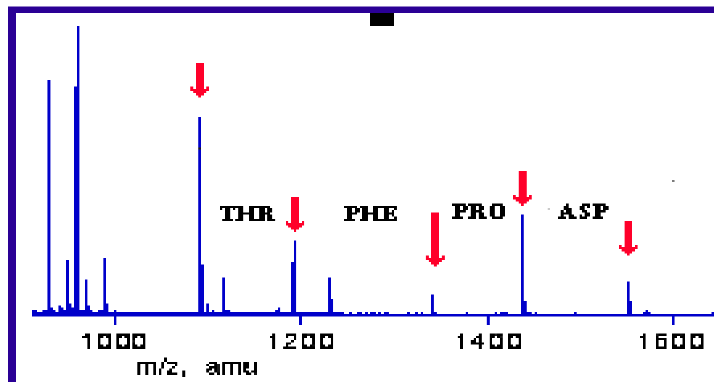
Lecture 9: Mass Spectrometry for Proteins

Microbiology 343

David Wishart, University of Alberta, Edmonton, AB

Mass Spectrometry

Mass spectrometry is an analytical technique for determining the mass (or atomic weight) of molecules. Mass spectrometers use a variety of techniques to create charged ions or charged ion fragments and separate these ions on the basis of their charge-to-mass (m/z) ratios. The ionic separation is done by using a combination of magnetic or electric fields. The charged ions are detected by the miniscule current they create when the hit a detector. The resulting mass spectrogram is a series of peaks, with each peak corresponding to a different ion with a different mass-to-charge ratio.



The intensity of the peaks is proportional to the number of ions with the corresponding m/z ratio. By analyzing the patterns of peaks seen in a mass spectrum it is possible to determine the mass of the parent ion (the original molecule) and to determine something about its chemical structure.

Over the past 10 years, mass spectrometry (MS) has largely overtaken microsequencing as a way of rapidly identifying peptides and proteins. Relative to microsequencing, mass spectrometry offers the advantages of higher throughput and greater sensitivity without the high costs (outside of the initial instrument purchase) and labour. These advantages to mass spectrometry have largely come about from a number of recent technical improvements including more robust high vacuum systems, better electronics, more sensitive detectors, more efficient ionization techniques and better methods for separating or filtering ions.

Mass spectrometry can be used in protein identification because it allows one to measure protein masses very precisely (< 0.1 dalton) using nanograms of material. For proteins, precise molecular weights are almost as distinctive as a person's name or phone number. They're kind of a molecular signature or calling card. The fact that protein masses can be predicted exactly from the amino acid sequences means that one can identify a protein purely on the basis of its sequence-predicted mass. Certainly as more and more protein sequences are deposited in databases like GenBank, it is becoming increasingly likely that proteins can be unambiguously identified on the basis of their spectrometrically measured molecular weights. Of course, just as with human names there can often be two or more proteins that share the same mass but have completely different sequences. Similarly because many proteins are post-translationally modified, their mass signatures will not always correspond to their sequence-predicted masses. In these cases it is necessary to use another technique -- called peptide mass fingerprinting.

Mass Spectrometry and Peptide Mass Fingerprinting

Mass spectrometry allows one to not only measure the mass of pure peptides and proteins, but it also allows one to measure the masses of mixtures of peptides and proteins. If a single pure protein is digested by an enzyme to generate peptides, the resulting mass spectrum will lead to a peptide-mass fingerprint that is far more unique signature than the mass of the pure protein alone. By correlating the collection of peptide masses generated from such a digestion with calculated masses derived from a sequence database one can very confidently ascertain the identity of the target protein. Peptide-mass fingerprinting is simply a method to identify proteins already contained within a sequence database using an algorithm to match a set of peptide-masses generated using specific cleavage reagents (either enzymatic or chemical) from the protein of interest with theoretical peptide masses calculated from each sequence entry in the database (which has been cleaved theoretically with the same cleavage agent. A ranking (or score) is then calculated to provide a measure of the fit between the observed and expected peptide masses. So to summarize, peptide mass fingerprinting involves the following steps:

1. Separate proteins by 1D or 2D gel electrophoresis or other suitable separation technique.
2. Cut the protein out from the gel and digest it with a protease (trypsin).
3. Separate the peptides generated by protease digestion, either by HPLC or CE

4. Inject the peptides into a mass spectrometer and determine the mass of each peptide.
5. If necessary, use MS/MS and CID to break each significant peptide ion into sub-fragments.
6. Take all of the peptide fragmentation data and use it to search protein databases to find the best match that would generate the same theoretical fragmentation pattern.
7. Interpret the data and decide if the best match generated by the computer is relevant

Obviously if the protein sequence is not represented in any database then there is no chance of identifying it using peptide mass fingerprinting. As a rule it is best to have data for at least 3 or more peptide fragments to have reasonable certainty about any given match. In other words, the more peptides one can identify, the better the chance one has of successfully identifying the protein.

The success of peptide mass fingerprinting is critically dependent on the specificity of the proteases used in the cleavage reactions. Trypsin, the most commonly used reagent for peptide-mass fingerprinting, specifically cleaves at sites C-terminal to Lys or Arg, (if not followed by a Pro). However, like most proteolytic enzymes, trypsin can cleave at other sites and it may not cleave its substrate to completion. Nonspecific cleavage is quite difficult to accommodate in most peptide mass search algorithms. In addition, trypsin will also autodigest. This can sometimes be useful as the trypsin derived peptides can be used for internal mass calibration. However, trypsin-derived peptides, if present in large excess over the peptides of interest, can suppress peaks or obscure target peptides by overlapping in the spectrum. Endoproteinase LysC (Lys-X, some Asn-X) is another frequently used enzyme for peptide-mass searching. While it can also miss cleavage sites and cleave nonspecifically, it has far fewer autolysis products than trypsin.

Mass Spectrometry and Peptide Sequencing

Mass spectrometry not only allows one to measure masses of pure and impure polypeptides, it also allows one to sequence them as well. This is where MS outshines microsequencing and it is also where unambiguous identification of proteins can be complete. Peptide sequencing by MS is normally done using tandem mass spectrometers (MS/MS instruments). A tandem mass spectrometer combines two mass analyzers in tandem with a CID cell. Tandem MS is the equivalent of 2D electrophoresis or 2D NMR. With a tandem MS, a particular m/z value (or ion collecion) can be selected or filtered away from all other ions entering a mass analyzer. Once isolated, this ion can be disassociated (in the CID) and the m/z values of the dissociated products can be

determined by the second mass analyzer. In the case of peptides, most fragmentation events or covalent bond breakage occurs around the amide bond. This produces a ladder of fragment ions that can be indicative of the amino acid sequence of a given peptide (isolated from the first mass analyzer).

A second approach to MS sequencing is to use MALDI-TOF and carboxypeptidase or amino-peptidases (so-called ladder sequencing). In this case, the protein of interest is mixed with the peptidase and the resulting cleavage products are isolated by HPLC or CE. The mass of each protein fragment (which may be missing 0, 1, 2 or more residues from the N or C terminal end) is determined and ranked. In this way the identity and sequence of the first (or last) few amino acids in the protein can be determined. The one limitation of MS sequencing is that it cannot identify amino acid mass isomers such as Asn and Asp or Gln and Glu. Nevertheless, peptide sequencing by MS is a powerful approach that is increasingly being used in proteomics applications.

Mass Spectrometry - Instrumentation and Glossary

Mass spectrometers consist of three key components: 1) an ionization source which converts molecules into gas-phase ions; 2) a mass analyzer which separates ions according to their mass-to-charge rations; and 3) an ion detector which converts the passage of charged ions through a medium into detectable currents or signals. Different types of spectrometer designs exist for different purposes. Some are more suitable for studying small molecules or for identifying isotopic abudances, others are more suitable for studying large biological molecules. The key development that has made mass spectrometry relevant to proteomics has been the development of soft ionization techniques to create large molecular ions (m/z ratios > 2000). These techniques are known as ESI, MALDI and FAB.

Electrospray ionization (ESI) - ESI is a soft ionization technique for generating multiply charged ions. The ESI source consists of a very fine needle and a series of skimmers. A sample solution is sprayed from the needle at high voltage into a source chamber to form droplets. The droplets carry charge when they exit the needle and, as the solvent evaporates, the droplets disappear leaving only the highly charged ions behind. This process is called nebulization. ESI is particularly useful for large biological molecules that are difficult to vaporize or ionize.

Fast-atom bombardment (FAB) - In FAB a high-energy beam of neutral atoms, typically Xenon or Argon, strike a solid sample causing desorption and ionization. FAB

causes little fragmentation and usually gives a large molecular ion peak, making it useful for molecular weight determination. The atomic beam used in FAB is produced by accelerating ions from an ion source though a charge-exchange cell. The ions pick up charges from collisions with neutral atoms to form a beam of high energy atoms. Like ESI, FAB is used for large biological molecules that are difficult to get into the gas phase.

Matrix-Assisted Laser Desorption Ionization (MALDI) - In MALDI a UV laser pulse is used to vaporize or ablate material from the surface of a sample. This creates a plasma that ionizes some of the sample constituents. The laser pulse accomplishes both vaporization and ionization of the sample. In MALDI the matrix typically consists of a UV absorbing material such as nicotinic acid. MALDI is an excellent method for vaporizing and ionizing large biological molecules such as proteins or nucleic acids.

Just as there are different types of ionization techniques, so too are there different types of mass analyzers and mass filters. Some of the most common mass analyzers are TOF and magnetic sector instruments. Among the common mass filters are quadrupoles and tandem mass spectrometers equipped with CID (collision induced decay) chambers.

TOF - A time-of-flight (TOF) mass analyzer uses the differences in transit time through a drift tube (a hollow metal tube under extremely high vacuum) to separate ions of different m/z ratios. The time required for ions to travel a set distance and strike a detector is used to calculate the m/z ratio seen in the spectrum. TOF systems operate best in a pulsed mode so ions must be produced or extracted in pulses, therefore TOF instruments are most compatible with matrix-assisted laser desorption (MALDI) or electrsopray ionization (ESI) sources. To understand the theory behind TOF, recall that when an electric field is applied to a free ion, it will give the ion a kinetic energy of zV, where z is the ion charge and V is the applied voltage.

4) Kinetic energy = $1/2mv^2 = zV$

Inverting the above equation, we get:

5) $V = (2zV/m)^{1/2}$

The transit time (t) through the drift tube is L/V where L is the length of the drift tube.

6) $t = L / (2V/m/z)^{1/2}$

As can be seen from (6) lighter ions will reach the detector at the end of the drift tube sooner than heavier ions. Therefore TOF instruments are better for measuring the masses of heavier ions (such as those found in proteins). By making the drift tube longer it is possible to increase the time of flight and improve the precision of the mass measurement.

Magnetic Sector - In magnetic sector instruments magnetic fields are used to deflect the ions according to their mass-to-charge ratio. The radius of curvature (instead of the time of flight) is used as a measure of the mass/charge ratio. Only ions with mass-to-charge ratios that have equal centrifugal and centripetal forces pass through the flight tube:

7) $mv^2 / r = Bzv$

Where r is the radius of curvature of the ion path, B is the magnetic field and v is the velocity:

8) $r = mv / zB$

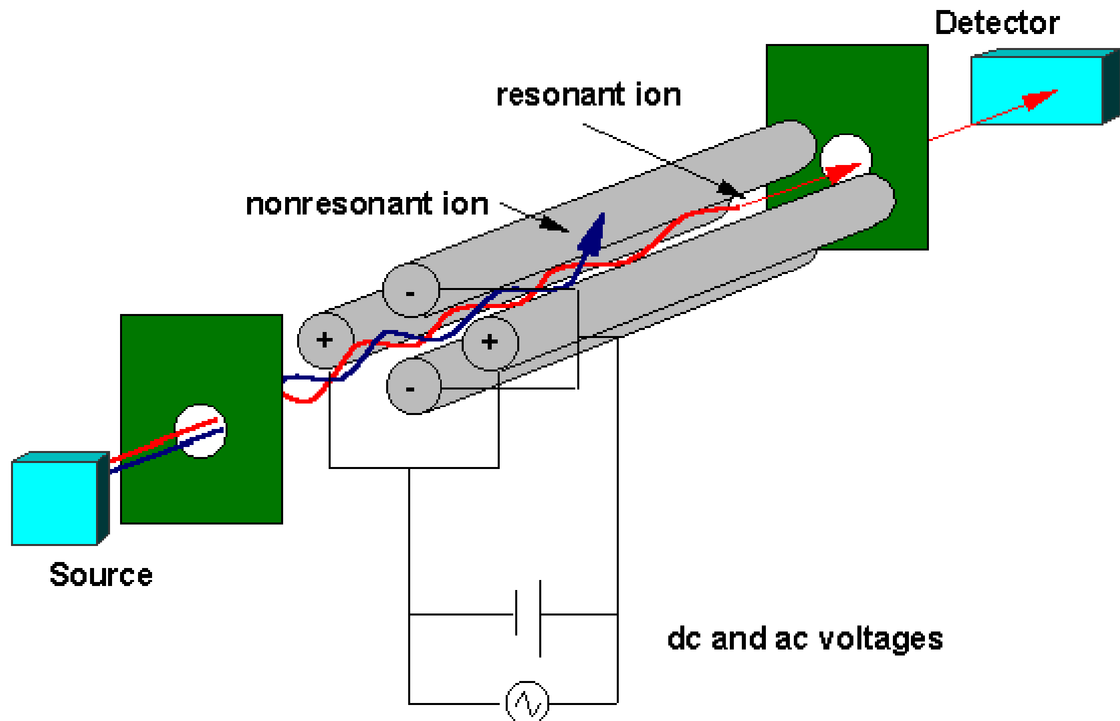
Recall that the kinetic energy of the ions entering the flight tube is: $1/2mv^2 = zV$

9) $m/z = B^2r^2/2V$

Therefore the m/z ratio is a function of the magnetic field and the applied voltage, both of which can be varied. In magnetic sector instruments a circular beam path of 180, 90, or 60 degrees can be used.

Quadrupole Filter - A quadrupole mass filter consists of four parallel metal rods arranged in parallel. The rods have opposing or out-of-phase time-varying (AC) voltages. These voltages affect the trajectory of ions traveling between the four rods. For given a

given set of voltages, only ions of a certain mass-to-charge ratio pass through the quadrupole filter and all other ions are thrown out of their original path. A mass spectrum is obtained by monitoring the ions passing through the quadrupole filter as the voltages on the rods are varied. Quadrupole filters reduce the background noise and generally improve the precision of mass measurements. They are frequently coupled with TOF mass analyzers.



Misccellaneous Mass Spectrometry Terms

MALDI-TOF - A Mass spectrometer with a MALDI ionization source and a time-of-flight mass analyzer. A MALDI-TOF will typically cost $300,000-400,000

MALDI-QTOF - A Mass spectrometer with a MALDI ionization source, a time-of-flight mass analyzer and a quadrupole mass filter. MALDI-QTOF systems typically cost $800,000.

LC-MS - A Mass spectrometer (usually an ESI instrument) fitted with an HPLC system. LC-MS systems can be purchased for $250,000.

LC-MS/MS - A Tandem mass spectrometer fitted with an HPLC system. An LC-MS/MS system can cost in excess of $500,000.

ESI-MS - A Mass spectrometer fitted with an ESI ionization source and a magnetic sector mass analyzer. ESI-MS systems can be purchased for under $200,000.

TQ-MS - A Triple qudrupole mass spectrometer contains three quadrupoles, termed Q1, Q2, Q3. The first (Q1) and last (Q3) quadrupoles serve to filter ions of specific m/z ratio by scanning the potentials applied to the quadrupole rods. The second quadrupole (Q2) serves as the collision cell (for CID) and is constructed to transmit ions without selection. Q1 is a conventional quadrupole analyzer, Q2 is the equivalent of a collision cell and contains an inert collision gas and Q3 records the m/z ratios of the fragment ions that originate from the fragmentation of the precursor ion selected by Q1. The advantage of the TQ-MS for acquiring MS/MS spectra is the fine control that the user can maintain over the entire process

CID - Collision Induced Disassociation is a method for dissociating ions through collisions between ions and inert gas atoms (usually Argon). CID normally occurs in a gas-phase collision cell that is conducive to low energy (10-50 eV) collisions. At these energies covalent bonds are broken which can lead to structurally informative fragment ions. In the case of peptides, CID can be used to determine sequences directly from MS data.

Tandem Mass Spectrometer (MS/MS) - A tandem mass spectrum combines two mass analyzers in tandem with a CID cell. Tandem MS is the equivalent of 2D electrophoresis or 2D NMR. With a tandem MS, a particular m/z value (or ion collection) can be selected or filtered away from all other ions entering the mass analyzer. Once isolated, this ion can be disassociated (in the CID) and the m/z values of the dissociated products can be determined by the second mass analyzer. In the case of peptides, most fragmentation events or covalent bond breakage occurs around the amide bond. This produces a ladder of fragment ions that can be indicative of the amino acid sequence.

Monoisotopic mass - the mass of the most abundant isotope for a certain element. The monoisotopic masses of peptides are the masses calculated from the mass of the most abundant isotopes.

Delayed-extraction/time-lag focusing ion sources - these are ionization techniques for MALDI that have been found to significantly improve the resolution

Mass-to-charge ratio (m/z) -This is the ratio of the mass of the ion to its charge. Ions can be multiply charged (1, 2, 3 or more charges). In protein mass spectrometry a peptide of molecular weight of 1400.1 daltons will have an m/z ratio of 1401.1 after ionization through the addition of one proton (M + H). With the addition of 2 protons $(M + 2H)^{2+}$ the m/z value is 701.1

Resolution - The ion separation power of a mass spectrometer is called the resolution, which is defined as: $R = m / \Delta m$, where m is the ion mass and Δm is the difference in mass between two resolvable peaks in a mass spectrum. For instance, a mass spectrometer with a resolution of 10 ppm can resolve an ion with a m/z of 1000.00 from an ion with an m/z of 1000.01.

Mass Spectrometry without 2D Gels

MS analysis of proteins is not limited to sample separated by PAGE. Essentially any separation method is suitable, provided that the peptides can be isolated in, or transferred into a solvent which is compatible with MS analysis. Recently, (McCormack et al. 1997) it has been shown that protein identification is feasible directly from proteolytic products of protein mixtures. The method was successfully used for the identification of components in several protein-protein complexes and immunoprecipitated complexes. All the proteins in the sample were co-digested and the resulting complex peptide mixture was separated by HPLC and analyzed by ESI-MS. Since the many peptides generated by the digestion of a relatively simple protein complex could easily overwhelm the resolution of a one-dimensional chromatography system, Link et al. (1998) have extended this approach to include multidimensional peptide separation by column chromatography. In this approach two chromatography systems with orthogonal separation properties are connected on-line with a tandem MS. Compared to single dimensional peptides separation this approach has several advantages which include simplified sample handling, process automation and increased number of identified proteins.

A second approach to studying protein mixtures without 2D gels is to use isotope-coded affinity tags (ICATs) developed by Rudy Aebersold (Gygi et al., 1999).

ICAT is a biotin-linked isotope labeled affinity tag. The biotin moiety is modified so that it reacts specifically with cysteine. Two types of ICATs are prepared, one with deuterated groups and the other with protonated groups. To study a mixture of proteins using ICAT and MS the protein mixture is first split into two groups and then reduced. The first group is reacted with the detuerated ICAT and the second is reacted with the protonated ICAT. The two sets of proteins are then mixed together and trypsinized. The peptide mixture is then passed over an avidin matrix and washed. Only the peptides with ICAT tags (only those peptides with cysteines) will remain attached to the matrix. This filtering step removes 98-99% of the peptides in the mixture, leaving only one or two peptide fragments that are characteristic of a single protein. This filtration step not only simplifies the peptide mixture it also simplifies the resulting MS spectrum. In fact, the resulting mass spectrum typically resembles the standard mass spectrum of a single pure protein that has been proteolyzed by trypsin. The use of deuterated/protonated biotin tags also allows one to quantitate the peaks and, indirectly, to measure the abundance of the proteins. In other words, ICAT allows one to identify as well as quantify protein mixtures without the need for gel electrophoresis or liquid chromatography. This technique is particularly sensitive (more so than 2D gels), fast and simple and it could greatly improve the process of protein identification in proteomics.

References

Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol.* **17**:994-999 (1999).

Li L, Golding RE, Whittal RM. Analysis of single mammalian cell lysates by mass spectrometry *J. Am. Chem. Soc.* **118**:11662-11663 (1996).

Link AJ, Eng J, Schieltz DM, Carmack E, Mize GJ, Morris DR, Garvik BM, Yates JR 3rd. Direct analysis of protein complexes using mass spectrometry. *Nature Biotech*. **17**:676-682 (1999).

McCormack, AL et al. Direct analysis and identification of protein in mixtures by LC/MS/MS and database searching at the low femtomole level. *Anal. Chem.* **69** 767-776 (1997)