

Protein identification and profiling with mass spectrometry

Andrew Meyers^a, Sunia Trauger^a, William Webb^a, Nichole Reisdorph^a, Cindy Wranik^a, Eric Peters^b and Gary Siuzdak^{a,*}

^a*The Scripps Research Institute, Center for Mass Spectrometry and Department of Molecular Biology, La Jolla, CA 92037, USA*

^b*Genomics Institute of the Novartis Research Foundation, La Jolla, CA 92037, USA*

1. Overview of the utility of protein ID

As a consequence of the completion of multiple genome projects, there has emerged a need for the efficient evaluation of complex samples of proteins on a broad scale [1–4]. Mass spectrometry (MS) has rapidly become one of the most important tools for the identification of proteins (Fig. 1) largely due to the introduction of new biomolecule-compatible ionization techniques such as electrospray ionization (ESI) [5] and matrix-assisted laser desorption/ionization (MALDI) [6] as well as high resolution mass analyzers. Using these technologies, proteolytic peptides can be ionized intact into the gas phase and their masses accurately measured. Based on this information, proteins can readily be identified using a methodology called protein mass mapping or peptide mass mapping, in which these measured masses are compared to predicted values derived from a protein database. Further sequence information can also be obtained by fragmenting individual peptides in tandem MS experiments. In addition, large scale changes in protein expression levels between two different samples can be assessed using quantitative tools such as two-dimensional gel electrophoresis (2D-GE) or stable isotope labeling in conjunction with mass spectrometry measurement.

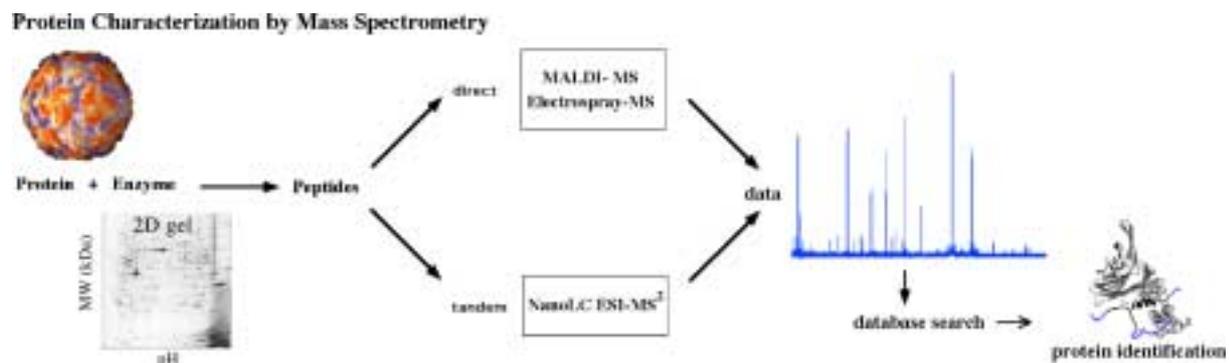


Fig. 1. Mass spectrometry strategies used for protein identification.

*Corresponding author. E-mail: siuzdak@scripps.edu.

Table 1

Protease specificity. Proteolysis experiments can use any of a number of enzymes to perform digestion. The cleavage specificity of some of the different enzymes is denoted by a slash (/) before or after the amino acid responsible for specificity. Combinations of proteases can be used to reduce specificity and to mimic other proteases. For example Lys-C and clostripain together are specific for the same sites as trypsin

Protease	Amino acid specificity	Exceptions
Trypsin	X- Lys /-Y X- Arg /-Y	Does not cleave if Y = Pro
Endoproteinase Lys-C	X- Lys /-Y	Does not cleave if Y = Pro
Clostripain	X- Arg /-Y	
Endoproteinase Asp-N	X- Asp /-Y X- cysteic acid bonds	Does not cleave if Y = Ser
CNBr	X- Met /-Y	Does not cleave if Y = Ser, Thr, or Cys
Glu-C (V8 Protease (E))	X- Glu /-Y X- Asp /-Y	Does not cleave if X = Pro
Pepsin	X- Phe /-Y X- Leu /-Y X- Glu /-Y	Does not cleave if Y = Val, Ala, Gly
Endoproteinase Arg C	X- Arg /-Y	Does not cleave if Y = Pro
Thermolysin	X- Phe -Y X- Ile -Y X- Leu -Y X- Ala -Y X- Val -Y X- Met -Y	Does not cleave if X = Pro
Chymotrypsin	X- Phe /-Y X- Tyr /-Y X- Trp /-Y X- Leu /-Y	Does not cleave if Y = Met, Ile, Ser, Thr, Val, His, Glu, Asp
Formic acid	X- Asp -Y	

1.1. Identification by peptide mass mapping

Protein identification has traditionally been done by subjecting proteolytic digests to high performance liquid chromatography (HPLC) followed by *N*-terminal (Edman) sequencing and/or amino acid analysis of the separated peptides. However, these techniques are relatively laborious, insensitive, and do not work with *N*-terminally modified peptides. More recently, mass spectrometry has been combined with protease digestion to enable peptide mass mapping. Definitively, peptide mass mapping combines enzymatic digestion, mass spectrometry, and computer-facilitated data analysis for protein identification. Sequence specific proteases or certain chemical agents (Table 1) are used to obtain a set of peptides from the target protein that are then mass analyzed. The enzyme trypsin is a commonly used protease that cleaves peptides on the *C*-terminal side of the relatively abundant amino acids arginine (Arg) and lysine (Lys). Thus, trypsin cleavage results in a large number of reasonably sized fragments from 500 to 3000 Daltons, offering a significant probability for unambiguously identifying the target protein. The observed masses of the proteolytic fragments are compared with theoretical “*in silico*” digests of all the proteins listed in a sequence database (Fig. 2). The matches or “hits” are then statistically evaluated and ranked according to the highest probability.

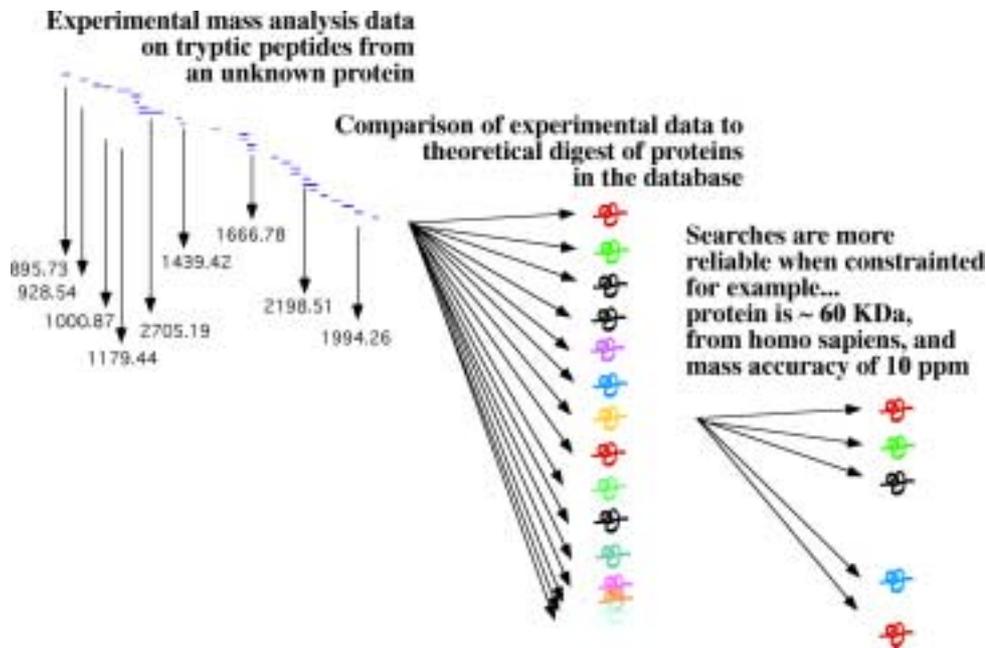


Fig. 2. Protein identification through the comparison of tryptic peptides of an unknown protein to the theoretical digest of known proteins. The identification can be made to be more reliable when constraints are added such as the source and size of the protein, as well as when high accuracy data and tandem mass measurements are used.

Clearly, the success of this strategy is predicated on the existence of the correct protein sequence within the database searched. However, the quality and content of such databases are continually improving as a result of genomic sequencing of entire organisms, and the likelihood for obtaining matches is now reasonably high. While exact matches are readily identified, proteins that exhibit significant homology to the sample are also often identified with lower statistical significance. This ability to identify proteins that share homology with poorly characterized sample species makes protein mass mapping a valuable tool in the study of protein structure and function.

Upon submitting a query to a search program, a theoretical digest of all the proteins in the database is performed according to the conditions entered by the researcher. Variables that can be controlled include taxonomic category, digestion conditions employed, the allowable number of missed cleavages, protein pI and mass ranges, possible post translational modifications (PTMs), and peptide mass measurement tolerance. A list of theoretical peptide masses is created for each protein in the database according to the defined constraints, and these values are then compared to the measured masses. Each measured peptide generates a set of candidate proteins that would produce a peptide with the same mass under the digestion conditions specified. The proteins in these sets are then ranked and scored based on how closely they match the entire set of experimental data.

This method of identification relies on the ability of mass spectrometry to measure the masses of the peptides with reasonable accuracy, with typical values ranging from roughly 5 to 50 ppm (5 ppm = ± 0.005 Daltons for a 1,000 Daltons peptide). The experimentally measured masses are then compared to all the theoretically predicted peptide digests from a database containing possibly hundreds of thousands of proteins to identify the best possible matches. Various databases (Table 2) are available on the Web, and can be used in conjunction with such computer search programs such as Profound (developed at Rockefeller University), ProteinProspector (University of California, San Francisco) and Mascot (Matrix

Table 2

Two of the protein databases available on the Internet

NCBIInr	A non-redundant database compiled by the National Center for Biotechnology Information (NCBI) by combining most of the public domain databases (Expressed Sequence Tags (EST's) not included).
Swiss Prot	A curated protein sequence database which strives to provide a high level of annotations, such as the description of the function of a protein, its domain's structure, post-translational modifications, variants, etc. This database offers a minimal level of redundancy and high level of integration with other databases.

Science, Limited). One obvious limitation of this methodology is that two peptides having different amino acid sequences can still have the same exact mass. In practice, matching 5–8 different tryptic peptides is usually sufficient to unambiguously identify a protein with an average molecular weight of 50 kDa, while a greater number of matches may be required to identify a protein of higher molecular weight. It is important to note that the term protein identification as used here does not imply that the protein is completely characterized in terms of its entire sequence as well as all its PTMs. Rather, this term typically refers to matching the sample to the base amino acid sequence as translated from the encoding gene.

In theory, accurate mass measurements of the undigested sample could also be used for protein identification. In practice however, the identification of a protein based solely on its intact masses is nearly impossible due to the stringent sample purity required, the need for extremely accurate mass measurements, and most importantly, the unpredictable variability introduced by numerous possible PTMs.

1.2. Identification using tandem mass spectrometry

A more specific database searching method involves the use of partial sequence information derived from MS/MS data (Fig. 3). As discussed later, tandem mass spectrometry experiments yield fragmentation patterns for individual peptides. Manual interpretation of a tandem MS experiment can often be quite difficult due to the number of different fragmentations that can occur, not all of which yield structurally useful information. However, in analogy to peptide mapping experiments, the experimentally obtained fragmentation patterns can be compared to theoretically generated MS/MS fragmentation patterns for the various proteolytic peptides arising from each protein contained in the searched database. Statistical evaluation of the results and scoring algorithms using search engines such as Sequest (ThermoFinnigan Corp.) and MASCOT (Matrix Science, Limited) facilitate the identification of the best match. The partial sequence information contained in tandem MS experiments is much more specific than simply using the mass of a peptide, since two peptides with identical amino acid contents but different sequences will exhibit different fragmentation patterns.

2. Prior to mass analysis: separation

2.1. The requirement for sample separation

Although these methodologies have greatly enhanced the ability to perform efficient protein identifications, they cannot directly be used to identify all the proteins present in a typical biological sample due to the significant signal suppression caused by complex mixtures in mass spectrometry. Trypsin digestion

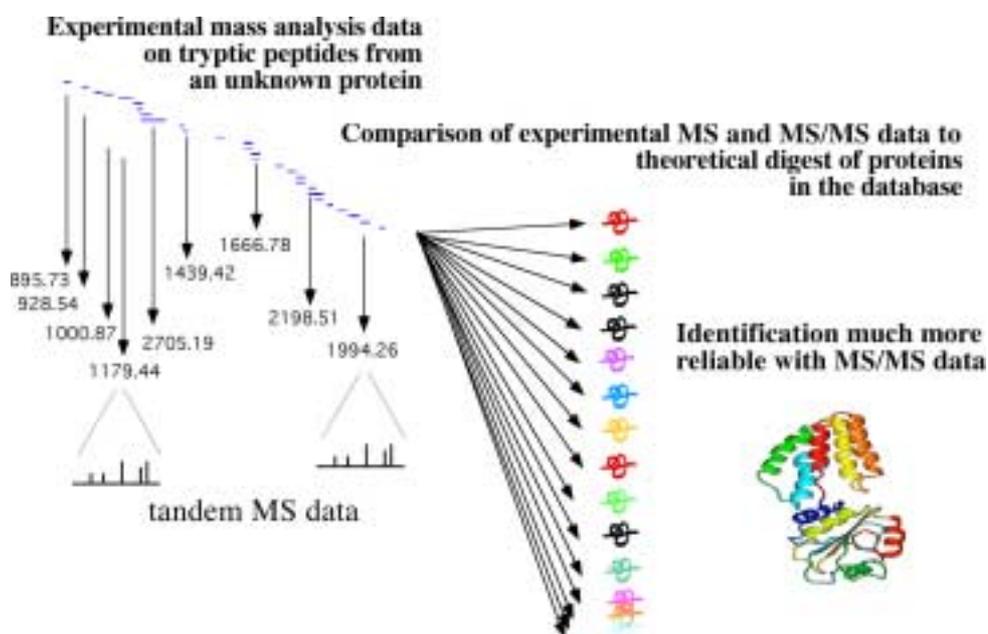


Fig. 3. Protein identification through the comparison of tryptic peptides of an unknown protein to the theoretical digest and theoretical MS/MS data of known proteins is even more reliable than just comparing the mass of the fragments.

Table 3
Protein identification with MALDI and LC-MS/MS

	MALDI TOF reflectron	LC-MS/MS with an ion trap analyzer
Advantages	Very fast Widely available Easy to perform analysis High accuracy (5–30 ppm) adds reliability to data Useful for a wide range of proteins	In addition to molecular mass data, tandem MS measurements are performed in real time. MS/MS information adds additional level of confirmation. Multiple proteins can be analyzed simultaneously with simple reversed-phase LC run Useful for PTM identification High coverage of proteins (30% to 90%) depending on the protein
Disadvantages	Problematic for mixtures of proteins Typically less coverage than LC-MS/MS approach.	Computationally intensive. Large database searches can take days. Cluster computer systems not easily available.

of a typical protein can result in the production of roughly fifty peptides, while miscleavages and various PTMs can give rise to many other unique species. Thus, biologically-derived samples can contain thousands to literally millions of individual peptides in the case of whole cell extracts. By comparison, the tryptic digestion of approximately 3–5 proteins results in a peptide mixture complex enough to cause considerable signal suppression. Thus, samples of proteins (or peptides in a proteolytic digest) must be separated [7–19] by gel electrophoresis or liquid chromatography prior to mass analysis.

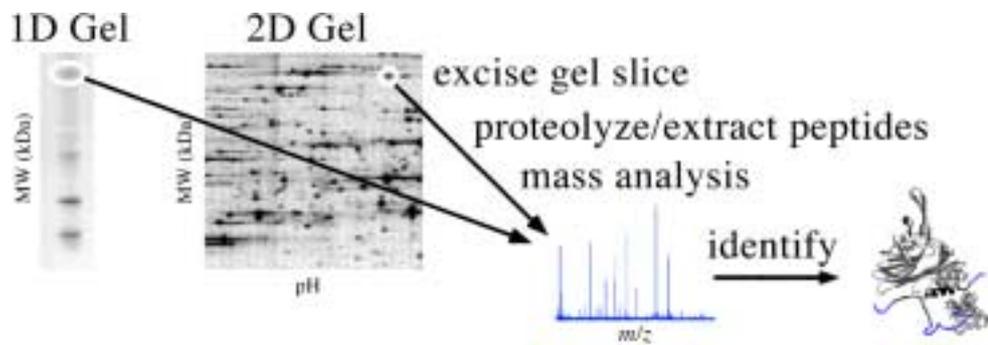


Fig. 4. After performing SDS PAGE separation on a 1D (or 2D) gel the stained portion of the gel representing the sample is cut out and then prepared for mass analysis by destaining, performing alkylation/reduction in-gel digestion, and spotting on a MALDI plate for analysis. This can be done manually or automatically performed with a robot.

2.1.1. Gel Electrophoresis and mass spectrometry for protein characterization/profiling

Gel electrophoresis is one of the most widely used techniques for separating intact proteins. In sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE), sometimes called one dimensional gel electrophoresis, the proteins are treated with the denaturing detergent SDS and loaded onto a gel. As a result, each protein becomes coated with a number of negatively charged SDS molecules directly proportional to its total number of amino acids. Upon application of an electric potential across the gel, all the proteins migrate through the gel towards the anode at a rate inversely proportional to their size. The separation is typically performed with a ladder containing multiple proteins of known masses run alongside the proteins of interest in order to provide a size reference. Upon completion of the separation, the proteins are visualized using any of a number of different staining agents, and the individual bands are physically excised from the gel. These samples are subjected to in-gel digestion, peptide extraction, and protein identification (Fig. 4). A typical procedure for protein in-gel digestion and extraction is given below.

2.1.2. A typical in-gel digestion protocol

1. Run gel and locate the protein bands of interest.
2. Excise bands of interest to 1 mm² pieces.
3. Destain with 25 mM ammonium bicarbonate : 50% acetonitrile (ACN) and shake 10 min. Discard supernatant. Repeat until gel slices are clear.
4. Alkylation/reduction. Add ~25 μ l 10 mM dithiothreitol (DTT) or enough solution to cover the slices, vortex, spin, and let reaction proceed 1 hr at 56°C. Remove DTT and add ~25 μ l 55 mM iodoacetamide. Vortex, spin, and allow reaction to proceed 45 min in the dark. Remove supernatant. Wash gel slice with ~100 μ l ammonium bicarbonate 10 min. Discard supernatant. Wash 2 \times with 25 mM ammonium bicarbonate/50% ACN.
5. Speed Vac to dryness.
6. Rehydrate the gel pieces in 25 mM ammonium bicarbonate pH 8.
7. Add 3 μ l of modified sequence grade trypsin at a concentration of 0.1 μ g/ μ l.
8. Incubate with agitation at 30°C overnight.
9. Transfer the supernatant to a new Eppendorf.
10. Extract the fragments using 10% acetonitrile/0.1% TFA for 10 minutes at 37°C.
11. Combine with the supernatant liquid from step 8.
12. Dry to approximately 10 μ l and then analyze.

[Adapted from *Anal. Biochem.* **224** (1995), 451–455 and the web site of the UCSF MS Facility.]

Table 4
Protein profiling with gel electrophoresis and LC/MS

	2D Gels with mass spectrometry ID	2D LC-MS/MS
Advantages	<p>Widely available and becoming more reproducible.</p> <p>Software available for differential expression as well as quantifying differences ($\pm 20\%$ at best).</p> <p>Easy visualization of up/down regulation demonstrated for highly complex mixtures</p>	<p>Good dynamic range especially for low expression level proteins. (Codon Adaptation Index, CAI, of 0–0.2) very difficult for 2D gels to look below $<CAI = 0.2$</p> <p>Good for a wide range of proteins with different hydrophobicities</p> <p>MS/MS experiments performed automatically for reliable identification.</p>
Disadvantages	<p>Problematic for proteins with extremes of PI and MW. Difficult for hydrophobic proteins.</p> <p>Sample handling extensive for MS analysis (stain/destain and extraction), multiple steps required and time consuming. Can be minimized with use of robots.</p> <p>Overlapping proteins possible</p> <p>Limited dynamic range</p>	<p>Computationally intensive. Database searches can take days. Cluster computer systems not easily available.</p> <p>Multi-dimensional protein identification technology (MudPIT) has the potential of coupling to quantitative approaches. Currently limited application shown for differential protein expression (relative quantitation) with 2D LC/MS/MS for highly complex mixtures.</p>

The combination of SDS-PAGE electrophoresis with an isoelectric focusing step also enables the separation of protein of similar mass. In two dimensional gel electrophoresis (2D-GE), proteins are first separated according to their isoelectric points by electrophoresis through a solution or gel containing an immobilized pH gradient, with each protein migrating to a position in the pH gradient corresponding to its isoelectric point. Once the isoelectric focusing step is complete, a secondary gel electrophoresis similar to SDS-PAGE is performed in an orthogonal direction to also separate the proteins by size. Like 1D gels, 2D gel spots can be cut out, enzymatically digested, and mass analyzed for protein identification. Using this technique, thousand of proteins can simultaneously be separated and identified. Additionally, this technique can help facilitate the analysis of certain PTMs. For example, differently phosphorylated forms of the same base protein may appear as a series of bands of roughly identical mass but different isoelectric points.

In addition to enabling the identification of thousand of different proteins, 2D-GE can also be used to assess large scale changes in protein expression levels between two different samples (i.e., healthy versus diseased samples). These protein profiling experiments (Fig. 5) rely on the fact that the chemicals used to visualize the separated protein bands produce responses roughly proportional to the total level of protein in the band. The experiments are typically performed by running 2D-GE on each of the two samples separately and comparing the resulting patterns. Proteins bands that appear in only one gel or that differ significantly in their intensity are excised and identified. Alternatively, the two samples can be treated with different visualization agents (i.e., two dyes with significantly different fluorescent emission spectra), combined, and run on the same gel.

Although historically difficult, the reproducibility of 2D-GE has improved with the availability of high quality pre-cast gels, immobilized pH gradients strips (IPG), sophisticated pattern recognition software, and laboratory automation. However, considerable limitations remain, including operational difficulty in handling certain classes of proteins, the co-migration of multiple proteins to the same position, and

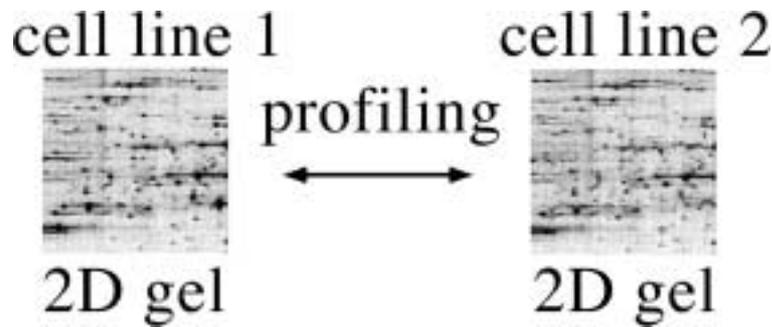


Fig. 5. Protein profiling can be performed by comparing the 2D gel from two different cell lines. The protein spot of interest is excised from the gel and an in-gel proteolysis of the protein is performed.

potential unwanted chemical modifications. An even greater potential shortcoming of the classic 2D-GE technique is its inability to accommodate the extreme range of protein expression levels inherent in complex living organisms due to sample loading restrictions imposed by the gel-based separation technology employed. Thus, 2D-GE separations often result in only the most abundant proteins being visualized and characterized. This limitation is of particular concern in that most interesting classes of regulatory proteins are often expressed at low copy numbers per cell.

2.2. High throughput protein ID with MALDI-MS

The ability to profile changes in the expression levels of thousands of proteins would be relatively meaningless without the ability to rapidly identify species of interest. To this end, automated liquid handling robots have been developed that perform all the sample preparation steps for peptide mapping experiments, including gel destaining, alkylation/reduction, in gel digestion, peptide extraction, and MALDI target plating. The benefits of such automation include less contamination during sample preparation, increased reproducibility, rapid protocol development, and the ability to prepare hundreds of proteins in the course of one day. Whereas manual preparation would require a full week to perform two-hundred analyses, a robotics station can complete the task in a matter of hours.

Mass spectral data acquisition systems have similarly been automated to acquire spectra, process the raw data, and perform database searches for numerous samples. Commercial MALDI-TOF systems are currently available that can perform 5,000 peptide mapping experiments in just twelve hours. These systems are able to perform automated calibrations, vary laser energies, and adjust laser firing location to maximize signal, with the entire data acquisition process requiring approximately 30 seconds or less. Similarly, automated data processing systems can recognize suitable signals, identify monoisotopic peaks, and submit summary peak list directly to a search engine.

Such high throughput proteomics systems enable researchers to investigate multiple unknown samples at once, whereas once it was too time consuming and costly to perform such analyses. Additionally, the flexibility of automated acquisition and data analysis software allows researchers to rapidly reacquire and/or reanalyze entire batches of samples with minimal user effort. Automated systems are however limited in that they are only as good as the programming given to them. For example, the detection and accurate mass assignment of species exhibiting low signal to noise ratios is often problematic. Such issues have led to a great deal of developmental work on post-acquisition data processing. Improvements

in these processes have enabled high throughput automated systems to achieve identification “hit” rates equal to or above those obtained manually.

2.3. Separation with liquid chromatography

2.3.1. Protein identification with liquid chromatography-tandem mass spectrometry

An alternative approach to gel electrophoresis techniques involves the use of analytical separation methods such as high performance liquid chromatography (HPLC). Although the rest of this chapter focuses specifically on liquid chromatographic techniques, it is important to note that the same advantages also apply to other separation methods such as capillary electrophoresis. Although fast and often effective for the identification of individual proteins, peptide mapping methods usually fail when dealing with more complex mixtures due to significant signal suppression as the sample becomes more complex. By contrast, LC-based methodologies fractionate the peptide mixtures before MS analysis, thus decreasing signal suppression and improving the analysis of any given peptide. More importantly, additional information can also be obtained on individual peptides by performing tandem MS experiments.

Whereas gel electrophoresis techniques separate intact proteins, liquid chromatography can be performed on both intact proteins as well as proteolytic peptides. However, the actual protein identification analysis is almost always performed using digested samples due to the reasons discussed earlier. One of the most popular means of performing peptide LC-MS/MS involves the direct coupling of the separation system to an ion trap mass spectrometer through an electrospray ionization interface. Other mass analyzers suitable for these experiments include triple quadrupoles, quadrupole time-of-flights, and quadrupole ion traps. However, ion traps remain the most popular because of their ease of use, relatively low cost, and rapid scanning capability that enables tandem mass measurements to be performed in real time. For example, the ion trap first performs MS measurements on all the intact peptide ions. Then, in a second scan, it performs a MS/MS experiment on a particular peptide ions detected in the first. This series of alternating scans can rapidly be repeated, with different ions selected for each tandem MS experiment. In this manner, single peptides from a complex mixtures can individually be addressed and analyzed.

Tandem MS experiments provide structural information for a given peptide by physically fragmenting it. This process is initiated by converting some of the kinetic energy of the peptide ion into vibrational energy, and is experimentally achieved by inducing the selected ion, usually an $(M + H)^+$ or $(M + nH)^{n+}$ ion, into physical collisions with neutral Ar, Xe, or He atoms. The resulting fragment ions are then monitored by MS. Fortunately, a large percentage of the fragment ions produced in this process result from cleavages along the linear backbone of the peptide, and can be separated into two major classes. One class retains the charge on the N-terminal fragment, and the ions types are designated as a_n , b_n , or c_n depending on the exact site of cleavage. The second class of fragment ions retains the charge on the C-terminal fragment, and are similarly designated as x_n , y_n , and z_n type ions. Of these species, the most frequently observed fragments are the b- and y-type ions that result from cleavage of the bond between the carbonyl carbon and the amide nitrogen (the amide bond). The differences in masses between the members of either the b- or y-ion series corresponds to the amino acid sequence of the fragmented peptide.

The additional sequence information provided by tandem MS experiments can be extremely powerful, sometimes enabling a definitive protein identification to be made on the basis of a single peptide. Obviously though, tandem MS spectra of multiple peptides that arise from the digestion of a given protein provides greater opportunity for obtaining a definitive identification. Generally, sequence information can be obtained for peptides with molecular masses up to 2500 Daltons. Larger peptides can reveal at least partial sequence information that often suffices to solve a particular problem.

Although powerful, tandem mass spectrometry has certain limitations with respect to its ability to obtain complete sequence information. For example, it is not possible to distinguish between leucine and isoleucine as they have the exact same mass. Similarly, lysine and glutamine can only be distinguished using high mass accuracy tandem analyzers, as they differ in mass by only 0.036 Daltons. Generally, complete ion series (y or b type) are usually not observed. However, the combination of the two series often provides more useful information and possibly the entire sequence. In addition, some amino acids as well as certain PTMs bias the fragmentation towards certain cleavages, dramatically decreasing the amount of sequence information obtained. Although chemical labeling techniques can partly compensate for these phenomena, it is important to note that not every peptide yields useful tandem MS spectra, thus further emphasizing the usefulness of attempting tandem MS spectra for multiple peptides arising from a given protein.

2.4. Protein identification with multi-dimensional liquid chromatography-tandem mass spectrometry (LC-MS/MS)

LC-MS/MS methodologies for protein identification have been extended to mixtures of even greater complexity by performing multi-dimensional chromatographic separations before MS analysis (Fig. 6). As its name suggests, extremely complex tryptic digests are first separated into a number of fractions using one mode of chromatography, and each of these fractions is then further separated using a different chromatographic method [14,15]. In theory, any combination of operationally compatible chro-

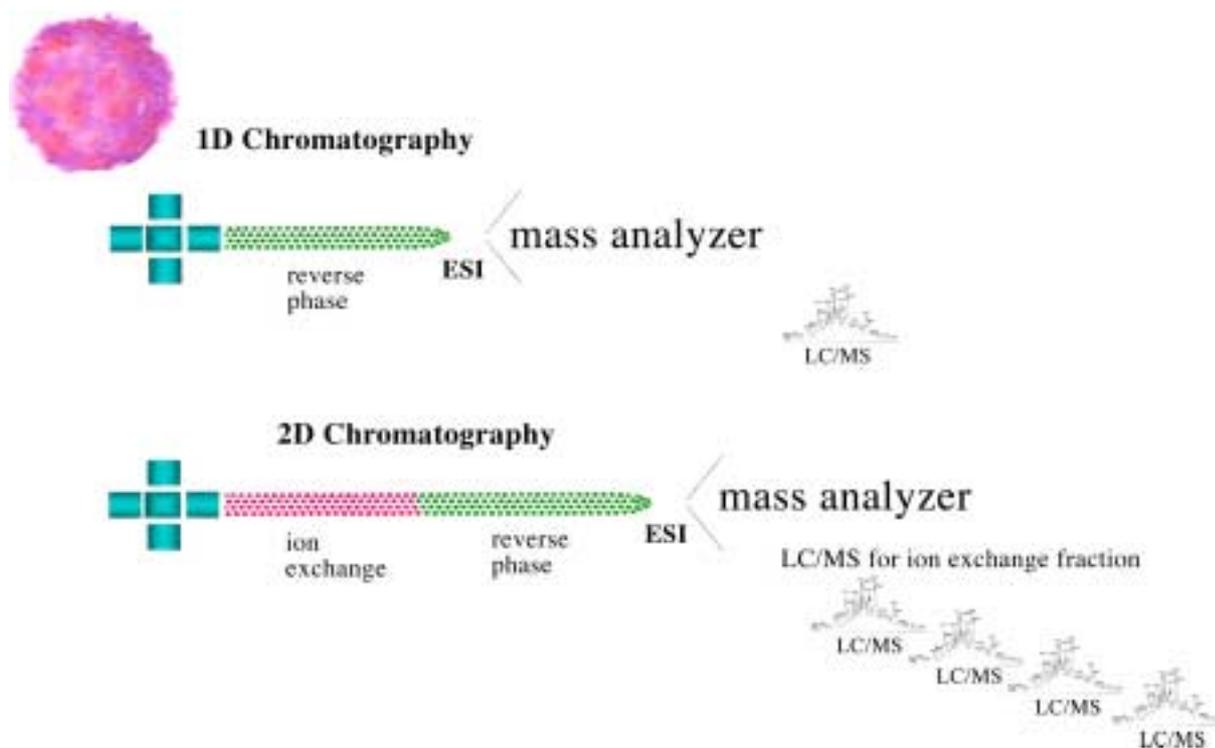


Fig. 6. The proteolytic peptides separated by liquid chromatography and 2D liquid chromatography are ionized using electrospray ionization and then subjected to tandem mass spectrometry (MS/MS) experiments. The data from the 1D experiments can analyze up to 200 proteins simultaneously while the 2D experiments are capable of handling thousands of proteins.

matographic methods possessing sufficiently orthogonal modes of separation can be utilized, and several different combinations have been described in the literature. However, the overwhelming number of studies to date have combined strong cation exchange (SCX) and reversed-phase (RP) chromatographies. More recently, further improvements have been realized by having both chromatographic beds in a single capillary column and directly coupling this column to an ion trap mass spectrometer. A step gradient of salt concentrations is used to elute different peptide fractions from the SCX resin onto the RP material, after which RP chromatography is performed without affecting the other peptides still bound to the SCX resin. The resulting nano-RP LC column effluent is directly electrosprayed into the mass spectrometer, making this method not only amenable to automation but also very sensitive. Using this “MuDPIT” methodology (Multi-Dimensional Protein Identification Technology), thousands of unique proteins have been identified from a whole cell lysate in a single 2D LC-MS/MS experiment. Additionally, recent studies have also indicated that this technique possesses a greater dynamic range than that obtained using 2D gel electrophoresis, enabling the detection of lower abundance proteins. However, one limitation slowing this methodology’s wide scale implementation is the massive computing power required to effectively compare and evaluate the statistical significance of similarities between the huge number of tandem MS spectra experimentally generated and the predicted fragmentation patterns of peptides resulting from the *in silico* digestion of all the entries in a given protein database.

2.5. Protein profiling with LC-MS/MS

Protein profiling studies can also be performed using multi-dimensional LC-MS/MS in conjunction with stable isotope labeling methodologies [16–18]. Specifically, two samples to be compared are individually labeled with different forms of a stable isotopic pair, and their tryptic digests are then combined before the final LC-MS analysis. This should result in every peptide existing as a pair of isotopically labeled species that are identical in all respects except for their masses. Thus, each isotopically labeled peptide effectively serves as its partner’s internal standard, and the ratio of the relative heights of two isotopically labeled species provides quantitative data as to any differential change that occurred in the expression of the protein from which the peptide arose.

One approach towards differential labeling involves growing cells in isotopically enriched media. For example, one set of cells would be cultured in media that contained ^{14}N as the only source of nitrogen atoms, while the comparative case would be grown in media that only contained ^{15}N . Although effective in incorporating different stable isotopes into the two samples, the determination of which two peptides comprise an isotopically labeled pair is severely complicated by the fact that each pair exhibits a different mass difference depending on the number of isotopic atoms incorporated. “Inverse labeling” methodologies have recently been introduced that cleverly address this issue at the cost of doubling the number of experiments that need to be performed. Ultimately though, this technique is limited in that it obviously cannot systemically be applied to higher organisms.

Alternatively, the recently introduced methodology of isotope-coded affinity tags (ICAT) (Fig. 7) provides a more generally applicable approach based on the *in vitro* chemical labeling of protein samples. Specifically, ICAT utilizes the high specificity of the reaction between thiol groups and haloacetyls such as iodoacetamide to chemical label cysteine residues in proteins with isotopically light or heavy versions of a molecule that differ only by the existence of eight hydrogen or deuterium atoms, respectively. The labeled protein samples are then combined and simultaneously digested, resulting in every cysteine-containing peptide existing as an isotopically labeled pair differing in mass by eight Daltons per cysteine

label and used in conjunction with a monomeric avidin column to affinity purify the cysteine-containing peptides. Alternatively, a solid phase capture and release strategy has more recently been described. Although these solutions enable the accurate identification of isotopically labeled peptide pairs, they obviously preclude the analysis of proteins that do not contain cysteines and also greatly reduce the number of opportunities to effect LC-MS/MS identification of the proteins that do contain cysteine residues.

2.6. LC-MALDI MS/MS: A potentially powerful paradigm

Although still relatively early in their development, these multi-dimensional LC-MS/MS-based approaches exhibit tremendous potential for the improved high-throughput identification and profiling of proteins from complex mixtures. However, the operational parameters of the ESI coupling methods overwhelmingly employed to date in their implementation also impose several limitations. Specifically, the separation system and mass spectrometer employed are coupled directly in real time, enabling tandem MS experiments to be performed on only a relatively small fraction of the species simultaneously eluting from the LC column. As a result, the list of proteins identified using this technology has been shown to vary significantly between repeat analyses of the same sample. More importantly, current instrument control and data analysis software is not sophisticated enough to allow real time data-dependent processing during the course of a chromatographic separation except when employing simple selection criteria such as peak intensity. This necessitates that upon the completion of a separation and subsequent analysis of the resulting data, the same sample must often be rerun to focus on those species that exhibited the desired selection criteria but were not randomly subjected to tandem MS.

In light of these considerations, several groups have begun to explore MALDI-based LC-MS/MS strategies that involve the creation of a “permanent record” of the multidimensional separation by depositing the effluents of the final separation columns directly onto MALDI target plates (Fig. 8) [19]. De-coupling the separation step from the mass spectrometer in this manner enables more thorough analyses of samples to be performed due to the removal of artificially imposed time restrictions. The resulting plates can also be reanalyzed as required without the need to repeat the separation step, thus decreasing sample requirements while focusing system resources only on the acquisition of tandem MS spectra of species of interest. Although more difficult to effectively implement than ESI-based methods, the commercial introduction of mass analyzers such as the MALDI-QTOF promise to further speed the development of MALDI-based LC-MS/MS platforms.

2.7. Protein identifications using extremely accurate mass measurements

Further improvements in mass analyzer performance should also enable new approaches for the effective identification of proteins [19,20]. For example, Fourier transform ion cyclotron resonance (FT-ICR) mass spectrometers can measure the masses of proteins and/or peptides with mass accuracies of 1 ppm or better. When combined with chemical labeling techniques to determine the number of specific amino acids contained in a given species, these extremely high mass accuracy can be used to unequivocally identify a protein from a single peptide without performing tandem MS. Although intriguing, this technique requires significant further development in order to demonstrate its general applicability.

3. Conclusion

Both MALDI and LC-MS/MS are playing important roles in protein identification and protein profiling. MALDI offers many advantages in terms of speed and ease of use for protein identification whereas

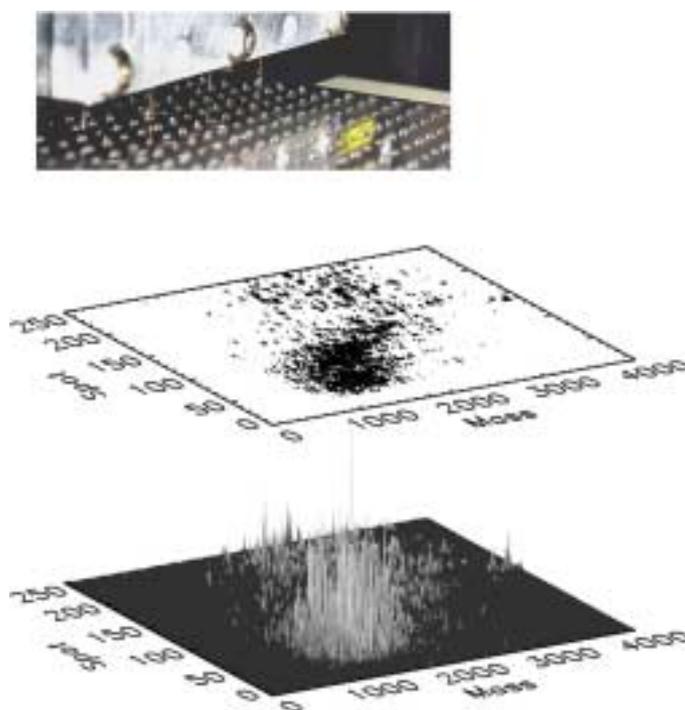


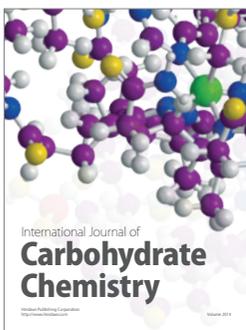
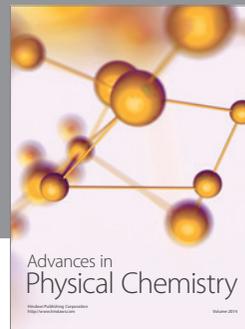
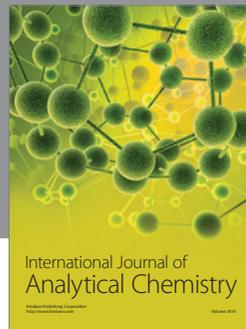
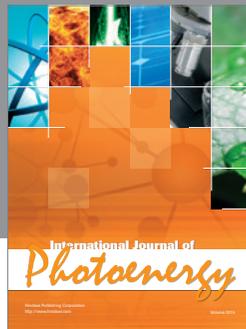
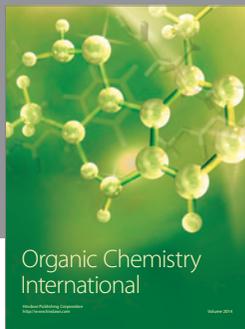
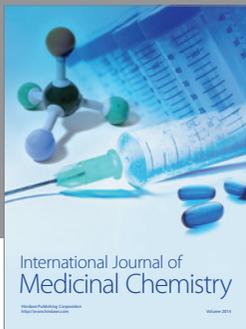
Fig. 8. An example of HPLC MALDI-MS. (Top) Four μ HPLC columns performing parallel deposition on 384 microtiter plate format for analysis by MALDI. (Middle & bottom) Three-dimensional plot of the reversed-phase μ HPLC–MALDI FT-ICR MS analysis of a tryptic digest of yeast proteins.

LC-MS/MS offers a more reliable protein ID as well as a greater potential for post-translational modification identification. MALDI is also finding utility in its application to traditional protein profiling with gels and protein profiling while LC-MS/MS appears to be a tool of the future with respect to quantitative analysis with isotope labelling. Finally, LC-MS/MS capabilities are not only limited to electrospray, with new approaches coupling LC to MALDI offer the potential of rapid analysis and highly accurate mass measurements. LC/MALDI also can provide a platform that offers re-analysis of samples.

References

- [1] A. Pandey and M. Mann, Proteomics to study genes and genomes, *Nature* **405** (2000), 837–846.
- [2] S. Broder and J.C. Venter, Sequencing the entire genomes of free-living organisms: the foundation of pharmacology in the new millennium, *Annual Rev. Pharmacol. Toxicol.* **40** (2000), 97–132.
- [3] J.R. Yates, 3rd, Mass spectrometry: From genomics to proteomics, *Trends in Genetics* **16** (2000), 5–8.
- [4] R. Aebersold and D.R. Goodlett, Mass spectrometry in proteomics, *Chem. Rev.* **101** (2001), 269–295.
- [5] J.B. Fenn, M. Mann, C.K. Meng, S.F. Wong and C.M. Whitehouse, Electrospray ionization for mass spectrometry of large biomolecules, *Science* **246** (1989), 64–71.
- [6] M. Karas and F. Hillenkamp, Laser desorption ionization of proteins with molecular masses exceeding 10,000 Daltons, *Anal. Chem.* **60** (1988), 2299–2301.
- [7] T. Voss and P. Haberl, Observations on the reproducibility and matching efficiency of two-dimensional electrophoresis gels: consequences for comprehensive data analysis, *Electrophoresis* **21** (2000), 3345–3350.
- [8] S.D. Patterson and R. Aebersold, Mass spectrometric approaches for the identification of gel-separated proteins, *Electrophoresis* **16** (1995), 1791–1814.
- [9] S.P. Gygi, G.L. Corthals, Y. Zhang, Y. Rochon and R. Aebersold, Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology, *Proc. Natl. Acad. Sci. USA* **97** (2000), 9390–9395.

- [10] J. Peng and S.P. Gygi, Proteomics: the move to mixtures, *J. Mass Spectrometry* **36** (2001), 1083–1091.
- [11] G.L. Corthals, V.C. Wasinger, D.F. Hochstrasser and J.C. Sanchez, The dynamic range of protein expression: a challenge for proteomic research, *Electrophoresis* **21** (2000), 1104–1115.
- [12] D.B. Wall et al., Isoelectric focusing nonporous RP HPLC: a two-dimensional liquid-phase separation method for mapping of cellular proteins with identification using MALDI-TOF mass spectrometry, *Anal. Chem.* **72** (2000), 1099–1111.
- [13] S.E. Martin, J. Shabanowitz, D.F. Hunt and J.A. Marto, Subfemtomole MS and MS/MS peptide sequence analysis using nano-HPLC micro-ESI fourier transform ion cyclotron resonance mass spectrometry, *Anal. Chem.* **72** (2000), 4266–4274.
- [14] M.P. Washburn, R. Ulaszek, C. Deciu, D.M. Schieltz and J.R. Yates 3rd, Analysis of quantitative proteomic data generated via multidimensional protein identification technology, *Anal. Chem.* **74** (2002), 1650–1657.
- [15] D.A. Wolters, M.P. Washburn and J.R. Yates 3rd, An automated multidimensional protein identification technology for shotgun proteomics, *Anal. Chem.* **73**(23) (2001), 5683–5690.
- [16] S.P. Gygi et al., Quantitative analysis of complex protein mixtures using isotope-coded affinity tags, *Nature Biotechnology* **17** (1999), 994–999.
- [17] S. Wang and F.E. Regnier, Proteomics based on selecting and quantifying cysteine containing peptides by covalent chromatography, *Journal of Chromatography A* **924** (2001), 345–357.
- [18] E.C. Peters, D.M. Horn, D.C. Tully and A. Brock, A novel multifunctional labeling reagent for enhanced protein characterization with mass spectrometry, *Rapid. Commun. Mass Spectrom.* **15** (2001), 2387–2392.
- [19] E.C. Peters, A. Brock, D.M. Horn, Q.T. Phung, C. Ericson, A.R. Salomon, S.B. Ficarro and L.M. Brill, An automated LC-MALDI FT-ICR MS platform for high-throughput proteomics, *LCGC Europe* (July 2002), 2–7.
- [20] S.E. Martin, J. Shabanowitz, D.F. Hunt and J.A. Marto, Subfemtomole MS and MS/MS peptide sequence analysis using nano-HPLC micro-ESI fourier transform ion cyclotron resonance mass spectrometry, *Anal. Chem.* **72** (2000), 4266–4274.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

