MS. MASS SPECTROMETRY OF PROTEINS

Protein sequencing is the first step towards a structural understanding of protein function. There are multiple techniques that may be applied to the problem. Edman degradation is a chemical technique that cleaves amino acids sequentially from the N-terminus of a protein. While somewhat old-school, it is still in use. You should Google it to get the gist of what is going on. A more common technique is to obtain the DNA sequence of the gene encoding a protein (see my notes on the Central Dogma) and then *translate* the DNA sequence to obtain the predicted protein sequence. We'll cover that at the end of the semester. Note however that it requires knowledge of the gene associated with a given protein. That information is not always present.

When you have an unknown protein and wish to identify its sequence, the most common technique in current use is mass spectrometry MS. As in organic chemistry, MS measures the charge to mass ratio of an ionic species. Also, as in organic chemistry, fragmentation products are key to making a positive ID of the protein's identity. Below we will discuss sample preparation, ionization, charge/mass measurement and sequence reconstruction.

Protein and Peptide Fractionation

Not surprisingly, one needs a pure protein sample in order to obtain meaningful mass information. There are an enormous number of techniques used in biochemistry to purify proteins, but we'll just look at a couple of them here.

Electrophoresis

The most commonly used separation technique in protein biochemistry is gel electrophoresis. The gel is a cross-linked polymer of acrylamide that functions as a sieve, allowing small molecules to pass easily while retarding the progress of larger molecules. The motive force dragging proteins through the mesh is an electrical potential. Typically proteins are solubilized with sodium dodecyl sulfate (Figure MS.1), a strong anionic detergent, which gives each protein molecule an overall anionic charge roughly proportional to its size (molecular weight). By placing the cathode at the top of the gel and the anode at the bottom, the anionic protein samples migrate to the bottom at a rate dependent on their molecular size.

Reverse Phase Liquid Chromatography

Ionization Methods

The principal challenge of protein mass spectrometry is getting proteins from the condensed phase to the gas phase in an ionized form. I've heard it described as "getting elephants to fly". Traditional mass spectrometry is limited to thermostable compounds that can be heated to a point at which they readily vaporize. That doesn't work well for proteins, which are cooked before

they enter the gas phase. Instead, two common techniques have been applied to this problem as described below.

Matrix Assisted Laser Desorption/Ionization (MALDI)

MALDI is commonly used to generate singly charged ions of intact proteins. The protein of interest is mixed with a solution of some UV-absorbing compound (commonly a derivative of cinnamic or benzoic acid). The solvent is evaporated off, leaving a sample of protein embedded in an acidic organic matrix (the "M" in MALDI). A laser, tuned to the absorption wavelength of the matrix, is fired at the sample causing rapid heating and evaporation of the matrix, essentially abandoning the protein in a naked, exposed (and gaseous) state. For reasons that aren't fully understood, the protein is typically released in a singly ionized state $(M+H)^+$ due to the transfer of one additional proton to the neutral protein molecule. That ion is ready for mass analysis.

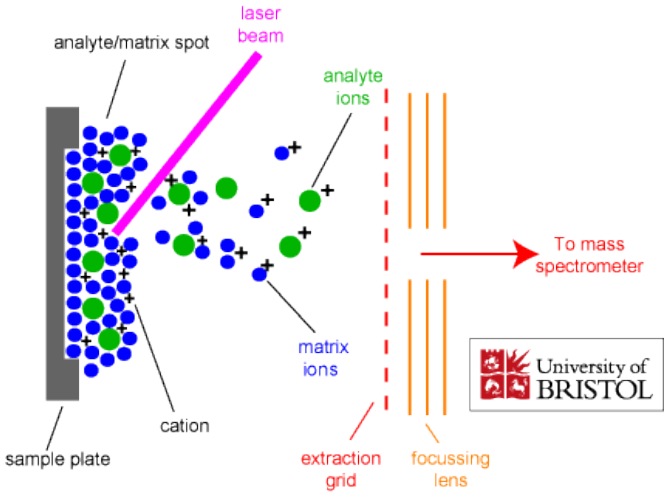

Figure MS.2 Stolen diagram of MALDI. Note solid mixture of analyte (protein) and matrix (organic compound) on the left. The cations are sadly drawn as though there are no anions around. That's what you get for stealing images. However, the protein molecules are typically released into the gas phase as cations and drawn to the negatively charged plate that acts as the extraction grid.

Electrospray ionization (ESI)

ESI operates by taking a protein from organic solution (instead of organic solid matrix like in MALDI) to the gas phase via evaporation. One additional difference between ESI and MALDI (beyond phase change) is the use of a high potential in ESI to impart charge to the evaporating liquid. ESI proceeds by spraying an acidic solution (acetic acid in acetonitrile perhaps) of the protein through a capillary tube (small internal diameter) that is held at a positive potential (2-4 kV). As a result of that potential, a positive charge is imparted to the drops thanks to an excess of cations – usually protons – and the solution disperses into a fine mist. As the solution evaporates from the droplets, the protein in the droplets absorbs the excess protons, and when evaporation is complete, the protein is left naked as a $(M + nH)^{n+}$ cation. The value of "n" is

variable, and so a collection of mass ions varying by charge will be directed to a negatively charged focusing plate (Figure MS.3).

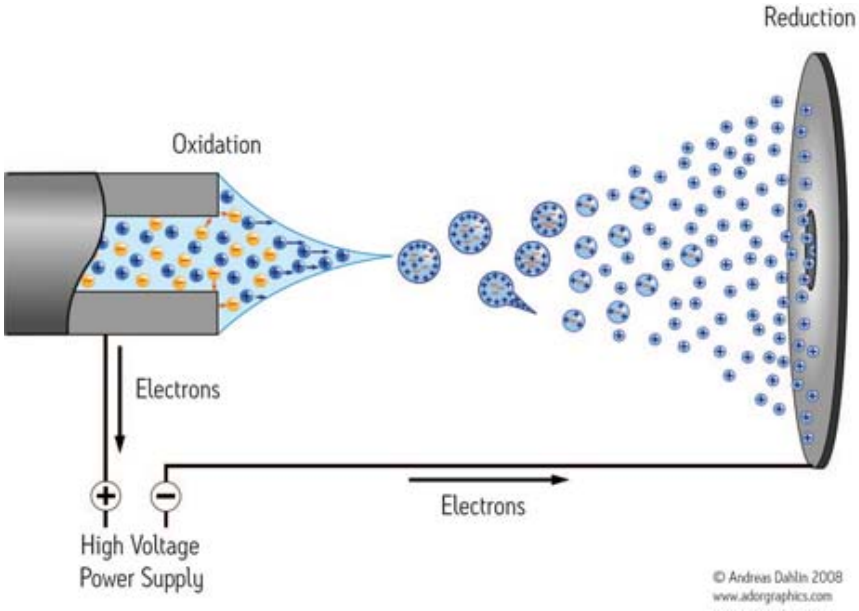

Figure MS.3. Diagram of electrospray ionization. The sample is emitted through a positively charged capillary. A positively charged spray is emitted and evaporates to yield cationic forms of the protein of interest. They are focused through a cathode (the extractor cone) and proceed to the mass analyzer. (Figure stolen from http://www.waters.com/waters).

Mass Determination

Once an ionized form of the protein has been introduced into the instrument via MALDI or ESI (or some other method), separation of ions by mass/charge ratio must take place to give the mass spectrum. Again a couple of techniques dominate.

Time of Flight (TOF)

TOF is the simplest form of mass analysis and was employed in the earliest mass spectrometers in the 1920's. The protein cation is accelerated by an electric potential that is applied across a "drift tube" through which the ion must "fly". Since the potential is fixed, the acceleration is related to the mass/charge ratio of the ion. Ions with a smaller m/z ratio accelerate more rapidly and their "time of flight" is shorter than for ions with larger m/z ratios. There is a nice derivation of the equation (Eq. MS.1) that relates the m/z ratio to time (t), length of tube (L), and voltage (U) in Wikipedia.[1]

$$(m/z) = 4t^2U/L^2 \qquad \text{(Eq. MS.1)}$$

[1] http://en.wikipedia.org/wiki/Time-of-flight_mass_spectrometry accessed on 8/25/12.

TOF spectrometers are often, but not uniquely, coupled to MALDI ion sources, which creates a MALDI-TOF instrument. Ions are produced in pulses by laser desorption from the matrix, and following each desorption event, the time delay to ion detection (in the micro- to millisecond range) is measured. A large number of pulses are collected and the results pooled to generate a mass spectrum for the cation of interest.

Quadrupole Mass Analyzers (Q)

While TOF measures the time required of all ions to transit the drift tube, Q analyzers behave more like filters, only allowing ions of a selected m/z ratio to be transmitted. The quadrupole is composed of four metal tubes arranged in parallel (Figure MS.4) such that they are maintained in pairs, with an oscillating potential field created between each pair. Ions adopt an oscillating path as a result of the field, but can only persist through the quadrupole if their acceleration in the field can be reversed with each oscillation of the potential. By tuning the frequency and potential of two simultaneous dipoles, only ions of a particular m/z ratio will survive. Others will spin out of control and strike the rods. The advantage of quadrupole detection is the high resultion of m/z ratio compared to TOF.

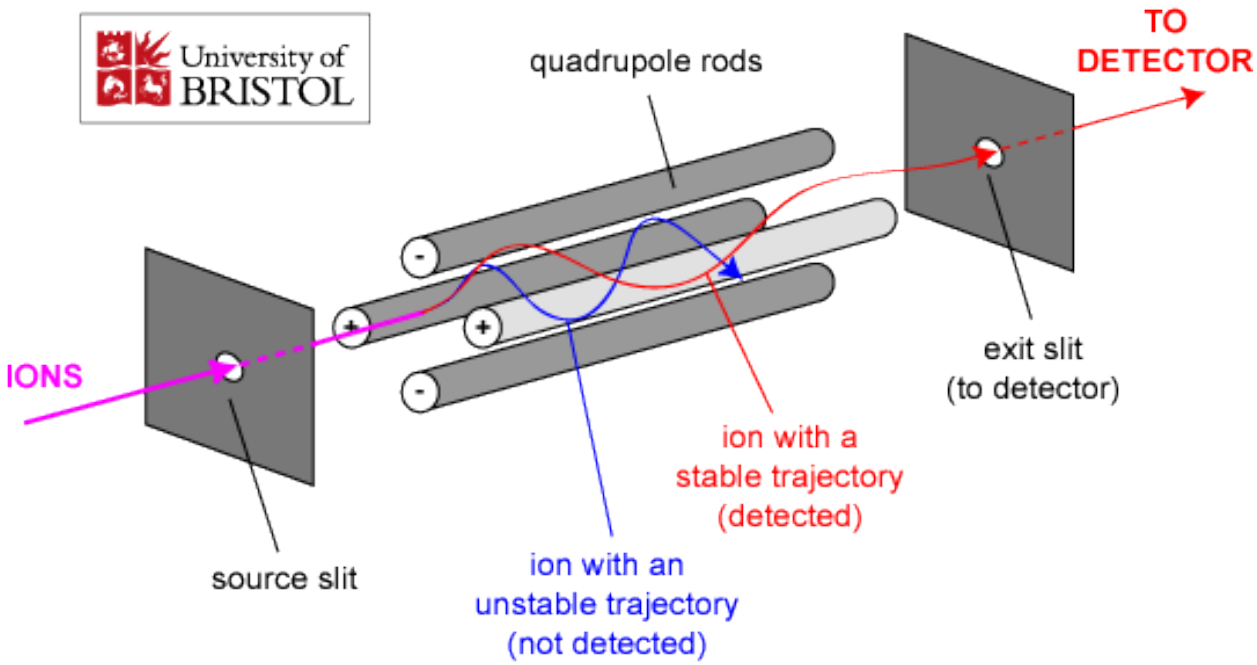

Figure MS.4. Quadrupole mass analyzer. Two sets of parallel rods set up simultaneously oscillating electric fields. Only ions with a selected m/z ratio will successfully navigate the analyzer. (Image purloined from http://www.chm.bris.ac.uk/ms/theory/quad-massspec.html)

MS/MS Sequencing

Mass spec is useful in a wide variety of applications, but has become a relatively common tool for determining the primary structures of unknown proteins. The sample protein is typically

isolated by electrophoresis and subjected to **tryptic digestion**. The latter process uses a **protease** called **trypsin** to hydrolyzed the protein of interest. For reasons to be discussed later this semester, trypsin hydrolyzes peptide bonds to the C-terminal side of lysine and arginine residues, creating a collection of smaller peptides from the initial intact protein (Figure MS.5).

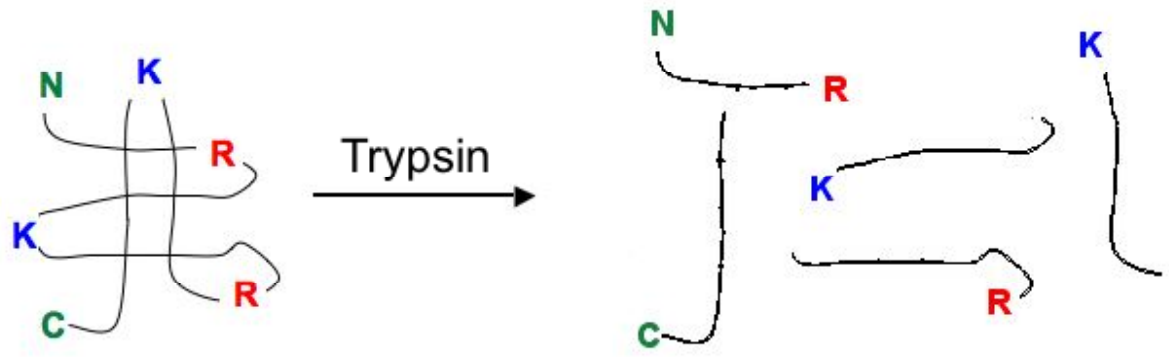

Figure MS.5. A protein with several arginine (R) and lysine (K) residues will be hydrolyzed by trypsin to yield peptides where the C-terminus of each peptide (except the C-terminal one!) will have Arg or Lys as the final residue.

The peptide fragments are commonly separated by reverse phase chromatography[2] and injected into a mass spectrometer (ESI is convenient in this instance because it ionizes samples already in the solution phase). The trick here is that the first mass spec is arranged in tandem with a second mass spec (aka MS/MS). The first device is used as a mass filter that delivers a fragment of desired m/z ratio to the second device. In this way, one can select for the $(M+2H)^{2+}$ ion of a given peptide fragment (Figure MS.6).

What happens next is that the selected peptide ion is subjected to fragmentation prior to entry into the second analyzer. There are a number of methods available, but a common choice is collision induced dissociation (CID), which bombards the ion with neutral atoms (such as a noble gas), which is like firing cannon balls at it. The kinetic energy imparted is sufficient to break bonds along the backbone of the peptide. The most common fragmentation pattern involves cleavage at the amide bonds, yielding so called "b" and "y" fragments (Figure MS.6). Essentiall, the "b" fragments are N-terminal chunks of the peptide with the b_1 fragment having one residue, b_2 has two residues, etc. In contract, the y fragments are from the C-terminus and are likewise labeled for the number of included residues.

2 Reverse phase chromatography is similar to silica gel chromatography, but reversed! The solid phase is a silica support coated with a hydrocarbon surface. C4 columns are coated with butyl groups, C18 with octadecyl groups. A polar solvent (acetonitrile/water mixtures are common) acts as the mobile phase. Non-polar molecules elute more slowly than polar molecules because they have higher affinity for the solid phase.

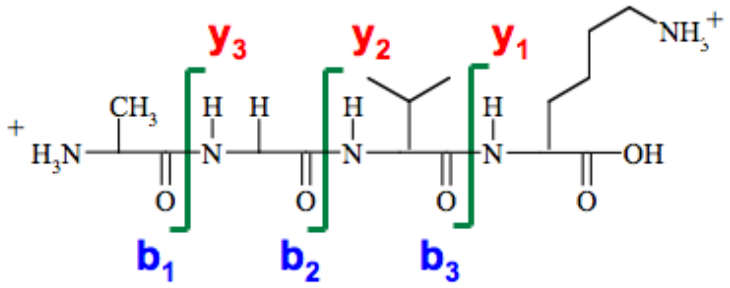

Figure MS.6. Fragmentation options for a dicationic peptide fragment. Note that tryptic digests always leave a basic residue at the C-terminus (Lys or Arg) and a free amino group at the N-terminus. The protons can migrate, but this way the $(M+2H)^{2+}$ ion looks sensible.

The mechanism of fragmentation requires that a proton migrate from the N-terminal ammonium group to the various carbonyls, which weakens the relevant amide bond and permits fragmentation at that site (Figure MS.7).

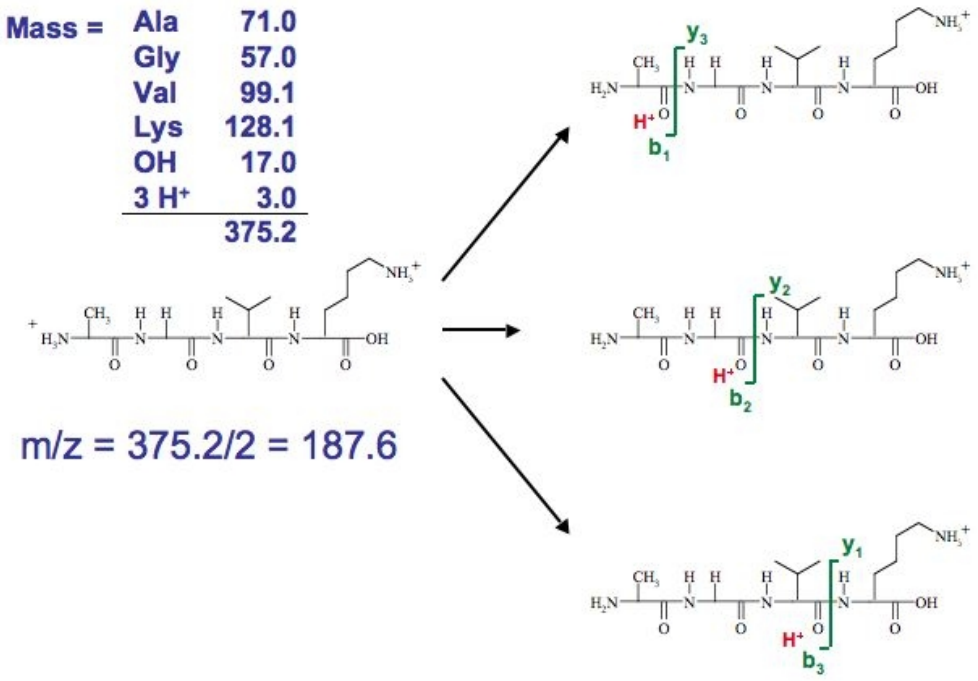

Figure MS.7. Note that the mass ion in this case has an m/z ratio of 187.6. The tabulation is performed by adding up the residue weights plus the –OH that is appended to the C-terminus, two H^+ that are appended to the N-terminus and one H^+ that is appended to the lysine. As the proton migrates, fragmentation is promoted at alternate sites.

Fragmentation then proceeds via a variety of mechanism, one of which is shown in Figure MS.8. The important thing to note is that one proton will always remain with the C-terminal Lys or Arg residue of the peptide (part of the y fragment) while the other will be retained by a nitrogen in the b fragment, so each will carry a single positive charge. Also, the masses of the fragments are adjusted by the presence of several atom groups above and beyond the residue masses. Remember that residue masses are for that fraction of the amino acid found in a protein. The backbone nitrogen normally contains a single bonded hydrogen and the C-terminal carbonyl is not a carboxylate. Thus we need to adjust our expectations of masses. The b fragment will always have one extra proton beyond its collection of residue masses, while the y fragment will have two protons and an OH group beyond the sum of its residue masses (Figure MS.8).

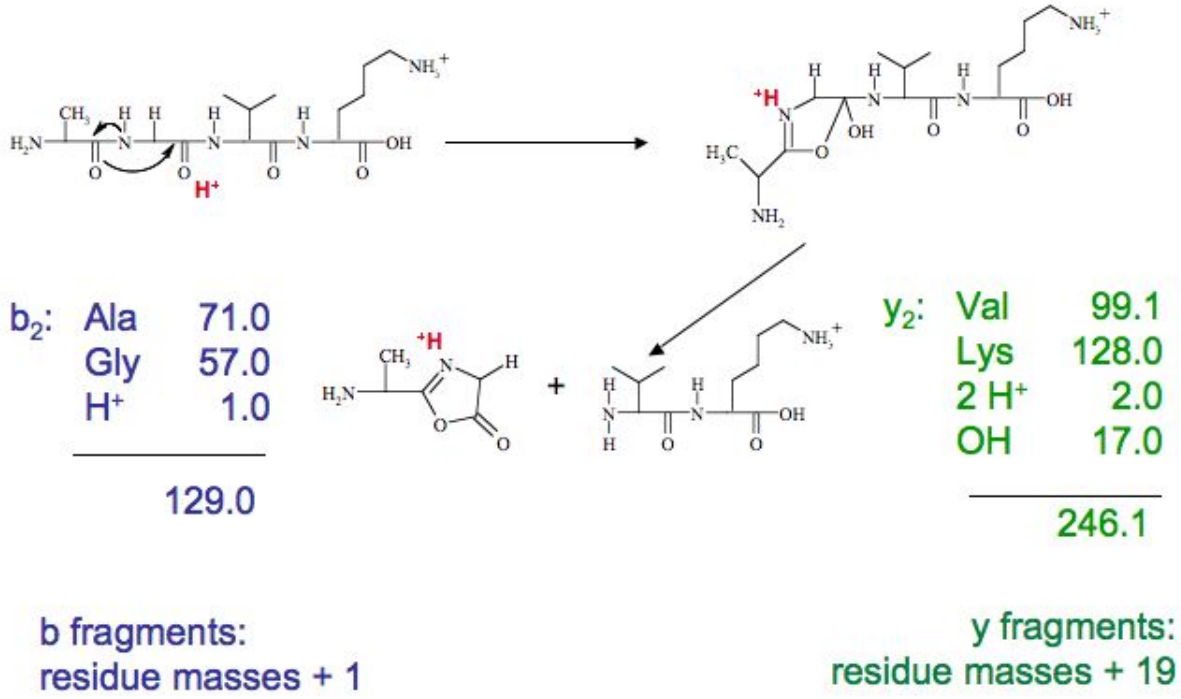

Figure MS.8. Note that amide bond with proton is cleaved in collisionally induced dissociation (CID). The b fragment will end up with one excess proton, and the y fragment with two excess protons and an OH group.