MS/MS in conjunction with capillary liquid chromatography (LC) has been used for identification of proteins excised from 1-D or 2-D gels. The automated nature of protein database searching using uninterpreted MS/MS data (4), the low femtomole sensitivity and the ability to identify a protein based on a single MS/MS product ion spectrum allow the identification of more than 100 proteins per day. Thus, extensive protein identifications have been performed for example to identify cancer-related proteins or to study human pathogens. During a study of MHC-associated peptides the use of LC-MS/MS methods for the analysis of complex peptide mixtures was developed.

In combination with novel stable isotope-tagging methods (metabolic labeling using amino acids, enzymatic incorporation of $^{18}O$, chemical reactions using ►ICAT) and multi-dimensional capillary LC (strong cation exchange/reversed phase), tandem mass spectrometry enables quantification of differences in protein expression. In order to study multi-protein complexes, techniques which include LC-MS/MS and tandem affinity purification (TAP) tags for identification of proteins have been established (1).

Since MS/MS-generated fragment ion spectra provide a "structural fingerprint" of particular peptides resulting from enzymatic degradation of the protein by specific endoproteinases, MS/MS can be used to determine the site of post-translational modification. Numerous reports show that functionally important modifications such as phosphorylation, glycosylation, methylation, myristoylation and palmitoylation are accessible to analysis *via* tandem mass spectrometry in such a way that sequences and modified amino acid side chains can be deduced from C- and N-terminal fragment ions of product ion spectra (Fig. 2).

## Clinical Relevance

Proteomics technologies which include highly sensitive and high-throughput MS/MS methods for efficient identification and characterization of proteins have the potential to increase our understanding of the molecular basis of diseases. This could lead to new disease-specific targets. Tandem mass spectrometry directly contributes to protein profiling, the identification of diagnostic markers and toxicity studies and thus has a profound impact on clinical diagnosis and drug discovery. In addition to its implications for proteomics, MS/MS is becoming an increasingly important method in therapeutic drug monitoring (immunosuppressants, antidepressants) and analysis of metabolic pathways. It is expected that substantial increase in MS/MS sensitivity as well as nano scale LC techniques will further expand the range of applications.

## References

1. Aebersold R, Mann M (2003) Mass spectrometry-based proteomics. Nature 422:198–207
2. Tang XJ, Thibault P, Boyd RK (1993) Fragmentation reactions of multiple-protonated peptides and implications for sequencing by tandem mass spectrometry with low-energy collision induced dissociation. Anal Chem 65:2824–2834
3. Wilm M, Neubauer G, Mann M (1996) Parent ion scans of unseparated peptide mixtures. Anal Chem 68:527–533
4. Eng JK, McCormack AL, Yates JR (1994) An approach to correlate tandem mass-spectral data of peptides with amino acid sequences in a protein database. J Am Soc Mass Spectrom 5:976–989
5. Kleuss C, Krause E (2003) Gαs is palmitoylated at the N-terminal glycine. EMBO J 22:826–832

# Mass Spectrometry: Quantitation

Oliver Fiehn[1], Wolfram Weckwerth[2]
[1]UC Davis Genome Center, Davis, CA, USA
[2]Max-Planck-Institute of Molecular Plant Physiology, Golm/Potsdam, Germany
ofiehn@ucdavis.edu

## Definition

Mass spectrometry is widely used in proteomics for protein identification, mostly after tryptic digestion into peptides (1). For decades, mass spectrometry has also served for determining the absolute or relative amounts of components of complex matrices. There are countless reports on absolute quantification of target molecules by external or internal standards, often by use of stable isotope markers. For some application areas, semi-quantitative methods have been added that aim at comparative assessments of a wide range of compounds rather than ultimate precision for a very few targets. However, in biological contexts, such classical quantitation strategies have only recently been adopted. When sample preparation is aiming at metabolites, such approaches have been named 'metabolomics' (2) (and, for historical reasons, 'metabonomics' in biomedical applications). In the last couple of years there are also reports on quantitative approaches to proteomics using mass spectrometry. All quantitative methods must follow certain guidelines, termed method validation (3). Starting from the definition of the validity area (i.e. the exact description of the scope and objective of the analysis), each method must give certain minimum characteristics such as how compounds of interest can be distinguished from chemical or biological background noise ('selectivity'),

the lower limit ('sensitivity') and the dynamic range where quantitation can be achieved, a robustness assessment and, eventually, the method reproducibility. For each of these prerequisites, mass spectrometric methods may face problems due to instrumental constraints such as ►duty cycles, or lack of tandem mass spectrometric capabilities. More severe, however, are fundamental constraints of ionization suppression and adduct formation that may ultimately undermine any precise comparative quantitation.

## Characteristics

There are more than 50 types of mass spectrometer used for biological determinations, each having specific characteristics limiting their use outside predefined applications. Some features of mass spectrometers, however, are common to all of them, enabling performance comparisons of the instruments and simultaneously, setting fundamental constraints on the analysis of complex mixtures. Firstly, and looking trivial at first sight, mass spectrometers can only detect ions, not neutral species. Therefore, the details of the ionization method are an integral part of any quantitative mass spectrometric method. The same argument is fundamental for another aspect of mass spectrometric detection; pure or 'naked' ions only survive a very short time if they undergo collision, for example at high gas pressures. Therefore, compounds must reach vacuum pressures immediately after, or even before, ion formation. Secondly, all mass spectrometers are able to distinguish different masses; this is called mass resolution $R = m/\delta m$. Since chemical and most physical properties of the molecules are almost identical, if only a few atoms are exchanged for stable isotopes ($^{13}C$, $^{18}O$, $^{2}H$, $^{15}N$), such stable isotope marker compounds are the ideal reference for exact quantitation in mass spectrometry. This is especially true when they are added at the earliest possible time point in sample preparation in order to account for potential losses due to adsorption, precipitation or oxidation. Today, all mass spectrometers are able to distinguish nominal masses (e.g. m/z 212 from m/z 213). A few types of mass spectrometers, for example some time-of-flight instruments, have a higher ►resolution of R = 5,000 to 50,000 or, ultimately, there are FT-ICR mass spectrometers with a resolution of up to R > 1,000,000. Such resolution enables the differentiation and quantitation of co-eluting isobaric compounds that have identical nominal masses but different exact masses due to small differences in the exact atomic weights ($^{16}O$ = 15,994915 u, $^{1}H$ = 1,008725 u, $^{14}N$ = 14,003074).

## GC/MS

One of the oldest and most classical methods is electron impact ionization of molecules in the gas phase. This technique gives robust ionization with reasonable efficiency, but obviously requires volatility of the compounds. Therefore, it is restricted to small molecules such as metabolites and is usually coupled to ►gas chromatography. In a single run, up to 20,000 components can be detected by GC x GC coupling prior to MS detection if time-of-flight mass spectrometers are utilized. Due to the large energy excess ('hard ionization'), the ionization efficacy is not hampered by co-eluting compounds over a wide range of concentrations. Therefore, quantitation can be reliably achieved even by external calibration. Higher precision can be achieved by adding stable isotope labeled compounds as outlined above and this technique has been used for over 40 years. A rougher estimation of differences in metabolite abundances uses direct comparisons of relative peak areas between two or many experiments; this is then called 'relative quantitation' for metabolomic use. This strategy is comparable to approaches undertaken in proteomics or transcriptomics where dye labeling or staining of 2D gel spots is compared without particular calibration curves. By relative quantitation, up to 1,000 metabolites can be compared between various biological tissues by GC/TOF analysis. Unlike 2D gel approaches in proteomics however, the sample throughput is much higher (and costs are lower). This enables a good estimation of reproducibility and its distinction from inherent biological variability – a field of the utmost importance, which has yet to be extensively covered in proteomics (see below). Quantitation in metabolomics by GC/MS regularly achieves 2–10%CV for identical extracts (technical error due to the instrument) or 20–30%CV for the total process (including errors in the extraction and sample preparation process), with typically some 3–4 orders of magnitude dynamic range, good robustness even in high-throughput operations, excellent selectivity due to the good chromatographic resolution (which is some 10-fold higher than in typical liquid chromatography applications) and good sensitivity due to low chemical noise. However, GC/MS is restricted to thermostable, small compounds (up to ~500 u) that can be made volatile, usually by chemical derivatization. Therefore, other approaches must complement GC/MS. As a complementary option, metabolomics by capillary electrophoresis coupled to MS detection detected 1,600 compounds from *B. subtilis* cultures with an overall method precision of 30►%CV for 65 identified metabolites. Other approaches utilizing novel column types for LC/MS are lacking method validation so far.

## LC/MS

The most difficult aspect of quantitation in LC/MS is the ionization process. Unfortunately, this has not always been sufficiently acknowledged, although there are numerous fundamental publications showing the

quenching effect of co-eluting compounds on the ionization efficiency of the target molecules. This phenomenon is called ion suppression and is fundamental to all 'soft' ionization techniques, especially for the most often used electrospray interface. Basically, the ionization efficiency of a particular molecule in a droplet eluting from the LC system is directly related to its hydrophobicity and its susceptibility to receiving a charge; more hydrophilic compounds will preferentially locate in the interior of the droplet, decreasing the likelihood of undergoing desolvation before reaching the skimmer of the mass spectrometer. If a slightly more hydrophobic molecule (or non-volatile component)) is present at the same time as the more hydrophilic one, it will suppress ionization of the hydrophilic compound. Quantitation between experiments is inevitably hampered (especially when many peak abundances are different between the two experiments) and a prediction of the severity of ion suppression is impossible. The same effect results from differences in the ability of molecules to become charged (e.g. basicity/acidity). This and other factors impairing ionization efficiency and quantitation, like geometry and size of the nozzle tip, have been nicely demonstrated by Karas and co-workers using model metabolites (4). This phenomenon is specifically a problem whenever the total composition of a specimen (often called the 'matrix') is believed to be altered, i.e. in any typical biomedical experiment such as comparing ►WT to mutant tissues or healthy to diseased samples.

On top of the difficulty of assessing the effects of ion suppression come alterations in adduct formation. Most biomolecules are only ionisable after adduct formation. The simplest form is protonation or deprotonation, but dozens of further adducts are known in LC/MS, ranging from adducts to the LC eluent (e.g. $[M+\text{acetonitrile}+H]^+$, or $[M+NH_4]^+$ in ammonium acetate buffered LC eluents) to more complex and less predictable adducts like $[M+M+H]^+$, or, in case of co-elution or direct infusion, $[M_1+M_2+H]^+$. For quantitation, the preferred adduct species must be known for each target molecule and its ratio to less abundant adducts must be monitored, since subtle alterations in matrix composition (e.g. minor quantities of Na- or K-salts in the LC buffer eluents) may have cause significant alterations in adduct ratios. The presence of basic amino groups in peptide mixtures diminishes the danger of variable adduct formations in proteomics compared to metabolomic studies.

Due to the unpredictability of the effects of changes in overall matrix compositions on ion suppression and adduct formation, claims that are made for direct LC/MS quantitation of peptides from complex mixtures without internal standards have to be treated with great caution (5). Instead of direct quantitation, use of stable isotope labeled samples has been extensively studied during the last four years. The idea is to combine a control sample with a test sample that consists of stable isotope labeled peptides. Then, both samples are combined and each peptide pair is quantitatively compared based on the abundance ratio of the labeled sample to the unlabeled control. Two basic approaches can be distinguished: either biological (*in vivo*) or chemical (*in vitro*) (6) incorporation of stable isotopes into the peptide test mixture. The advantage of the *in vivo* strategy is the uniformity of protein labeling which is independent of the protein primary structure but it must be seen as a disadvantage that *in vivo* labeling is restricted to use in growing cell cultures. *In vitro* labeling has gained more interest in the past few years. The idea here is to carry out chemical derivatisations with and without stable isotope linkers for specific peptide moieties (such as carboxylic acids, cysteine residues or amino groups). This labeling can also serve to reduce the complexity of mixtures before the LC/MS runs are carried out, without loss in proteomic information. For example, at least one cysteine is present in more than 90% of all predicted yeast ►ORFs and the thiol group of cysteines can readily be used for chemical derivatizations. A biotinylated derivate of iodoacetamide that included a stable isotope marked linker was used for quantitative proteomics. After derivatization, crude protein (or peptide) mixtures can then be purified over streptavidin columns, resulting in a much less complex mixture prior to LC/MS quantitation of each labeled/unlabeled peptide pair. However, disadvantages have been recognized in that thiol residues tend to be so reactive (specifically against oxidation), that the recovery from 2D gels may be irreproducible and that slight differences in the oxidative states of the control/test comparisons may result in high method errors. A variety of other reagents have been tested for proteomic surveys, for example by applying $D_2O$ during tryptic digestion. Other approaches have been specifically focused on post-translationally modified targets such as phosphorylated proteins. However, important parts of the validation of quantitative proteomic methods are still lacking, including in-depth robustness testing. So far, no study has been published focusing on the overall method % CV from different matrices. Reproducibility testing has so far been restricted to the LC/MS errors alone, which have been found to be 10–20%CV. However, it is known that sample preparation usually accounts for the largest error in analytical methods and this can also be assumed to be true for proteomics.

If specific proteins are targeted instead of trying to gain proteomic overviews, the classical analytical method guaranteeing selectivity and sensitivity involved using antibodies for Western blots. It has been shown, that even in this area of protein quantitation, classical

strategies can be complemented by LC/MS approaches. The concomitant characteristics of specific retention times for peptides in LC as well as the unique mass and mass fragmentation have been exploited by a 'mass Western' method. This method utilizes the selectivity power of tandem mass spectrometers by carrying out the quantitation on the ►MS/MS level, i.e. on fragment masses that are highly specific for each compound. If the mass spectrometric detection is then set to the characteristic masses of a peptide and to one of its most abundant daughter ions at the typical LC retention time of this peptide, the increase in signal-to-noise ratios enables the reliable quantitation even of very low abundance proteins after tryptic digestion, because any chemical noise molecule is very unlikely to undergo the same fragmentation at the same fragmentation energies. However, due to potential ion suppression effects, the synthesis and spiking of a stable isotope form of this 'mass Western peptide' is still advantageous for exact quantifications.

## Matrix Assisted Laser Desorption/Ionization-TOF

One of the most often applied techniques for identifying proteins after 1D or 2D gel separations is the in-gel digestion of proteins followed by peptide mass fingerprinting using matrix assisted laser desorption/ionization ►TOF mass spectrometry. For the MS detection, the sample is co-crystallized with a UV absorbing matrix on a plate. When a high-energy UV laser is then pulsed onto such crystals, the matrix immediately sublimes concomitantly with the desorption and ion formation of the target molecules. It is commonly believed that quantitation cannot be done using ►MALDI-TOF, due to differences in analyte/matrix distribution from crystal to crystal. However, modern MALDI-TOF instruments are capable of collecting and averaging spectra from many crystals per sample, enabling at least semi-quantitative fingerprint analysis in mutant screens. If internal standards were added, quantitation by MALDI-TOF was shown to result in identical data to that garnered through LC/MS applications, for example for toxic secondary plant metabolites like potato glycoalkaloids. If stable isotope labeling is included, MALDI-TOF was demonstrated to result in 5–20%CV on protein quantification after tryptic digestion from 2D gels.

## Significance in Quantitation

Apart from the actual choice of direct or stable isotope labeling quantifications in LC/MS, the next and very practical question is how many samples do I need to run to find significant differences between control and test samples? Just a few studies have been published on this topic, although this ultimately should be done in any method validation. Finding statistically significant differences relates to comparing mean values and their corresponding deviations from the means. These deviations are now composed of errors of the analytical method (instrument errors + sample preparation errors) and the inherent biological variation. Unfortunately, the latter is very often underestimated by biologists, and furthermore, it may vary from genotype to genotype and from organ to organ. Molley et al. have found a large contribution of biological variation on top of instrumental errors, which may reach a total error of up to 70%CV in proteomic applications (7). In some cases, appropriate sample pooling strategies may compensate for unwanted biological variation. On the other hand, however, there is inherent information in biological variation that is lost if only mean values are compared. For example, if co-regulation of gene expression transcripts, protein and metabolite abundances are to be analyzed in system biology approaches, the possibility of finding and the significance testing of such co-regulations require a certain amount of variability in the network data sets. For such cases, the analysis of biological snapshots from minute amounts of samples would be more suitable than pooling strategies.

## Instrumentation

For many purposes, the choice of the actual type of mass spectrometer being used for proteomic (or metabolomic) studies is less important than is commonly believed. Each instrument has its own advantages and disadvantages, depending on the actual application. For example, if quantitation is supposed to be verified by peptide identifications, a fast switching between full scan MS and (data dependent) MS/MS fragmentations must be achievable, the so-called duty cycle. Some of the instruments do not allow a convenient way to do MS/MS experiments but are very fast at acquiring full scan spectra (like TOF MS), whereas others have good mass resolution and mass accuracy for *de novo* peptide identification (like QTOF hybrid instruments) but lack high duty cycles. Others like FT-ICR-MS instruments are known for ultimate sensitivity, mass resolution and mass accuracy, but again come with low duty cycles and additionally, prohibitive prices and technical problems causing long down times over the year. Further, mass accuracy may depend on frequency perturbations and charge repulsions and quantitation may also be very difficult by FT-MS if too many ions are present at one time point. MS/MS spectra may vary dramatically between different instrument types due to the differences in the methods used for compound fragmentations. The best comparable spectra across instruments from different manufacturers are gained by ion trap mass spectrometers, and compilation of MS/MS libraries have therefore started with ion traps. Ultimately, the chromatographic (or electrophoretic) resolution, the type and maintenance of the electrospray interface and the sample

preparation method including stable isotope markers are more important than the actual instrument that is eventually employed.

## Clinical Relevance

Modern clinical research relies in all areas on the precision of quantitative and qualitative biochemical analysis for diagnostic purposes, biomarker validation and mechanistic studies. With current methodologies, quantitative metabolite and protein profiling may go far beyond classical diagnostic tests like automated recognition of metabolic diseases or newborn screening. It is the task of analytical chemists and biochemists to serve clinicians by accurately assessing standard operating procedures and critical steps in the process in order to achieve validated quantitative methods with low overall method %CVs. From a technical perspective, current mass spectrometers are able to fulfill such requirements.

## References

1. Ferguson PL, Smith RD (2003) Proteome analysis by mass spectrometry. Annu Rev Biophys Struct 32:399–424
2. Weckwerth W (2003) Metabolomics in systems biology. Annu Rev Plant Biol 54:669–689
3. Krull IS (1999) Analytical method development and validation for the academic researcher. Anal Lett 32:1067–1080
4. Schmidt A, Karas M, Dülcks T (2003) Effect of different solution flow rates on analyte ion signals in nano-ESI MS, or: When does ESI turn into nano-ESI? J Am Soc Mass Spectrom 14:492–500
5. Bondarenko PV, Chelius D, Shaler TA (2002) Identification and relative quantification of protein mixtures by enzymatic digestion followed by capillary reversed-phase liquid chromatography-tandem mass spectrometry. Anal Chem 74:4741–4749
6. Gygi SP, Rist B, Gerber SA et al (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nat Biotechnol 17:994–999
7. Molloy MP, Brzezinski EE, Hang J et al (2003) Overcoming technical variation in quantitative proteomics. Proteomics 3:1912–1919

# Mass Spectrometry: SELDI

Guru Reddy, Richard Rubin, Scot Weinberger
Ciphergen Biosystems, Inc., Fremont, CA 94555, USA
sweinberger@ciphergen.com

## Synonyms

SELDI; SELDI-TOF-MS; SELDI-ProteinChip technology; ProteinChip technology

## Definition

SELDI (surface-enhanced laser desorption and ionization) is a method of protein capture and enrichment on a chemically or bioaffinity active solid phase surface, often followed by selective washing steps and then followed by laser "elution" of the proteins into a detector, usually a time-of-flight mass spectrometer. The whole method is often referred to as SELDI-TOF MS or SELDI ProteinChip technology. With similarities to ►MALDI technology (in which the sample probe is not an active binding partner to protein analytes), an energy absorbing molecule or protein co-crystallization matrix is typically added on top of the captured proteins to assist ionization *via* laser excitation. SELDI is commonly used to discover disease associated protein ►biomarkers and to develop diagnostic assays.

## Characteristics

The key components of SELDI ProteinChip technology are the ►ProteinChip Arrays, the ►ProteinChip Reader and the associated software. ProteinChip Arrays contain various chromatography or biological surfaces to capture proteins from complex biological mixtures. Chromatographic surfaces are composed of hydrophobic, hydrophilic, ion exchange, immobilized metal or other chemistries. These surfaces are often used for profiling of proteins from biological mixtures, for biomarker discovery and for assay development. The activated surfaces contain covalently immobilized specific bait molecules such as antibodies, receptors or oligonucleotides and are often used for bio-molecular interaction studies such as protein-protein and protein-DNA interactions. Users of the activated surface arrays are enabled to customize them with their own bait molecules *via* a simple incubation procedure.

The primary ►Expression Difference Mapping™ application of SELDI involves applying a biological sample such as serum, urine, cell lysates or tissue extracts to a spot on a ProteinChip Array and allowing proteins to bind to the surface. The unbound proteins and mass spectrometric interfering compounds are washed away and the proteins that are retained on the array surface are analyzed and detected by laser desorption/ionization TOF-MS using a ProteinChip Reader (Fig. 1). A variety of statistical techniques and bioinformatic software systems are then used to analyze the data and to detect the differences between the protein profiles of two sample sets (1).

The SELDI ProteinChip technology is a versatile platform that can be used for the discovery of disease associated protein biomarkers and for their purification and identification. In addition, the same platform can be used for ►Interaction Discovery Mapping™ applications to detect proteins that interact with the newly discovered biomarker and also to develop predictive