Lecture 3 Mass Spectrometry

Principles of Mass Spectrometry

Mass spectrometry separates molecules based on the mass to charge ratio. This is often written as the m/z ratio. The limit for the technique is 1.5-2 kDa/unit charge. The separation chamber of the mass spectrometer is keep under a high vacuum and so all samples must be ionized to give a charge and be in the gas phase. For proteomics there are two major methods for ionization. The first is matrix assisted laser desorption ionization (MALDI).

Let's begin our discussion with ESI. The main advantage of this is that samples can be in solution. The other important factor is that multiple charging can occur allowing the measurement of the mass of large proteins and even complexes. This method has been used to look at protein-protein interactions as well as identify the mass of multisubunit complexes. Typically the system is run with a reverse phase HPLC to separate the various subunits or peptide fragments for identification. The HPLC system used now is call a ultra high pressure liquid chromatography (UHPLC). Flow rates are in microliters per minute and volumes are a few microleters. When ionization occurs most the ions do not even enter the mass spectrometer. Still the sensitivity is quite good and can allow the identification of proteins from whole cell digests. The main disadvantage is that not all molecules will ionize and so no matter what you do some information will be lost.

The ESI system is usually a HPLC system that is coupled to a nozzle. In positive mode ESI the nozzle is kept at a high positive voltage relative to ground. Small droplets are sprayed out of the end of the nozzle into a region with in from of the mass spectrometer inlet. As the droplets travel towards the vacuum the water molecules start to evaporate creating a more positive charge on the surface. At some point the charges on the surface get so great that they blow apart the droplet into smaller droplets. This continues until we end up with only one or a few proteins in a droplet which finally dries leaving the charges on the protein. If salt is present in the solutions adducts of the positive ions with the molecule can be observed making identification more difficult.

As I said multiple charging occurs with the protein. Here is a spectrum of cytochrome c that has been aquired through ESi-MS. Each peak represents the same molecule with different numbers of charges. Since each peak represents the addition of one charge, formulas can be worked out to determine the parental mass from the masses of the multiply charged ions.

For an example let's take the two lowest masses 1766.6 and 1545.7. These represent the mass divided by N and the mass divided by N+1.

$1766.6 = M/N$ $\qquad N = M / 1766.6$

$1545.7 = M/(N+1)$

$$1545.7 = \frac{M}{\left[\frac{M}{1766.6} + 1\right]} \qquad \left[\frac{M}{1766.6} + 1\right] 1545.7 = M$$

$$M\left[\frac{1545.7}{1766.6}\right] + 1545.7 = M \qquad 1545.7 = M\left[1 - \frac{1545.7}{1766.6}\right]$$

$$\frac{1545.7}{\left(1 - 1545.7 / 1766.6\right)} = M = 12361$$

This is done for all of the peaks and the best fit is obtained. There are programs that take into account all of the various peaks and create the singly charged spectrum. As I said this can be done with large complexes as long as they can be brought into solution as pure complexes and that they ionize.

In general the ion charge is determined by the isotope series. Since each protein is made up of many carbon and nitrogen atoms there is a probability that stable isotopes (C^{13}, N^{15}) will be incorporated into the molecule. This instrument is sensitive enough that for many molecules a single mass unit difference is quite easy to detect. The spacing between the isotope series will tell one the charge. If the spacing is one mass unit then the charge is 1. If it is ½ a mass unit the charge is 2 and so forth.

MALDI

In MALDI samples are mixed with a "matrix" and placed on a "target". The sample is dried down and then placed into the vacuum area of the MS and a UV or IR laser is shot at the target causing the ionization of the sample. The ions are then analyzed by the MS. The advantage of MALDI is that many samples can be spotted and put into the machine and it can automatically measure the MS of each spot. Also, since it is a laser, it actually requires less maintenance than the UHPLC. The next slide has a picture of an automated target sampler coming from a UHPLC.

The types of matrices differ greatly and are used for different types of samples. This is shown on the next slide. All of these have at least a single aromatic ring to absorb the UV light. Three of them are acids and so can be easily charged and the other two have multiple OH groups which can also be ionized.

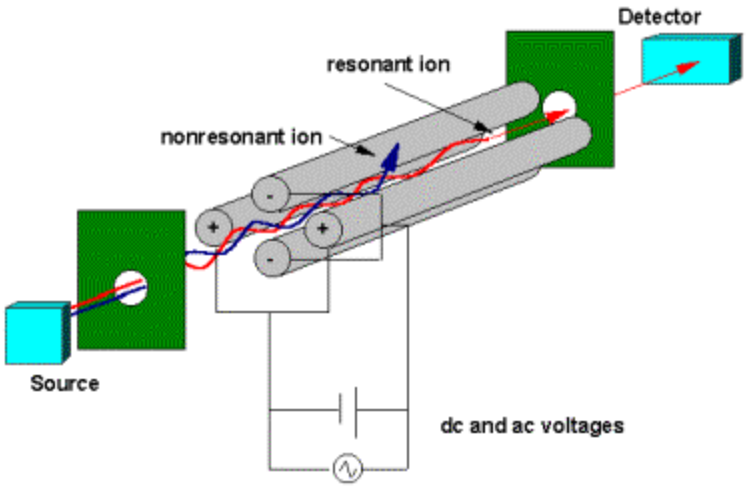
Mass Analyzers

The other part of the instrument is the mass analyzer. This is the part that determines the mass of the ion. This is accomplished by accelerating the ion and then sweeping a magnetic field across the accelerated ions. Since the ions are charged they

will bend in a magnetic field. Since the molecules have all been accelerate to the same velocity the angle it takes in the magnetic field will depend on its mass. By sweeping the mass we can change which ions are detected at the detector. Several different mass analyzers have been developed and each has some advantages.

Quadrapole

The quadrapole analyzer uses four rods with that are either + or – ve charge and an AC voltage to cause the ions to basically go in a spiral path through the rods until is reaches the detector. The ions that get all the way through have a mass that allows them to remain within the four poles until it reaches the end. By varying the frequency of the AC we can select different ions. This type of analyzer can be only two to three feet long.



The quadrapole mass spectrometer as you can see has four parallel rods. To these rods is applied a alternating voltage with the function below:
(U+Vcos(wt)) for +
-(U+Vcos(wt)) for -

U is a dc voltage
Vcos(wt) is an ac voltage
w is the frequency

These voltages change the path of the ions and at a given dc and ac voltage only ions with a certain m/z+ will pas through the filter. To get a mass spectrum the voltages are varied and the ions that pass through are measured. This can be accomplished by varying w and holding U and V constant. Mass resolution can be increased by putting quadarpoles in series or using an octapole.

Time of Flight

The time of flight analyzer is the most basic and the first designed. The ions are accelerated using an electric grid. Those ions that have a larger mass will move slower than those with a smaller mass since they were all accelerated with the same force. The distance between the various groups of ions will depend on the path length for its resolution. Therefore these can be very long. As the ions are accelerated some of the ions of the same mass will be accelerated less and some more depending on how far they were from the screen used to accelerate them. Therefore the ions that are closer to the screen will come out sooner but will be slower. Those that are farther from the screen will come out later but will move faster. At some point the ions will all get together and this is called the focal point. As they pass through the focal point the ions will tend to spread out again. To increase the resolution without increasing the instrument length, a reflector is put at the end and it bounces back to another detector. The reflector is made up of a series of grids. When the ions enter the reflector the faster ones will enter first and go farther into the reflector than the slower ones. The slower ones will enter later

and be reflected and exit the reflector before the faster ones. So now the faster ones are behind the slower ones and if things are set up right they will catch up to the slower ones just before they reach the detector. This then improves the resolution by narrowing the packet of ions.

Ion trap

This is the smallest of the mass analyzers and will be 4-6 inches wide. These are also the lightest of the instruments. Mass spectrometers were sent up in early satellites and balloons to try and analyze the molecules present in those environments. This consists of a ring and two end cap electrodes. By using both constant and varing electric fields the ion trap can accumulate a large number of different ions and store them. By changing the electric fields the accumulated ions can then be released as packets of different m/z values. Due to the accumulation the sensitivity of the instrument is increased over other instruments. The mass resolution was increased by filling the cavity with 1mTorr of helium. This instrument has a wide m/z range of 670 to 70,000 which is much wider than most other instruments. The main limitation of the instrumentation has been the mass resolution. The resolution is limited by how well you can control the AC and DC voltages to separate the ions for release from the ion trap.

FTICR-MS

This is an amazing machine due to all the things that are going on in it. The ions are generated by one of the two methods discussed before. These ions enter a series of quadrapoles that can act as ion transporters and in some cases as mass analyzers. The ions can then be stored in an octapole sometimes called a Penning trap. From the trap then can then be transported to the ion cyclotron for analysis. The ion cyclotron is very interesting because it can be used to accumulate spectra and therefore increase the signal to noise ration and sensitivity. When the ions enter the cyclotron they begin to move in a circular motion. The position within the cyclotron tube is determined the two RF electrodes that are used to accelerate the ions in the cyclic path. The RF is then turned off and a second set of electrodes are used to sense the RF field created by the cycling ions. Since the ions have different masses and charges they will cycle at different speeds. Each ion will then give a different sign wave depending on its rate of circling. These different sign waves are all detected at once resulting in a interference type pattern which can be analyzed by Fourier transformation of the signals to give the frequency and amplitude of each ions. The frequency is related to the mass and charge by the equation

$$freq = \frac{zB}{2\pi m} \qquad \frac{m}{z} = \frac{B}{2\pi\ freq}$$

This stimulation and detection process can be performed many times a second resulting in the ability to signal average, along with the storage aspect, increases the sensitivity of the method. The sensitivity can be as low as one atomole (10^{-18} moles) The mass resolution in this case is very good and is at least 1ppm.

Use of Mass spectrometers

So now that we have these instruments what use are they? The different instruments allow use to determine several things. If we have a pure protein we can determine the total mass of the protein if we use the multiple charging of the ESI. The same can be done with looking at multiple subunit complexes. As long as we can keep the complexes together when going into the ESI system they can be detected. Using a combination of the ESI quadrapole MS followed by a second quadrapole that we allow inert gas ions into to collide with the selected ions we can fragment the ions to get daughter ions. We can then choose those ions that fragment at the peptide bond and use the changes in the masses to sequence the peptide. Using MALDI-TOF we can look at the peptide map of a protein. Interestingly we can also use MALDI-TOF to look at what proteins are present in different regions of a cell. Lastly using additional labeling techniques we can quantify the proteins present in the sample.

Protein sequencing

The process of getting a protein sequence is fairly straight forward. The spot of protein of interest is first excised from the gel and digested with trypsin. Trypsin cuts proteins at positive charges and so at any lysine or arginine. If you use trypsin to cleave the proteins you will be using positive ESI since you will be assured of at least on positive charge. The best results are usually obtained with doubly charge fragments but singly charged ion will work. The protein digest is then injected into a nano-HPLC with a C-18 reverse phase column. The typical volume injected will be only a few microliters. The total column volume will be only 100-200 microliters. The peptide fragments are separated by a gradient and elute into the cone for the ESI-MS/MS.

Several problems exist with sequencing proteins. Although trypsin cleaves most proteins into segments of 10 -15 amino acids, in some cases the fragments are larger than 20 amino acids. At this point, if it is singly charged, the mass to charge ratio will be outside the range of most quadrapole systems. If the fragment is too small, 1-2 amino acids, it may also not be seen by the spectrometer. Another problem is that some peptides do not ionize well and so are at low concentrations or not seen at all. Lastly, post translational modifications will alter the mass of the fragment, may reduce the charge on the molecule such that it will not have the proper charge to see, the modification may come off during the fragmentation, and when the amino acid with the modification comes off it will have a different mass than expected and may need to be analyzed by hand.

Let’s go through the process. The peptide goes into the ESI source and is ionized obtaining a charge. The ion of interest is then selected and sent into the collider. The collider is a quadrapole MS where the ion goes through but is exposed to a small amount of argon gas. The collisions between argon and the peptide result in the breakage of the molecule at several locations. Some of these will be at the peptide bond. When the cleavage is at the peptide bond we obtain Y an B ions. The Y1 ion will be the mass of a single amino acid. The Y2 ion will be two amino acids and will differ from the Y1 ion

by the mass of a single amino acid and so on. The masses of the various ions are measured by the TOF mass spectrometer. This is used since it gives a very good mass resolution with fairly good sensitivity. The use of the HPLC also concentrates the sample to increase sensitivity.

What you see is a bunch of line representing the various masses. I should point out that the peak height does not correlate with the concentration of the ions. It is a reflection of the breakage efficiency and the charge on the ion.

There are several difficulties in the sequencing process. In this slide is shown the masses of the various amino acids listed according to their mass order. The masses listed here can be determined to near this resolution with a good instrument. Notice that there is only one atomic mass unit difference between K and E and N, L, and D. There is also only a two atomic mass unit difference between P and V, T and C, and E and M. Therefore we will need the resolution of the TOF to see these differences. Another problem is that I and L have the same mass and Q and K differ by only 0.04 mass units. In most instruments we are not going to be able to distinguish these amino acids.

On the next slide is another problem. Sometimes a dipeptide has the same mass as a single aminon acid. Listed in this table are some of the examples. As you can see a diglycine has the same mass as N. There are several more examples listed here. Therefore in some case you may think that you had a single amino acid come off when in fact there were two amino acids. When you identify any of the amino acids that could come from the combination it is important to be sure that there are no ions that could represent the amino acids that could make this dipeptide.

So how do we deal with this ambiguity? The easiest way is to have the genome present and use it to help fill in the gaps. A program can search a genome predicting all of the possible trypsin fragments. From these fragments the sequence is used to suggest the ion pattern that would be observed. This pattern is compared to the experimental pattern and if a match is observed then you have identified the protein and then fragment from that protein. Sometimes the fragment could have multiple proteins with that sequence and then all possible proteins are predicted. If sufficient protein is present it could also be Edman sequenced. In this system there are also several drawbacks to obtaining an accurate sequence.

There are several programs that are sued to identify proteins from LC-MS/MS. The most common are Sequest and Mascot. Each of these has a different way of coming to the assignment. These two programs are actually combined in a program call Scaffold and both analysis are run. When they agree it usually means that the assignment is correct.

For these programs to work you need to know the sequence before you try to sequence it with MS/MS. This means that you will need a genome that is annotated, a purified and sequenced protein, the sequence in an EST library, or that it has high identity with another known protein.

Quantification by MS

There are two main methods for quantifying the concentration of the proteins in samples. The First is call stable isotope labeling of amino acids in cell culture or SILAC. The next two methods are also very similar. The first is ICAT, isotope coded affinity tag, and the other is iTRAQ, isobaric tag for relative and absolute quantitiation. We will go through each of these to explain how they work. The only other competing technology is the differential gel electrophoresis system.

In SILAC the cells are grown in culture. To one culture a heavy isotope is added, N15 or C13, and the cells allowed to uptake the isotope. After a period of time with the treated, both cells are lysed, the proteins extracted and they are mixed. Care must be taken to mix the same amounts of protein from each culture. The proteins are separated by 1D gels, bands sliced, digested with trypsin, and analyzed by LC-MS/MS. In this case the relative size of the peaks is an indication of the relative concentrations of each species. Therefore the relative concentrations between control and treated can be determined.

The next method is similar to SILAC since the differences in the relative peak intensity is used to quantify the proteins. In ICAT (isotope coded affinity tag) The proteins are modified after their isolation. The protein label has 4 parts. The first is the iodoacetamide group that will link the tag to the sulfur of cysteine. On the other end is a biotin which is used to pull the tagged proteins out of the mixture by binding it to an avidin linked to a magnetic bead. Next is an isotopically labeled linker. This is a chain of 10 carbons that can have up to 9 of them changed to C13 resulting in a mass difference. The last aspect is an acid labile cleavage site to remove the biotin the linker. The proteins extracted from the samples are labeled with either the heavy or the light isotope chain and then mixed. The labeled proteins are then digested with a protease and the labeled peptides are isolated by the avidin beads. The labeled peptides are them identified and sequenced by LC-MS/MS.

iTRAQ is similar to ICAT in that we are labeling the proteins but this time we are labeling the primary amines found in lysine and the N terminus of peptides. Here the proteins are reduced, iodoacetamide blocked, and then digested with trypsin. The lysines and N termini are labeled with the iTRAQ label. In ITRAQ the total labels are all of the same mass so that they all appear at the same mass in the MS. The tag has the reactive group, a reporter group of different masses and a linker that will balance the mass such that all the tags have the same mass. The proteins are then separated by reverse phase nano-HPLC and run through the MS/MS. The peptide is identified by its sequence and the relative concentration of the different samples is determined from the reporter groups and their relative peak heights.