enhance fragmentation towards the peptide bond, as described by Wang *et al.* (2004), a sulphonic acid group was introduced at the N-terminus of the tryptic peptides using 4-sulphophenyl isothiocyanate reagent. Two of the peptides derivatized were sequenced (Fig. 3). After a search in the database for identification of the MS/MS ions generated, one of the peptides revealed 100% homology to a highly conserved peptide from a cutinase precursor from *C. gloeosporioides*. Additionally, in the peptide mass map of the 21-kDa protein, a peptide with a monoisotopic mass of 1.653 ([M+H]$^+$) is present, suggesting the blockage of the N-terminus with a glucuronamide residue.

Several software packages for *de novo* sequencing interpretation of MS/MS spectra are available, either proprietary, publicly available academic interfaces or commercial programs. Bringans *et al.* (2008) published a comparative study of the accuracy of three commonly used programs DeNovo Explorer™ (Applied Biosystems), Peaks Studio 4.5 (Bioinformatics Solutions, Inc) and PepNovo (Pitzer *et al.*, 2007). Combining the number of correct residues assigned for Q TRAP and MALDI-TOF/TOF data, Pep-Novo (64.9%) and PEAKS (64.5%) presented a higher accuracy than De Novo ExplorerTM (38.2%). The obtained results indicated that the middle of the peptide was more accurately sequenced and the optimal peptide length is 10-12 residues. As expected, the percentage accuracies were generally related to the quality of the data. PEAKS and PepNovo although presenting very similar accuracy results, often made different choices, consistent with the difference in their algorithm design. For increased confidence in the accuracy of a derived sequence, a combination of algorithms is advised. More recently several other programs for *de novo* analysis have been developed (Chi *et al.*, 2010, He & Ma, 2010, Tessier *et al.*, 2010). The determined amino acid sequences can be used for further functional analysis by homology search of protein sequence databases using adequate programs, namely BLAST-related algorithms (Gaeta, 1998), as "bait for gene fishing" or as a probe for antibody production.

The above-described workflow (MS and MS/MS spectra acquisition, *de novo* sequencing and Blast search) was used for the identification of proteins present in sea urchin (*Paracentrotus lividus*) tube foot adhesive (Santos *et al.*, 2009). Beside the six proteins identified by homology–database search, other MALDI-MS/MS spectra did not allowed direct protein identification and were further used to automatically generate *de novo* sequences in an attempt to obtain more information on these proteins. Five *de novo*-generated peptide sequences were found that were not present in the available protein databases, suggesting that they might belong to novel or modified proteins.

4. Aplications of peptide tandem mass spectrometry

4.1 Protein Identification in non-sequenced organisms

As previously mentioned, the inherent characteristics of tandem MS-based protein identification provides a high-throughput and efficient methodology of utmost importance in the context of research in a wide range of research subjects: Plant, Marine, Agronomical, Human and Veterinary Medicine, Physiology, Parasitology, etc. Despite the efficiency and reliability frequently associated to tandem MS, identifications obtained are nevertheless hindered and highly dependent on how further the proteins of a specific study organism have been described and entered in public databases. In this section we will address such

issue and the possible solutions that might overcame when a researcher is faced with the need of conducting proteomics studies using non-model organisms.

The number of protein entries per organism in public databases is extremely varied. In fact, if the organism is either sequenced or a model organism, chances are that the number of entries in public databases are representative enough to lead to a robust and publishable tandem MS-based proteomics study. On the contrary, non-sequenced or non-model organisms are however frequently poorly represented in databases, which may strongly condition the success of any tandem MS-based proteomics study. A quick analysis on the number of entries per organism in a public database, such as NCBI (http://www.ncbi.nlm.nih.gov/pubmed/), is presented in table 1 and is highly illustrative of the above mentioned.

When working with a poorly studied or poorly described organism, protein searches have to rely on protein entries that show a closer level of similarity with those of the protein of interest to the studied biological issue. Such searches are termed homology searches as they are dependent on the existence of homologies between the protein of interest and information available on the internet. An illustrative example of a homology search conducted by our research group, was the identification of tannin binding proteins in the saliva of sacred baboon (*Papio hamadryas*) using MS/MS data (Mau *et al.*, 2011). Sacred baboons are very poorly represented in public databases and, concerning NCBI, a mere total of 641 entries exist (September 2011). As a consequence, all proteins submitted for identification were identified in primate species other than *Papio hamadryas*, nevertheless with higher levels of representation in databases: Macaque, Colobus and Pan species (respectively 70.000, 1.057 and 43.890). Another striking example was the study involving Mediterranean mussel (*Mytilus galloprovincialis*) and the Vietnamese clam (*Corbicula fluminea*) when exposed to *Cylindrospermopsis raciborskii* cells (Puerto *et al.*, 2011). Again, both studied species are very poorly represented in public databases (1575 for *M. galloprovincialis* and 107 for *C. fluminea*). Consequently, all 26 identifications were obtained with higher level of confidence in other species, with special relevance to the blue mussel (*Mytilus edulis*), a species with 3.154 entries in NCBI.

Interestingly, public generalist databases may have remarkable differences regarding the number of entries for a specific organism. Such difference will have necessary consequences on the success rate of the identifications achieved. For instance, our research laboratory has conducted a study on the proteome of cattle pathogen *Ehrlichia ruminantium*, the agent of the tick-borne disease heartwater or cowdriosis (results unpublished). Identifications were extremely low when conducting the search using the UNIPROT KB curated database (2.369 entries for *Ehrlichia ruminantium*). On the contrary, when the search was conducted resorting to NCBI database, success rate of the identifications was substantially increased, as NCBI more than triples the number of entries for *E. ruminantium*. Such difference is probably the result of the fact that the number of laboratories conducting research on those specific bacteria is extremely low. Consequently, the choice on the public database to deposit protein sequences has necessarily significant implications. Additionally, curated databases only allow entries for verified protein sequences. Such rule limits strongly the access to entries based on, for instance, theoretical sequences obtained from genome-based information, often the most relevant sources of information.

Resorting to public databases may itself strongly limit the probability of obtaining robust and reliable identifications. In fact, the number of proteins entered in public databases is ultimately dependent on the availability of the researcher and host institutions to enter the

Sequenced or Model Organisms		Non-Sequenced or Non-Model Organisms	
Organism	**Number of Entries**	**Organism**	**Number of Entries**
Arabidopsis thaliana	552.550	Red cabbage (*Brassica oleracea*)	2.290
Barrel Medic (*Medicago truncatulla*)	10.223	Lucerne (*Medicago sativa*)	2.222
Rat (*Rattus norvegicus*)	113.981	Black rat (*Rattus rattus*)	937
Mouse (*Mus musculus*)	266.662	Ryukyu mouse (*Mus caroli*)	205
Cattle (*Bos taurus*)	108.129	Water Buffalo (*Bubalus bubalis*)	2.089
Rabbit (*Oryctolagus cuniculus*)	40.707	European hare (*Lepus europaeus*)	478
Human (*Homo sapiens*)	1.963.679	Gorilla (*Gorilla gorilla*)	2.100
Chimpanzee (*Pan troglodytes*)	43.890	Pigmy Chimpanzee (*Pan paniscus*)	1.117
Rhesus macaque (*Macaca mulatta*)	47.029	Celebes macaque (*Macaca nigra*)	73
Saccharomyces cerevisiae	225.898	*Saccharomyces bayanus*	541
Search performed on the 20th September 2011			

Table 1. Protein entries in NCBI database – sequenced and model organisms vs. non-sequenced and closely related non-model organisms.

information obtained in the database. Consequently, and imagining a putative consortium or laboratory would be sequencing a given organism, if the information generated would not be included in public databases, all such information would not be accessible to the general research community, particularly what concerns protein identification using mass spectrometry. Such information is, however, frequently compiled in the form of non-publicly disclosed dedicated databases that are organized and updated by the same consortium that generated it. Access to such databases may be granted in the form of collaboration or by ceding it to interested researchers. A pertinent example of the above-mentioned is a study we have conducted in our research group related to somatic embryogenesis in *Medicago truncatula* (results not published). In this study, protein identifications were conducted first using generalist databases, namely NCBI, with rather poor identification rates of around 50%. However, when the same information was submitted to a dedicated database, generated, kept and kindly ceded by the Samuel Roberts Noble Foundation (Ardmore, OK, USA), identifications rates increased significantly to rates over 85% of the searched proteins.

An interesting solution to the lack of information available on a specific organism may rely on the use of a dedicated database on a specific organism to which the studied organism is closely related and to which it would be expected that proteins would share a high level of homology. Such a strategy would associate the specificity of a dedicated database to high homologies of a very close species. It is likely that such strategy would only be surpassed, in successful protein identification rates, by the use of dedicated databases specific to the studied genus or species. Our research group has conducted one of such strategies to the proteome characterization of coelomocytes in sea star (*Marthasterias glacialis)* (Franco *et al.*,

2011b). *Marthasterias glacialis* sea stars are poorly represented in public databases with only 90 and 26 entries in NCBI and UniprotKB, respectively. Our research group has therefore opted to conduct protein identification using a dedicated database on *Strongylocentrotus purpuratus* (purple sea urchin), a sequenced organism, interestingly with over 1.500 entries in Uniprot and over 6.000 in NCBI. With this highly successful strategy and using both nanoLC-MALDI-TOF/TOF MS and two-dimensional electrophoresis with MALDI-TOF/TOF, we were able to identify over 350 proteins of a highly complex proteome.

Conducting research in non-sequenced or non-model organisms poses interesting challenges to proteomics researchers resorting to a uniquely reliable method for protein identification, such as tandem MS. To overcome such difficulties, the use of either homology searches or dedicated specific databases or even the combination of both may be a useful and interesting strategy. Alternatively, the use of *de novo* sequencing may also be an interesting and effective strategy that was the object of a previous section (section 3.2) in this chapter.

4.2 Top-down proteomics

So far, top-down mass spectrometry has been a less common application in tandem mass spectrometry. This global-view technique has been of great interest as a first approach on protein characterization, without the need of prior knowledge, for the determination of protein's exact molecular mass, post-translational modifications, etc. This examination is made without the need of protein digestion, intact proteins are directly fragmented in the mass spectrometer preventing artificial modifications that can occur during sample handling and preserving some PTM information.

In a recent work by Zhang and co-workers (2011), top-down mass spectrometry has been used for the analysis of human and mouse cardiac Troponin T. In this work, top-down mass spectrometry and fragmentation techniques such as CID and ECD were used to identify and characterize the modification sites of phosphorylation, acetylation, proteolysis, and spliced isoforms of the highly acidic N-terminal of cardiac Troponin T. In a preliminary approach, there has also been an attempt to employ this technology to clinical assays, as it was described by Théberge and co-workers (2011) where the software, BUPID (Boston University Protein Identifier) was developed in order to determine variant and/or modified protein sequencing. This analysis was applied to patient samples for the identification of transthyretin and haemoglobin in an automated and fast method, although it was concluded that at this stage it works mainly for small and abundant clinically relevant proteins.

A new technology has now been applied to top-down mass spectrometry, namely travelling wave ion mobility. A recent work by Halgand and colleagues (2011) involved the study of the heterogeneity of the recombinant phosphoprotein domain of the measles virus expressed in *Escherichia coli*. The use of this technique was particularly useful regarding the reduction of top-down MSMS spectra and the elucidation of the origin of subtle sample heterogeneity, regardless of the mechanisms responsible for the amino acid substitutions.

4.3 Post-translational modifications of proteins and peptides containing adducts

Another application of tandem mass spectrometry with biological and clinical relevance is the identification and characterization of post-translational modification (PTMs) of proteins, where mass spectrometry is the method of choice for this purpose. The modification of proteins post-translationally is of great importance in protein activity and cellular metabolism regulation. The protein's amino acids may be modified after translation, altering

the protein's molecular mass, and therefore it cannot be predicted from genome information.

There are over 400 post-translational modifications of proteins described (Creasy & Cottrell, 2004). The most common and naturally occurring PTMs include phosphorylation, glycosylation, cleavage, formylation, methionine oxidation and ubiquitination (see table 2, as reviewed by Farley & Link (2009). We will focus mainly on applications of the most common and also labile PTMs, phosphorylation, glycosylation, and also on glycation.

Phoshporylation and glycosylation are considered labile PTMs. In these cases, the site of modification is lost, including during the fragmentation process, and more precisely with CID. In a work performed by Carapito and co-workers (2009) a method to detect modified peptides in a complex peptide sample and establish the nature of the modification was developed based on the alternation between MS spectra acquisition using different collision conditions. This experiment led to the cleavage of the substituents and hence to the detection of modified peptides based on their specific fragmentation and on the detection of low mass reporter ions. The described approach allows the detection of multiple modifications without prior knowledge on its type.

Phosphorylation is a PTM that is particularly challenging, as it is present in low-stoichiometry amounts in samples, therefore enrichment techniques are usually required prior to mass spectrometry analysis, as reviewed by Leitner *et al.* (2011). Presently, there are several techniques effective and available for this purpose, namely IMAC, titanium dioxide, among others, and recently a highly efficient new enrichment technique using lathanum ions for the precipitation of this type of modified peptides was developed (Pink *et al.*, 2011). The advantage of the lanthanum enrichment is that it is based on a single step precipitation and phosphoproteins can be isolated from frozen tissue or cells, after cell lysis, using several buffer systems compatible with tandem mass spectrometry analysis.

Phosphorylated peptides can be analyzed using tandem mass spectrometry with CID. However, this type of modified peptides can undergo neutral loss, particularly if phosphorylated in serine or threonine amino acid residues. This means that the MSMS spectrum would not contain much information regarding the site of modification and peptide sequence as the energy for peptide fragmentation is directed towards the phosphoric acid. To overcome this disadvantage, neutral loss LC-MS experiments can be done, where MS^n experiments are established. An advantage of this type of methods is that the peptide after the loss of phosphoric acid can be fragmented again and information regarding the modification site and amino acid composition can be obtained. A disadvantage however is that the duty cycle of the equipment is long and the information obtained is more complex for bioinformatic analysis (Leitner *et al.*, 2011). The development of ETD, prevents this type of problem, as the PTM is preserved during fragmentation, allowing more informative MSMS spectra. Top-down has also been extensively used in the identification and characterization of phosphorylation events, where in this case exact mass of the phosphoprotein is obtained followed by direct characterization of the site of modification, without the need for proteolytic digestion.

The other most common PTM is glycosylation. This PTM is important in protein function and cell fate differentiation. Glycosylation, like phosphorylation, is regulated by enzymes and it involves the binding of glycans to proteins being attached in linked or branched chains with various glycan composition and length (Lazar *et al.*, 2011), hence having an impact on charge, conformation and stability of proteins. Glycosylation has been shown to have an important role in several diseases, including cancer, inflammatory diseases and

congenital disorders, etc. The study of glycosylation of proteins has been useful for the development of novel vaccines and biomarkers discovery for diagnosis (Lazar *et al.*, 2011).

PTM	Nominal mass shift (Da)	Stability	Proposed biological function
Phosphorylation			
pSer, pThr	+ 80	Very labile	Cellular signaling processes, enzyme activity, intermolecular interactions
pTyr	+ 80	Moderately labile	
Glycosylation			
O-linked	203, >800	Moderately labile	Regulatory elements, *O*-GlcNAc
N-linked	>800	Moderately labile	Protein secretion, signaling
Proteinaceous			
Ubiquitination	>1000	Stable	Protein degradation signal
Sumoylation	>1000	Stable	Protein stability
Nitrosative			
Nitration, nTyr	+ 45	Stable	Oxidative damage
Nitrosylation, nSer, nCys	+ 29	Stable	Cell signaling
Methylation	+ 14	Stable	Gene expression
Acetylation	+ 42	Stable	Histone regulation, protein stability
Sulfation, sTyr	+ 80	Very labile	Intermolecular interactions
Deamidation	+ 1	Stable	Intermolecular interactions, sample handling artifact
Acylation			
Farnesyl	+ 204	Stable	Membrane tethering, intermolecular interactions, cell localization signals
Myristoyl	+ 210	Stable	
Palmitoyl	+ 238	Moderately labile	
Disulfide bond	− 2	Moderately labile	Protein structure and stability
Alkylation, aCys	+ 57	Stable	Sample handling
Oxidation, oMet	+ 16	Stable	Sample handling

Table 2. Common PTMs of proteins, showing the mass difference of the modified protein as well as its stability and biological function (Farley & Link, 2009).

Due to the recent advances in technology, particularly in high performance liquid chromatography and mass spectrometry, these methods have been the preferred choice for the identification and characterization of glycosylated proteins. Just like with phosphorylation, glycosylation modified protein/peptides need to be enriched prior to mass spectrometry analysis. A common methodology for this purpose includes lectin enrichment, which can selectively separate complex carbohydrate structures, as reviewed by Lazar *et al.* (2011).

In mass spectrometry, the analysis of glycosylated peptides is particularly challenging, due to the heterogeneity of oligossacharide moieties and to proteolysis difficulties, since glycan motifs can prevent access of site-specific endoproteases, reducing protein coverage and identification. A solution to prevent this is to remove the glycans chemically or enzymatically, for example with PNGase F, prior to endoproteinase digestion. Another problem with this type of PTM is the fact that there is an increase of peptide mass due to the glycan moiety, which can lead to outside the adequate resolution range of most mass spectrometers. To overcome this, a combination of broad specificity proteases can be used in order to increase protein coverage and identification confidence, as well as characterization of glycosylation events. The ionization of these modified peptides is also challenging due to the fact that they can have low ionization efficiency, which implies lower sensitivity in peptide detection. The ions generated can be due to the sequential loss of the sugar residues, however, the ions from the fragmentation of the peptide backbone can be in low abundance or even absent (Lazar *et al.*, 2011).

Glycation is a type of chemical modification that can occur *in vivo*, but it is not controlled enzymatically. Glycation is involved in alterations of structure and stability of proteins, hence affecting protein function. This PTM has been shown to be involved in several diseases, such as diabetes and amyloidotic neuropathies. In a study performed in collaboration with our group, the effects of glycation *in vivo* and *in vitro* were studied on yeast enolase. Peptide mass data were used to determine methylglyoxal-derived advanced glycation end-product (MAGE) nature and location. Since only lysine and arginine residues are modified, tryptic digestion of glycated proteins will produce peptides with at least one miss cleavage associated to a defined mass increase corresponding to a specific MAGE. Using this approach for MS data interpretation important differences were observed between *in vivo* and *in vitro* glycation, that is the same residues were consistently modified *in vivo* suggesting that it is a specific process, whereas *in vitro* this is not the case, that is several residues are modified with different glycation end products (Gomes *et al.*, 2008). Another recent work by Oliveira and co-workers (2011) studied the mechanism of insulin fibril formation in the presence of methylglyoxal, the most significant glycation agent *in vivo*. To unequivocally identify glycated peptides and amino acid residues, non-glycated and glycated insulin were digested using chymotrypsin followed by MS and MS/MS analysis. A modified glycated peptide should be exclusively present in the MS spectrum of glycated insulin with a mass value corresponding to the insulin peptide plus the specific mass increment characteristic of a MAGE modification. This information was used to construct an inclusion list of modified peptides to be fragmented by an additional MS/MS experiment using the MALDI-TOF/TOF instrument. The sequence information thus obtained allowed the unequivocal identification of MAGE-modified peptides and also assignment of specific modified amino acids. This study showed that glycation by methylglyoxal agent stabilizes soluble aggregates that retain native-like structures of insulin.

Another application of tandem mass spectrometry of peptides is the study of protein adducts which is of great importance in toxicology studies. In a recent review by Rappaport and co-workers (2011) it is described a strategy for the identification of adducts in human blood, more precisely in haemoglobin and human serum albumin. In this review, a recent mass spectrometry methodology is described, named fixed-step selected reaction monitoring (FS-SRM), that involves a list of theoretical parent and product ions for the detection of all modifications of targeted nucleophile within a range of adduct masses. Another approach for this type of tandem MS application is described by Switzar and colleagues (2011), where the protein digestion protocol conditions were optimized (pH, temperature and time) in order to increase protein coverage and signal intensity of the modified peptides.

This tandem mass spectrometry application has also been employed in the study of the toxicity effect of anti-viral drugs, more precisely nevirapine (NVP) employed against human immunodeficiency virus type-1 (HIV-1) (Antunes *et al.*, 2010). MALDI-TOF-TOF-MS of tryptic digests was used to identify which human serum albumin and human hemoglobin amino acid residues were bound to NVP upon incubation with the synthetic model electrophile 12-mesyloxy-NVP, used as a surrogate for the Phase II metabolite 12-sulfoxy-NVP. The adopted strategy consisted of (i) comparison of the MS spectra of unmodified and NVP-modified HSA and Hb digests. The presence of new m/z peaks in the latter was presumed to correspond to potential NVP-amino acid adducts. (ii) The m/z values observed exclusively in the MS spectra of the tryptic digests of the modified proteins were compared to the theoretical tryptic peptide mass list for each protein, taking into account the mass increase characteristic of NVP modification. (iii) This information was used to construct an inclusion list of possible NVP-modified peptides to be fragmented by an additional MS/MS experiment using the MALDI-TOF-TOF instrument. The amino acid sequence information thus obtained allowed the unequivocal identification of NVP-modified peptides and the assignment of the specific NVP-modified amino acids. This study prompts to the identification of multiple modification sites suggesting several possible biomarkers of nevirapine toxicity that can be useful for monitoring the toxicity of this drug in patients.

Reference should also be made to the characterization of disulfide bonds using tandem mass spectrometry as these covalent bonds are important in protein folding and aggregation and hence the detection of this type of PTM and identification of the involved cysteine residues is particularly important for monitoring recombinant protein production, namely in the biopharmaceutical industry. The wrong establishment of disulfide bonds may have a remarkable importance in protein 3-dimensional structures and, consequently in their functional activity, thus affecting protein drug selection, assay development and drug testing. The conventional methodology for studying this PTM is, with HPLC or mass spectrometry, to compare protein enzymatic digests using the target protein in their reduced and native forms, where the chromatographic peaks or masses obtained are compared and the differences obtained in the native form are considered due to possible disulfide bonded peptides. In a recent study by Janecki & Nemeth (2011) an efficient method using MALDI-TOF-TOF and high-energy CID was applied to identify disulfide-bonded peptides in proteins with well-documented disulfide bond networks (namely bovine insulin and human serum albumin) and on recombinant proteins where disulfide bonds were not defined, without previous separation of the protein native digests. This method was based on the fact that a number of fragmentation processes happen around the S-S bond leading to a "triplet peak" signature in the spectrum (Fig. 4), which results from the symmetric cleavage of the S-

S bond originating a cysteine fragment (middle peak of the "triplet peak signature") and asymmetric cleavage of the S-S bond originating dehydroalanine and thiocysteine fragments with a mass difference from the middle peak of -34 and +32 Da, respectively (Fig. 4; Janecki & Nemeth, 2011). Sometimes a smaller peak identified as the dehydrocysteine peak can occur around the cysteine *m/z* peak. Hence this "triplet peak" signature can be used for identifying peptides containing one or more inter-disulfide bonds.
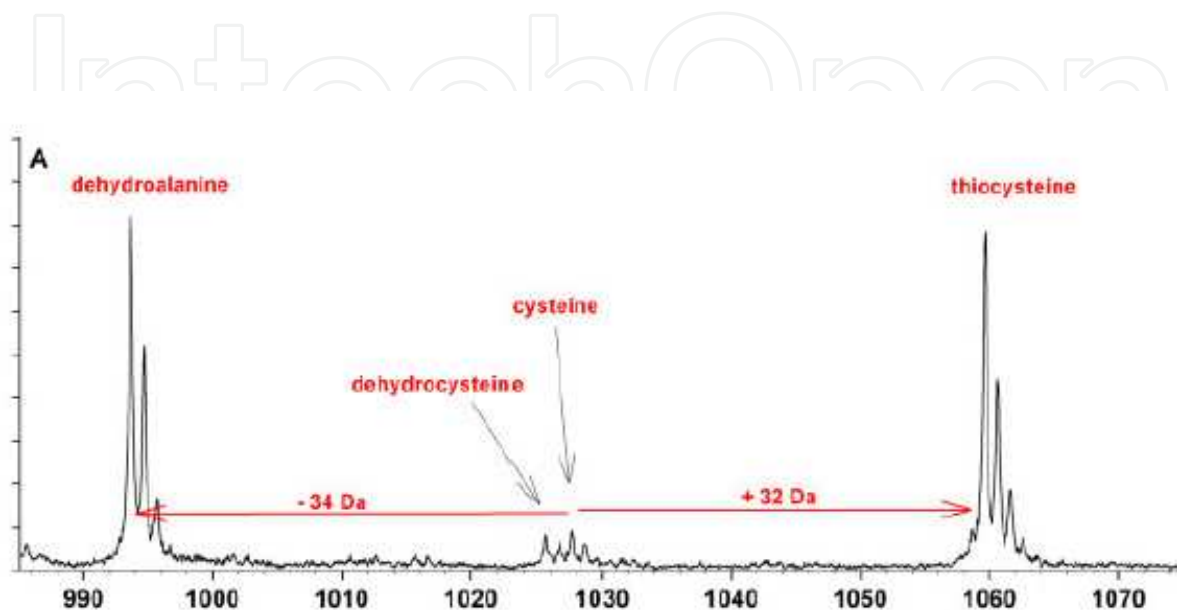
Fig. 4. Zoomed region of the tandem mass spectrum of the *m/z* peak 2071.01 of the native peptic digest of bovine insulin showing the "triplet peak" signature and the characteristic S-S bond fragments (from Janecki & Nemeth, 2011).

Top-down mass spectrometry has also been applied for the identification and characterization of disulfide bonds, however this technique is still limited. In a recent study by Chen and co-workers (2010), native chicken lysozyme was used to evaluate CID fragmentation of protein disulfide bonds by LTQ Orbitrap in positive ion mode. They identified fragments for low-charged protein precursor ions that correspond to the breakage of disulfide-bonds and of protein backbone. These related disulfide-bond fragments resulted from the addition or subtraction of a hydrogen atom or sulfhydryl group with mass changes of -32, -2, +2 and +32 Da for –SH, -H, +H, and +SH, respectively, similar to the patterns obtained by Janecki & Nemeth (2011) for tryptic peptides.

5. Concluding remarks

In this chapter we have reviewed the tandem mass spectrometry approach and its relevance as a technique for the analysis of proteins. To date there are several methodological developments namely, at the levels of sample preparation and experimental procedure and on instrumental and software innovation. Additionally, the number of applications regarding peptide identification, characterization of post-translational modifications, peptide/protein adduct identification, *de novo* studies and proteome studies (including non-sequenced organisms) are still being improved and diversified.

6. Acknowledgments

7. References

An, M., Dai, J., Wang, Q., Tong, Y. & Ji, J. (2010). Efficient and clean charge derivatization of peptides for analysis by mass spectrometry. *Rapid Communications in Mass Spectrometry*, Vol.24, No.13 (July 2010), pp.1869-1874, ISSN 1097-0231

Antunes, A.M.M., Godinho, A.L.A., Martins, I.L., Oliveira, M.C., Gomes, R.A., Coelho, A.V., Beland, F.A. & Marques, M.M. (2010). Proteins Adducts as Prospective Biomarkers of Nevirapine Toxicity. *Chemical Research in Toxicology,* Vol.23, No.11, (November 2010), pp.1714-1725

Bai, L., Romanova, E.V. & Sweedler, J.V. (2011). Distinguishing Endogenous D-Amino Acid-Containing Neuropeptides in Individual Neurons Using Tandem Mass Spectrometry. *Analytical Chemistry*, Vol.83, No.7, (March 2011), pp.2794-2800, ISSN 1520-6882

Bakhtiar, R. & Guan Z. (2006). Electron Capture Dissociation Mass Spectrometry in Characterization of Peptides and Proteins. *Biotechnology Letters,* Vol.28, No.14, (July 2006), pp.1047-1059, ISSN 1573-6776

Budnik, B.A., Haselmann, K.F. & Zubarev, R.A. (2001). Electron Detachment Dissociation of Peptide Di-Anions: An Electron-Hole Recombination Phenomenon. *Chemical Physics Letters*, Vol.342, No.3-4, (July 2001), pp.299-302, ISSN 0009-2614

Carapito, C, Klemm, C., Aebersold, R. & Domon B. (2009). Systematic LC-MS Analysis of Labile Post-Translational Modifications in Complex Mixtures. *Journal of Proteome Research*, Vol.8, No.5, (March 2009), pp.2608-2614, ISSN 1535-3907

Chen, J., Shiyanov, P., Zhang, L., Schlager, J.J. & Green-Church, K.B. (2010). Top-Down Characterization of a Native Highly Intralinked Protein: Concurrent Cleavages of Disulfide and Protein Backbone Bonds. *Analytical Chemistry*, Vol.82, No.14, (July 2010), pp.6079-6089, ISSN 1520-6882

Chi, H., Sun, R.X., Yang, B., Song, C.Q., Wang, L.H., Liu, C., Fu, Y., Yuan, Z.F., Wang, H.P., He S.M. & Dong, M.Q. (2010). pNovo: *de novo* peptide sequencing and identification using HCD spectra. *Journal of Proteome Research*, Vol.9, No.5, (May 2010), pp.2713-2724, ISSN 1535-3907

Cottrell, J.S. (2011). Protein Identification using MS/MS data. *Journal of Proteomics*, Vol.74, No.10, (September 2011), pp.1842-1851, ISSN 1874-3919

Creasy, D.M. & Cottrell, J.S. (2004). Unimod: Protein Modifications for Mass Spectrometry. *Proteomics*, Vol.4, No.6, (June 2004), pp.1534-1536, ISSN 1615-9861

Cui, W., Rohrs, H.W. & Gross, M.L. (2011). Top-Down Mass Spectrometry: Recent Developments, Applications and Perspectives. *Analyst*, Vol.136, No.19, (August 2011), pp.3854-3864

Dagda, R.K., Sultana, T. & Lyons-Weiler, J. (2010). Evaluation of the Consensus of Four Peptide Identification Algorithms for Tandem Mass Spectrometry Based Proteomics. *Journal of Proteomics and Bioinformatics*, Vol.5, No.3, (February 2010), pp.39-47

Eng, J.K., McCormack, A.L. & Yates III, J.R. (1994). An Approach to Correlate Tandem Mass Spectral Data of Peptides with Amino Acid Sequences in a Protein Database. *Journal of the American Society for Mass Spectrometry*, Vol.5, No.11, (November 1994), pp.976-989, ISSN 1879-1123

Eng, J.K., Searle, B.C., Clauser, K.R. & Tabb, D.L. (2011). A Face in the Crowd: Recognizing Peptides through Database Search. *Molecular and Cellular Proteomics*, in press

Farley, A.R. & Link, A.J. (2009). Identification and Quantification of Protein Post-Translational Modifications, In: *Methods in Enzymology*, Burgess, R.R. & Deutscher, M.P., pp.725-763, Elsevier, ISBN 978-0-12-374536-1

Franck, J., Ayed, M.E., Wisztorski, M., Salzet, M. & Fournier, I. (2010). On tissue protein identification improvement by N-terminal peptide derivatization. *Methods in Molecular Biology*, Vol.656, No. 3, pp.323-338, ISSN 1064-3745

Franco, C.F., Santos, R. & Coelho, A.V. (2011a). Exploring the Proteome of an Echinoderm Nervous System: 2DE of the Sea Star Radial Nerve Cord and the Synaptosomal Membranes Subproteome. *Proteomics*, Vol.11, No.7, (April 2011), pp.1359-1364, ISSN 1615-9861

Franco, C.F., Santos, R. & Coelho, A.V. (2011b). Proteome Characterization of the Sea Star Coelomocytes – the Innate Immune Effector Cells of Echinoderms. *Proteomics*, Vol.11, No.17, (September 2011), pp.3587-3592, ISSN 1615-9861

Gaeta, B.A. (1998). Database searching using short peptide sequences. In: *ABRF News*, September 2011, Available from: http://www.abrf.org/ABRFNews/1997/December1997/dec97tips.html

Ganisl, B., Valovka, T., Hartl, M., Taucher, M., Bister, K. & Breuker K. (2011). Electron Detachment Dissociation for Top-Down Mass Spectrometry of Acidic Proteins. *Chemistry – A European Journal*, Vol.17, No.16, (April 2011), pp.4460-4469

Gomes, R.A., Oliveira, L.M., Silva, M.C., Ascenso, C., Quintas, A., Costa, G., Coelho, A.V., Sousa Silva, M., Ferreira, A.E., Ponces Freire, A. & Cordeiro, C. (2008). Protein Glycation *in vivo*: Functional and Structural Effects on Yeast Enolase. *Biochemical Journal*, Vol.416, No.3, (December 2008), pp.317-326

Halgand, F., Habchi, J., Cravello, L., Martinho, M., Guigliarelli, B. & Longhi S. (2011). Dividing to Unveil Protein Microheterogeneities: Travelling Wave Ion Mobility Study. *Analytical Chemistry*, e-pub, (July 2011), ISSN 1520-6882

He, L. & Ma, B. (2010). ADEPTS: advanced peptide *de novo* sequencing with a pair of tandem mass spectra. *Journal of Bioinformatics and Computational Biology*, Vol.8, No.6, (December 2010), pp.981-994, ISSN 1757-6334

Hennrich, M.L., Mohammed, S., Altelaar, A.F. & Heck, A.J. (2010). Dimethyl isotope labeling assisted de novo peptide sequencing. *Journal of American Society for Mass Spectrometry*, Vol.21, No.12, (December 2010), pp. 1957-1965, ISSN 1044-0305

Janecki, D.J. & Nemeth, J.F. (2011). Application of MALDI TOF/TOF Mass Spectrometry and Collision-Induced Dissociation for the Identification of Disulfide-Bonded Peptides. *Journal of Mass Spectrometry*, Vol.46, No.9, (September 2011), pp.677-688, ISSN 1096-9888

Jia, W., Wu, H., Lu, H., Li, N., Zhang, Y., Cai, R. & Yang, P. (2007). Rapid and Automatic On-Plate Desalting Protocol for MALDI-MS: Using Imprinted Hydrophobic Polymer Template. *Proteomics*, Vol.7, No.15, (August 2007), pp.2497-2506, ISSN 1615-9861

Koči, K., Binetti, A., Soares, R., Haag, E., Vinderola, G., Costa, M.C., Reinheimer, J. & Coelho, A.V. (2010). Insights in the Peptidomics of a Functional of a Functional Food Prototype – The Buttermilk Fermentation Product by *Lactobacillus delbrueckii* subsp. *lactis*, presented at 4th Portuguese Mass Spectrometry Meeting, Lisbon, Portugal, December, 2010

Larsen, M.R. (2007). Preparation of Reversed-Phase Microcolumns. *Cold Spring Harbor Protocols*, Vol.11, (November 2007), ISSN 1559-6095

Lazar, I.M., Lazar, A.C., Cortes, D.F. & Kabulski, J.L. (2011). Recent Advances in the MS Analysis of Glycoproteins: Theoretical Considerations. *Electrophoresis*, Vol.32, No.1, (January 2011), pp.3-13, ISSN 1522-2683

Mau, M., Almeida, A.M., Coelho, A.V. & Südekum, K.H. (2011). First Identification of Tannin-Binding Proteins in Saliva of *Papio hamadryas* using MS/MS Mass Spectrometry. *American Journal of Primatology*, Vol.73, No.9, (September 2011), pp.896-902, ISSN 1098-2345

Mikesh, L.M., Ueberheide, B., Chi, A., Coon, J.J., Syka, J.E.P., Shabanowitz, J. & Hunt, D.F. (2006). The Utility of ETD Mass Spectrometry in Proteomic Analysis. *Biochimica et Biophysica Acta*, Vol.1764, No.12, (December 2006), pp.1811-1822

Miyashita, M., Hanai, Y., Awane, H., Yoshikawa, T. & Miyagawa, H. (2011). Improving peptide fragmentation by N-terminal derivatization with high proton affinity. Rapid Communications of Mass Spectrometry, Vol.25, No.9, (May 2011), pp.1130-1140, ISSN 1097-0231

Nakajima, C., Kuyama, H., Nakazawa, T., Nishimura, O. & Tsunasawa, S. (2011).A method for N-terminal de novo sequencing of Nα-blocked proteins by mass spectrometry. *Analyst*,Vol.136, No.1, (January 2011), pp.113-119

Oliveira, L.M.A., Lages, A., Gomes, R.A., Neves, H., Família, C., Coelho, A.V. & Quintas, A. (2011). Insulin Glycation by Methylglyoxal Results in Native-like Aggregation and Inhibition of Fibril Formation. *BioMed Central Biochemistry*, Vol.12, No.41, (August 2011), pp.1-13, ISSN 1471-2091

Paizs, B. & Suhai, S. (2005). Fragmentation Pathways of Protonated Peptides. *Mass Spectrometry Reviews*, Vol.24, No.4, (July/August 2005), pp.508-548

Park, C.Y., Klammer , A.A., Käll, L., MacCoss, M.J. & Stafford Noble, W. (2008). Rapid and accurate peptide identification from tandem mass spectra. *Journal of Proteome Research*, Vol.7, No.7, (May 2008), pp.3022-3027, ISSN 1535-3907

Perkins, D.N., Pappin, D.J., Creasy, D.M. & Cottrell, J.S. (1999). Probability-Based Protein Identification by Searching Sequence Databases using Mass Spectrometry Data. *Electrophoresis*, Vol. 20, No.18, (December 1999), pp.3551-3567, ISSN 1522-2683

Pink, M., Verma, N., Polato, F., Bonn, G.K., Baba, H.A., Rettenmeier, A.W. & Schmitz-Spanke, S. (2011). Precipitation by Lanthanum Ions: A Straightforward Approach to Isolating Phosphoproteins. *Journal of Proteomics*, in press, (August 2011), ISSN 1874-3919

Pitzer, E., Masselot, A. & Colinge, J. (2007). Assessing peptide *de novo* sequencing algorithms performance on large and diverse data sets. *Proteomics*, Vol.7, No.17, (September 2007), pp. 3051-3054, ISSN 1615-9861

Puerto, M., Campos, A., Prieto, A., Cameán, A., de Almeida, A.M., Coelho, A.V. & Vasconcelos, V. (2011). Differential Protein Expression in Two Bivalve Species; *Mytilus galloprovincialis* and *Corbicula fluminea*; Exposed to *Cylindrospermopsis raciborskii* cells. *Aquatic Toxicology*, Vol.101, No.1, (January 2011), pp.109-116, ISSN 0166-445X

Rappaport, S.M., Li, H., Grigoryan, H., Funk, W.E. & Williams, E.R. (2011). Adductomics: Characterizing Exposures to Reactive Electrophiles. *Toxicology Letters*, in press, ISSN 0378-4274

Reid, G.E. & McLuckey, S.A. (2002). "Top-Down" Protein Characterization via Tandem Mass Spectrometry. *Journal of Mass Spectrometry*, Vol.37, No.7, (July 2002), pp.663-675, ISSN 1096-9888

Ren, D., Pipes, G.D., Liu, D., Shih, L.Y., Nichols, A.C., Treuheit, M.J., Brems, D.N. & Bondarenko, P.V. (2009). An Improved Trypsin Digestion Method Minimizes Digestion-Induced Modifications on Proteins. *Analytical Biochemistry*, Vol.392, No.1, (September 2009), pp. 12-21, ISSN 0003-2697

Santos, R., da Costa, G., Franco, C., Alves, P., Flammang, P. & Coelho, A.V. (2009). First Insights into the Biochemistry of Tube Foot Adhesive from the Sea Urchin *Paracentrotus lividus* (Echinoidea, Echinodermata). *Marine Biotechnology*, Vol.11, No.6, (December 2009), pp.686-698, ISSN 1436-2236

Seidler, J., Zinn, N., Boehm, M.E. & Lehmann, W.D. (2010). *De novo* sequencing of peptides by MS/MS. *Proteomics*, Vol.10, No. 4, (February 2010), pp.634–649, ISSN 1615-9861

Shilov, I.V., Seymour, S.L., Patel, A.A., Loboda, A., Tang, W.H.,Keating, S.P., Hunter, C.L., Nuwaysir, L.M. & Schaeffer, D.A. (2007). The Paragon Algorithm, a Next Generation Search Engine That Uses Sequence Temperature Values and Feature Probabilities to Identify Peptides from Tandem Mass Spectra. *Molecular and Cellular Proteomics*, Vol.6, No.9, (May 2007), pp.1638-1655, ISSN 1535-9484

Switzar, L., Giera, M., Lingeman, H., Irth H. & Niessen, W.M.A. (2011). Protein Digestion Optimization for Characterization of Drug-Protein Adducts using Response Surface Modelling. *Journal of Chromatography A*, Vol.1218, No.13, (April, 2011), pp.1715-1723, ISSN 0021-9673

Tessier, D., Yclon, P., Jacquemin, I., Larré, C. & Rogniaux, H. (2010). OVNIp: an open source application facilitating the interpretation, the validation and the edition of proteomics data generated by MS analyses and *de novo* sequencing. *Proteomics*, Vol.10, No.9, (May 2010), pp.1794-1801, ISSN 1615-9861

Théberge, R., Infusini, G., Tong, W., McComb, M.E. & Costello, C.E. (2011). Top-Down Analysis of Small Plasma Proteins using an LTQ-Orbitrap. Potential for Mass Spectrometry-Based Clinical Assays for Transthyretin and Hemoglobin. *International Journal of Mass Spectrometry*, Vol.300, No.2-3, (March 2011), pp.130-142, ISSN 1387-3806

Wang, D., Kalb, S. & Cotter, R. (2004). Improved procedures for N-terminal sulfonation of peptides for matrix-assisted laser desorption/ionization post-source decay peptide sequencing. *Rapid Communications in Mass Spectrometry*, Vol.18, No.1, (January 2004), pp.96–102, ISSN 1097-0231

Washburn, M.P., Wolters, D. & Yates, J.R.3rd. (2001). Large-scale analysis of the Yeast Proteome by Multidimensional Protein Identification Technology. *Nature Biotechnology*, Vol.19, No.3, (March 2001), pp.242-247, ISSN 1087-0156

Wells, J.M. & McLuckey, S.A. (2005). Collision-Induced Dissociation (CID) of Peptides and Proteins, In: *Methods in Enzymology,* Burlingame A.L. (Ed.), pp.148-185, ISBN 978-0-12-182807-3

Xu, B., Zhang, Y., Zhao, Z., Yoshida, Y., Magdeldin, S., Fujinaka, H., Ismail, T.A., Yaoita, E. & Yamamoto, T. (2011). Usage of Electrostatic Eliminator Reduces Human Keratin Contamination Significantly in Gel-Based Proteomics Anaysis. *Journal of Proteomics*, Vol.74, No.7, (June 2011), pp. 1022-1029, ISSN 1874-3919

Yague, J., Paradela, A., Ramos, M., Ogueta, S., Marina, A., Barahona, F., Lopez de Castro, J.A. & Vasquez, J. (2003). Peptide Arrangement during Quadrupole Ion Trap Fragmentation: Added Complexity to MS/MS Spectra. *Analytical Chemistry*, Vol.75, No.6, (March 2003), pp.1524-1535, ISSN 0003-2700

Zhang, J., Zhang, H., Ayaz-Guner, S., Chen, Y.C., Dong, X. & Xu, Q. (2011). Phosphorylation, but Not Alternative Splicing or Proteolytic Degradation, is Conserved in Human and Mouse Cardiac Troponin T. *Biochemistry*, Vol.50, No.27, (June 2011), pp.6081-6092, ISSN 1520-4995

Zhenjia, C., Franco, C.F., Baptista, R.P., Cabral, J.M.S., Coelho A.V., Rodrigues, C.J. & Melo, E.P. (2007).Purification and identification of cutinases from *C. kahawae* and *C. Gloesporioides. Applied Microbiology and Biotechnology*, Vol.73, No.6 (January 2007), pp.1306-1313, ISSN 1432-0614

Zubarev, R.A. (2004). Electron-Capture Dissociation Tandem Mass Spectrometry. *Current Opinion in Biotechnology*, Vol.15, No.1, (February 2004), pp.12-16
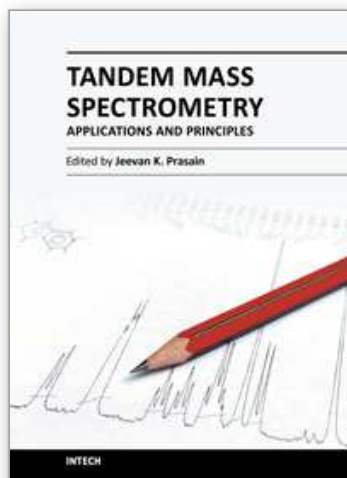
Tandem Mass Spectrometry - Applications and Principles

Edited by Dr Jeevan Prasain

Tandem Mass Spectrometry - Applications and Principles presents comprehensive coverage of theory, instrumentation and major applications of tandem mass spectrometry. The areas covered range from the analysis of drug metabolites, proteins and complex lipids to clinical diagnosis. This book serves multiple groups of audiences; professional (academic and industry), graduate students and general readers interested in the use of modern mass spectrometry in solving critical questions of chemical and biological sciences.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Renata Soares, Elisabete Pires, André M. Almeida, Romana Santos, Ricardo Gomes, Kamila Koči, Catarina Ferraz Franco and Ana Varela Coelho (2012). Tandem Mass Spectrometry of Peptides, Tandem Mass Spectrometry - Applications and Principles, Dr Jeevan Prasain (Ed.), ISBN: 978-953-51-0141-3, InTech, Available from: http://www.intechopen.com/books/tandem-mass-spectrometry-applications-and-principles/tandem-mass-spectrometry-of-peptides

INTECH
open science | open minds