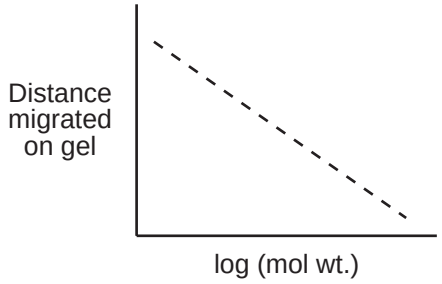
Methods for Protein Analysis

1. Protein Separation Methods

The following is a quick review of some common methods used for protein separation:

SDS-PAGE (SDS-polyacrylamide gel electrophoresis) separates proteins mainly on the basis of molecular weight as opposed to charge (which is 'swamped out' by the excess of protein-bound SDS) or folding (proteins are largely denatured in SDS). In the ideal picture, the distance migrated by the protein in a given time is inversely related to the logarithm of its molecular weight:



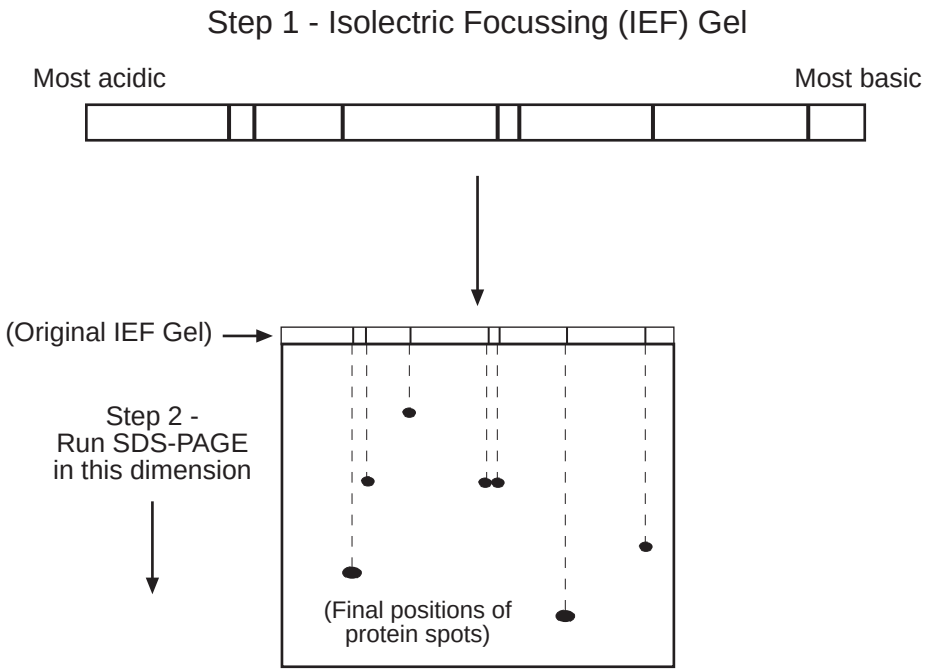
However, factors other than molecular weight alone (e.g., glycosylation or a strongly hydrophobic character) can also affect the migration of a protein on an SDS-PAGE gel, as discussed in class.

Isoelectric focussing separates proteins on the basis of their balance of acidic (negatively charged) vs. basic (positively charged) amino acid residues, which determines a property known as the protein's **isoelectric point** [pI], the pH at which the protein's net charge becomes exactly zero. Proteins that differ by as little as one charged residue (e.g., the mono- and di-phosphorylated forms of a given protein) can be separated using this method. Conversely, however, proteins with very similar pI values but very different sizes can run at the same position.

Two chromatographic methods are frequently used for protein or (more often) peptide separation. **High-performance liquid chromatography (HPLC)** can be used to separate and to purify proteins/peptides based on size, charge or overall hydrophobicity. **Thin-layer chromatography (TLC)** can also be used to separate out peptides (e.g., derived from proteolytic digestion of a protein) based on similar properties. Both of these methods are very useful adjuncts to gel-based approaches, particularly for peptide analysis and for preparative purposes (e.g., to purify rather than simply to analyze proteins).

Two-dimensional (2-D) gel electrophoresis is a powerful gel-based method commonly used for 'global' analysis of complex samples (i.e., when we are interested to characterize the full range of proteins in a sample, not just a few specific proteins). In this technique the protein is run first in a narrow (often tube-shaped) isoelectric focussing gel, which as already noted separates proteins on the basis of their isoelectric points (acid vs. basic character). The isoelectric

focussing gel is then placed over an SDS-PAGE gel and run on the latter in the perpendicular dimension:



In the final gel, proteins run not as bands but as spots, and the position of each protein spot depends on both its size and its charge properties. This provides a more powerful means to separate proteins than does either SDS-PAGE or isoelectric focussing alone.

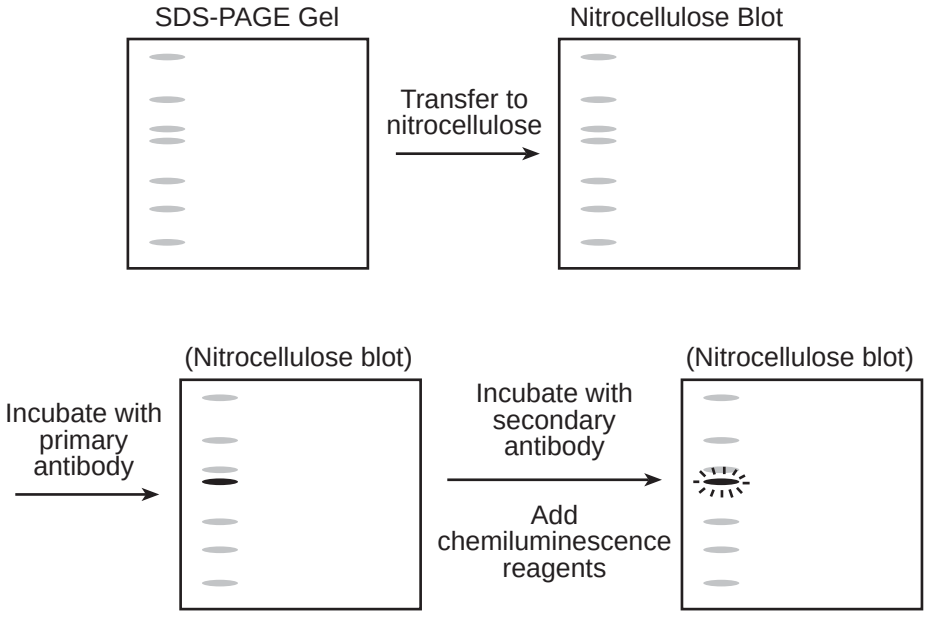
2. Western Blotting

The most common version of Western blotting is known as **immunoblotting**. In this technique a sample of proteins is first electrophoresed by SDS-PAGE to separate the proteins on the basis of their molecular weights. The wet gel is then placed against a sheet of **nitrocellulose** and placed in a special type of electrophoretic chamber. The gel (sandwiched against the nitrocellulose) is then subjected to an electric field which causes the proteins to migrate out of the gel and onto the nitrocellulose sheet, to which the proteins become tightly (in effect irreversibly) adsorbed. The nitrocellulose with its tightly bound proteins can then be 'blotted' or 'probed' with an antibody (known as a **primary antibody**) which (in theory) will bind to the nitrocellulose sheet only in places where the protein(s) recognized by the antibody are adsorbed. We can then determine where on the nitrocellulose blot the primary antibody binds (a clue to the identity of the protein) and how intense the binding is (a clue to the amount of the protein present).

The above conceptual description neglects a few technically important steps and some key factors that determine how well the method works in practice. Since insufficient control of these factors can create both false negative and false positive results, we will briefly discuss them here. To help make things concrete,

let us suppose that we are studying a specific kinase (MyKinase) that is recognized by a particular mouse monoclonal antibody (anti-(MyKinase), of the immunoglobulin G1 [IgG1] class of mouse antibodies).

The first complexity that we should note in immunoblotting concerns how we detect binding of the primary antibody (anti-(MyKinase)) to the blot. In principle the primary antibody could be radioactively or otherwise labeled in a manner that would allow us to determine directly (e.g., with a simple autoradiogram) where on the blot it had bound. A more sensitive and efficient, and therefore much more common, method uses a two-step approach with two different antibodies. In this approach the nitrocellulose blot is first incubated with anti-(MyKinase), then washed and further incubated with a **secondary antibody** which binds to the primary antibody (say, a goat antibody which binds to any mouse IgG1 antibody) and which is modified in some manner to make it easy to detect. The most common secondary antibodies in current use are chemically modified so that they catalyze **chemiluminescence** (production of small amounts of light) when the appropriate reagents are added. When a nitrocellulose blot treated with anti(MyKinase) and secondary antibody is soaked in a solution of the chemiluminescence reagents, then placed against a piece of film and left in the dark, the film will become exposed (due to local light production) *only* where the secondary antibody is bound to the blot, and hence (in theory) only where (MyKinase) is present on the nitrocellulose blot. The key steps in immunoblotting at a conceptual level can thus be summarized as shown schematically below:



The second complexity in immunoblotting is one whose origin may already have occurred to you. Nitrocellulose is a very 'sticky' matrix, which is why it can efficiently adsorb all of the proteins transferred to it from the original gel. However, after the initial protein sample is adsorbed to the nitrocellulose, we must have some way to reduce the 'stickiness' of the nitrocellulose before we

incubate it with the primary antibody, anti-(MyKinase, so that the antibody will bind to the nitrocellulose sheet only where it specifically recognizes (MyKinase). This is accomplished by **'blocking'** the nitrocellulose after it adsorbs the proteins from the gel (and without displacing these proteins from the nitrocellulose) by incubating it in a solution rich in proteins that our primary (or secondary) antibodies does *not* recognize. The most commonly used blocking solution for most purposes is reconstituted nonfat dry milk. Once the nitrocellulose is properly 'blocked,' meaning that all the remaining 'sticky sites' have been tied up by proteins from the blocking solution, the primary and secondary antibodies will bind tightly to the blot *only* where they can bind to proteins already on the blot, not to the nitrocellulose itself.

Pay close attention to my last statement: "[After blocking] the primary and secondary antibodies will bind tightly to the blot *only* where they can bind to proteins already on the blot, not to the nitrocellulose itself." This does *not* mean however that the antibodies may not bind weakly to the blot elsewhere through relatively nonspecific interactions. Therefore, a further important technical element in immunoblotting is to **wash the blots** thoroughly after each incubation with antibody. This (if properly done) will minimize the level of nonspecific binding of antibodies to the blot and hence reduce both the 'background staining' and the possible identification of 'false-positive' spots. **Careful blocking, and later washing to remove primary and secondary antibodies that are not tightly bound to the nitrocellulose blot, are crucial for the success of an immunoblotting experiment.**

Let's now outline the sequence of steps in a real immunoblotting experiment:

- Run SDS-PAGE
- Transfer to nitrocellulose
- **BLOCK**
- Incubate with primary antibody
- **WASH**
- Incubate with secondary antibody
- **WASH**
- Soak blot with chemiluminescence reagents
- Expose film to reagent-soaked blot
- Develop film
- Observe darkened spots/bands on film (align with markers on original gel)

The quality of the experimental technique, and also on some cases the quality of the antibodies, can have a huge effect on the quality of the final results.

Two additional 'complexities' in immunoblotting experiments are probably worth noting. First, a given primary (or more rarely a secondary) antibody may not be very pure or very specific, and hence it may bind to proteins other than the 'target' of interest to us. In this case we will see extra bands on an immunoblot that have nothing to do with our protein of interest. (In fact, immunoblots frequently have extra bands around the region where antibody

heavy chains run, even if they are otherwise 'clean'). For this reason among others, authors will usually show in their papers only the small region of an immunoblot where the protein of interest is found. Second, if our primary antibody is a monoclonal antibody, in some cases it may fail to bind to the 'target' protein if the antibody-binding region of the latter (the **epitope**) has been cleaved off or is posttranslationally modified in some manner that significantly changes its reactivity with the antibody. This can be a useful effect (for example, we can determine whether a protein has lost a small region at its N-terminus by probing replicate blots with two primary antibodies, one of which recognizes the N-terminus while another recognizes another region of the protein).

A final useful feature of immunoblots is that on a given blot (or on a given set of blots that are run, transferred, blocked, etc. at the same time with exactly the same solutions), the integrated intensity of a given protein band is proportional to the amount of the protein present. This can be very useful, for example to compare the levels of a given protein present in different cell samples. It is however difficult to compare intensities of bands between blots run in different experiments (unless one uses purified protein as an internal standard In each experiment, which few researchers do) - too many factors can differ between them.

A variation of the above method is **Far-Western blotting**, in which the nitrocellulose blot is 'probed,' not with an antibody but rather with a labeled soluble protein of interest (we'll call it **P**). The probe protein **P** is radiolabeled or modified in some other manner that allows it to be easily and sensitively detected where it has bound to proteins on the blot. If for example the probe protein **P** iis radiolabeled calmodulin, we can detect which proteins in our sample bind to calmodulin. One point to keep in mind about Far-Western blotting, however, is that proteins may often bind their normal protein 'partners' only when the proteins are in an undenatured conformation. Some proteins are relatively inefficient at refolding from their SDS-denatured state to a native conformation when transferred from an SDS-PAGE gel to nitrocellulose. It is thus quite possible to have 'false negative' results in which we find that a protein of interest to us does *not* bind the probe protein **P** on a Far-Western blot, even though the two proteins actually *do* interact under physiological conditions.

3. Protein Identification Methods

Let us suppose that you have found a protein of interest as a band/spot on a gel or as a peak from an HPLC separation, and you wish to determine its molecular identity. Two methods, **Edman degradation** and **mass spectrometry**, are commonly used for this purpose.

3A. Edman degradation uses reagents known as **isothiocyanates** to react with the N-terminus of a protein or peptide. Under suitable conditions, a single 'round' of Edman degradation will cleave off the N-terminal amino acid residue to produce a derivative of that amino acid plus a new free amino terminus corresponding to the next amino acid in the polypeptide chain:

NH_2-AA1-AA2-AA3-AA4-… + (isothiocyanate) → **X**-AA1 + NH_2-AA2-AA3-AA4-

where **X** is a chemical 'tag' that the reagent attaches to the N-terminal amino group as part of the cleavage process. We can then separate **X**-AA1 from the mixture and identify what amino acid it represents. The reaction cycle shown above can then be repeated to determine the second amino acid in the chain, AA2:

NH_2-AA2 -AA3-AA4-… + (isothiocyanate) → **X**-AA2 + NH_2-AA3-AA4-…..

and so on. The isothiocyanate reagent may be fluorescent or radioactive, allowing determination of very small amounts of released amino acid derivatives **X**-AA and hence allowing sequencing of very small amounts of protein (nanograms or less).

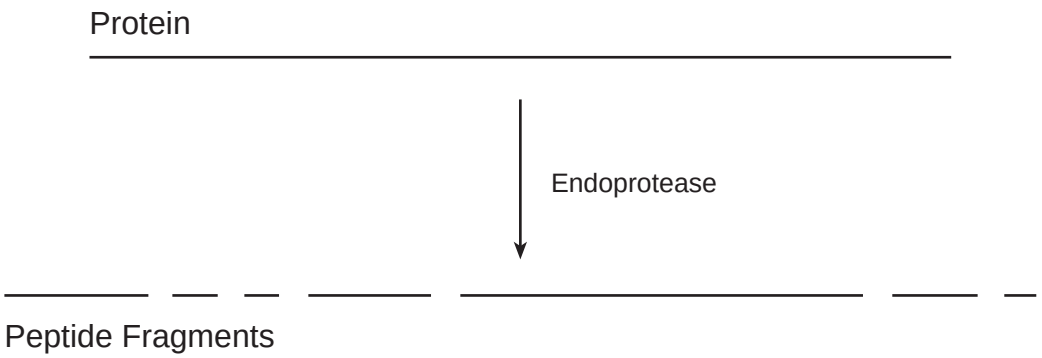
Once the amino-terminal sequence of a protein has been determined for at least 8-10 residues, the deduced sequence can be compared to a protein or genomic database for the organism of interest in order to determine the protein's identity. If more than one protein from the database fits the experimental sequence, additional experimental criteria such as the protein's molecular weight can be used to refine the assignment. In some cases it may be possible to cleave the protein (chemically or enzymatically) into peptides that can themselves be sequenced to facilitate identification of the protein.

Edman degradation has a few significant limitations; it does not work well for very hydrophobic proteins or peptides (a potential problem for membrane proteins), and it cannot work when the N-terminus of the protein or peptide is chemically blocked (e.g., with an acetyl group). In the latter case, however, it may be possible to cleave the protein (chemically or enzymatically) into peptides, most of which (except for the one derived from the original protein N-terminus) will have free amino-termini and can therefore be sequenced.

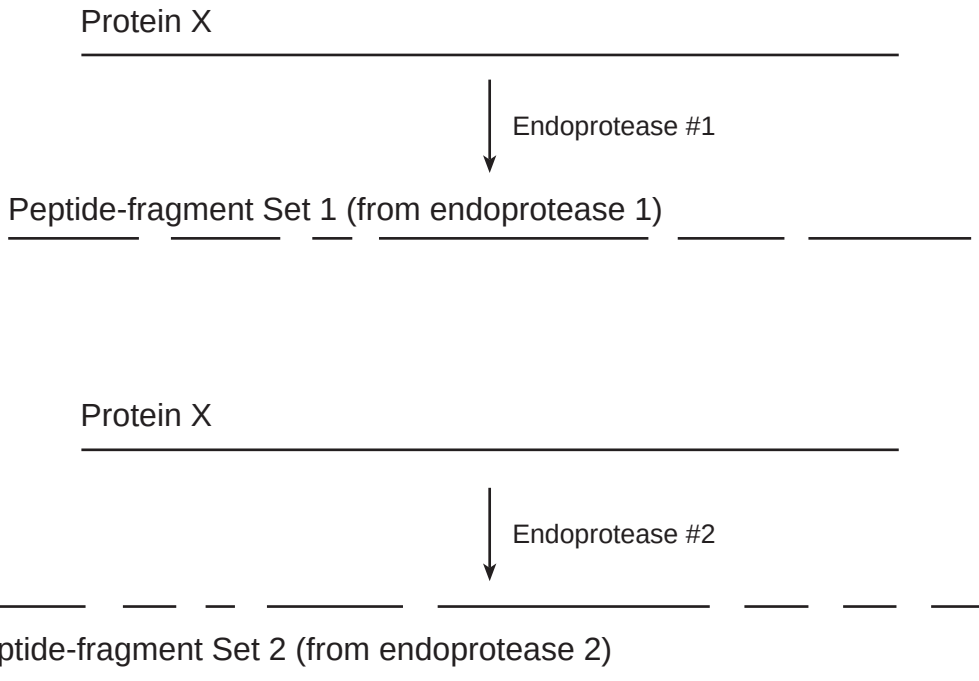
3B. Mass spectrometry is a very sensitive and accurate method to determine the precise molecular weight of a protein. Because it is very sensitive and is amenable to rapid processing of many samples, it is becoming very popular for determining the different proteins present in complex samples (e.g., the proteins in a given cellular organelle, separated by two-dimensional gel electrophoresis), in what is known as a **proteomics** analysis. Mass spectrometry can be used to determine protein sequences and thus to determine the identities of proteins in a manner conceptually similar to that described above based on Edman degradation. A more powerful method (because it can be applied to large numbers of samples more rapidly) is however to use mass spectrometry to determine the precise molecular weights of specific cleavage fragments derived from a given protein, as described below.

Endoproteases (enzymes that cleave peptide bonds in the middle of a polypeptide chain rather than 'chewing off' amino acid residues one at a time

from one end) typically have significant specificity. Trypsin, for example, specifically cleaves peptide bonds on the C-terminal side of lysine and arginine residues. When a given protein is cleaved with an endoprotease of this type, it will thus yield a series of well-defined fragments, the number and masses of which are directly determined by its amino acid sequence. Mass spectrometry is perfectly adapted to provide such information about the products obtained upon cleaving an isolated (unknown) protein with a particular endoprotease:



We can then feed our experimental results into a protein (or more often a genomic) database to determine what proteins in our organism of interest would give a pattern of fragments exactly matching those found in our digested sample. There are of course a huge number of proteins in any organism, and the pattern of cleavage fragments obtained by treating a given protein with a single endoprotease often does not allow the protein to be identified uniquely.



However, and perhaps remarkably, if as shown above we digest samples of our protein with as few as two or three different endoproteases with different

specificities, we can usually use the resulting digestion patterns (again, analyzed by mass spectrometry to provide highly accurate determination of the fragment molecular weights) to produce a unique identification of our unknown protein:

Again, mass spectrometry is uniquely well-suited for such analyses because it can yield very accurate determinations of molecular weights from very even very small amounts of fragments resulting from the digestion of a particular protein. Techniques have now been developed by which proteins separated in two-dimensional gels can be digested within the gels and then injected directly into a mass spectrometer for analysis of the resulting fragments.