Bioanalytical chemistry

4. Gel Electrophoresis

Required reading: **Sections 9.1, 9.2.3, 9.2.4, 9.5.1, 10.1 to 10.7, 11.1 to 11.5, and 15.5** of Mikkelsen and Cortón, *Bioanalytical Chemistry*

Some objectives for this section

⇒ You will know what DNA agarose gel electrophoresis is
⇒ You will know what the difference between normal and pulsed field electrophoresis of DNA
⇒ You will know what SDS-PAGE is.
⇒ You will understand how the basis for molecular basis for size-based separation of proteins by SDS-PAGE.
⇒ You will know what IEF is.
⇒ You will know how SDS-PAGE and IEF can be combined in 2-dimensional gel electroophoresis
⇒ You will know what a Western Blot is.
⇒ You will appreciate how these techniques can be used in the analysis of DNA and proteins.

Primary Source Material

- Chapter 4 and 6 of Biochemistry: Berg, Jeremy M.; Tymoczko, John L.; and Stryer, Lubert (NCBI bookshelf).
- Chapter 3 and 7 of Molecular Cell Biology 4th ed. (Ch. 9, 5th ed.): Lodish, Harvey; Berk, Arnold; Zipursky, S. Lawrence; Matsudaira, Paul; Baltimore, David; Darnell, James E. (NCBI bookshelf).
- Chapter 12 of Introduction to Genetic Analysis Anthony: J.F. Griffiths, Jeffrey H. Miller, David T. Suzuki, Richard C. Lewontin, William M. Gelbart (NCBI bookshelf).
- Some animations are from http://www.wiley-vch.de/books/info/3-527-30300-6/.
- http://www.piercenet.com/

Electrophoresis

The velocity of migration *(v)* of a molecule in an electric field depends on the electric field strength *(E),* the net charge on the protein *(z),* and the frictional coefficient *(f)*.

$$v = \frac{Ez}{f}$$

The frictional coefficient *f* depends on both the mass and shape of the migrating molecule and the viscosity (η) of the medium. For a sphere of radius *r,*

$$f = 6\pi\eta r$$

The speed of migration is therefore proportional to the charge:mass ratio.

$$v \propto \frac{z}{r}$$

- Electrophoresis is a technique for separating, or *resolving,* molecules in a mixture under the influence of an applied electric field. Dissolved molecules in an electric field move, or migrate, at a speed determined by their charge:mass ratio.
- For example, if two molecules have the same mass and shape, the one with the greater net charge will move faster toward an electrode.

Agarose gel electrophoresis is typically used for the separation of DNA molecules

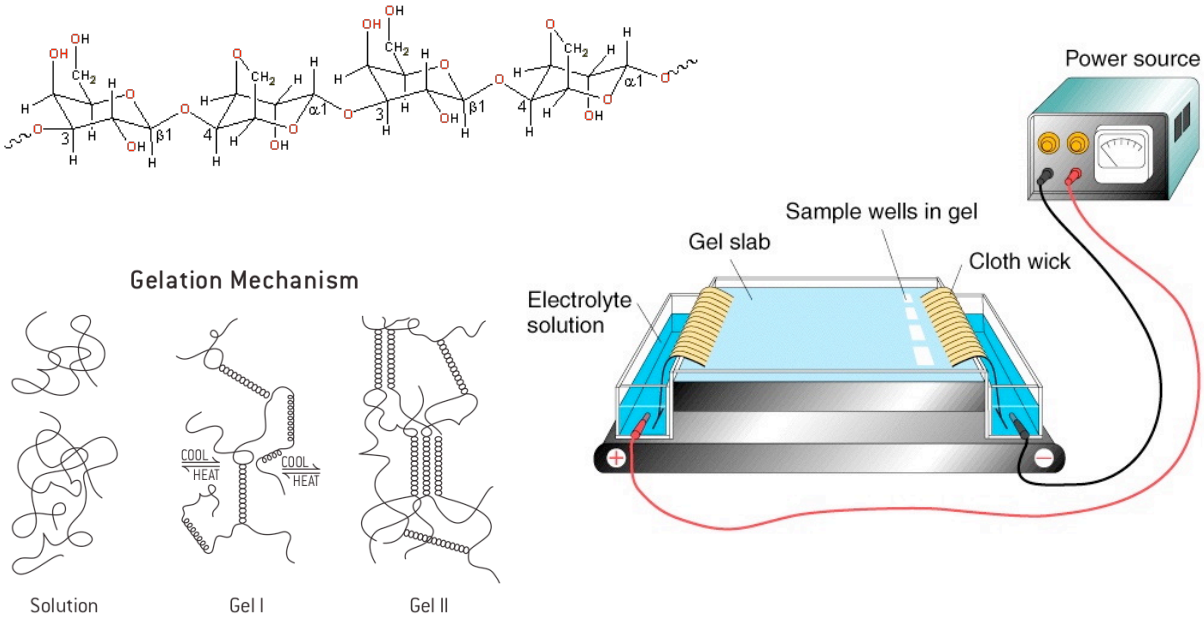


Figure 4: Gelation of agarose by formation of double helices connected in three dimensions by zones of random coil configuration.

- Agarose is a polysaccharide that is obtained from the cell wall of some seaweed species.
- The basic repeating unit of agarose is 1,3-linked β-D-galactopyranose and 1,4-linked 3,6-anhydro-α-L-galactopyranose. Two units of this repeating structure, known as agarobiose, are shown on the slide.
- The average molecular mass of the agarose polysaccharide is ~120,000 daltons, which corresponds to ~400 agarobiose units. Some of the sugar groups in the polysaccharide are sulfated.
- Agarose is essentially insoluble at room temperature but can be dissolved in water at high temperature. By heating and cooling a solution of agarose, a firm gel can be formed. The pores within this gel depend on the concentration of agarose used, but can be in the range of 100 to 300 nm. This pore size is sufficient to allow DNA molecules to pass through in an applied electric field. However, there is very little passive diffusion of DNA through the gel.
- The mechanism for gelation of agarose upon cooling is thought to involves the formation of double helix structures in the initial stages of gelation. These double helixes then pack together into bundles to give the final gel structure.
- Some information and the gelation mechanism figure from www.Lonza.com (MB_SB_appendixB_207-209.pdf)
- What is agar and agarose? See http://www.lsbu.ac.uk/water/hyagar.html

DNA is typically fragmented with restriction enzymes prior to electrophoresis

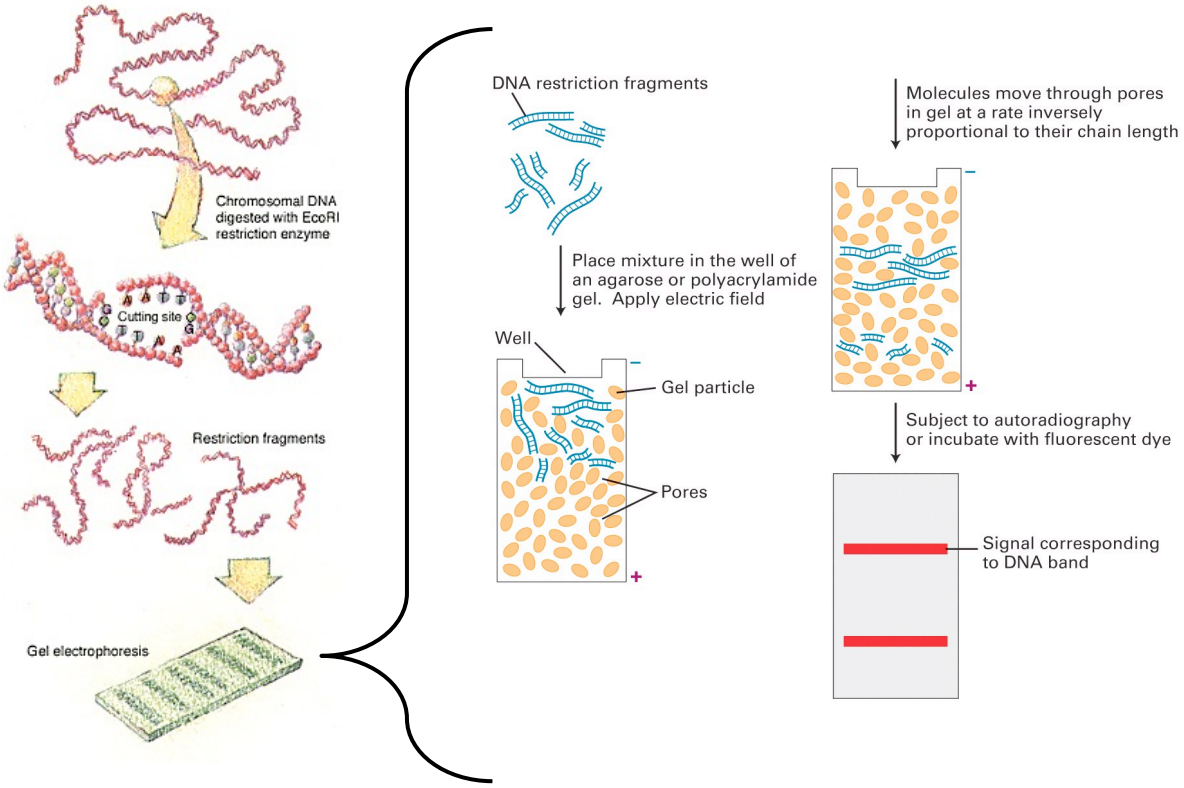
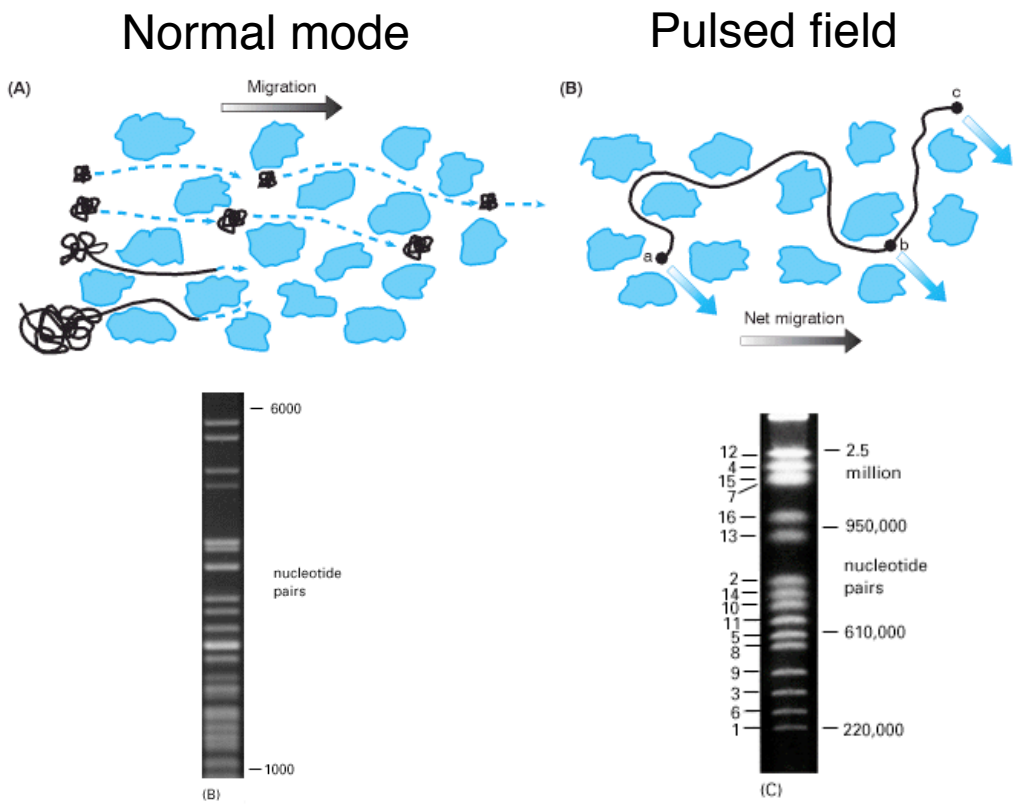


image credit: U.S. Department of Energy Human Genome Program, http://www.ornl.gov/hgmis.

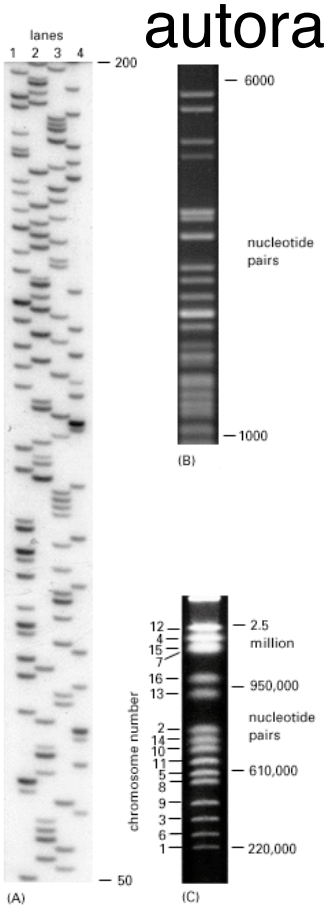
- Gel electrophoresis is used to separate DNA and RNA molecules by size and to estimate the size of nucleic acid molecules of unknown length by comparison with the migration of molecules of known length.
- DNA and RNA molecules are highly charged near neutral pH because the phosphate group in each nucleotide contributes one negative charge. As a result, DNA and RNA molecules move toward the positive electrode during gel electrophoresis. Smaller molecules move through the gel matrix more readily than larger molecules, so that molecules of different length, such as restriction fragments, separate. Because the gel matrix restricts random diffusion of the molecules, molecules of different length separate into “bands” whose width equals that of the well into which the original DNA mixture was placed.

Pulsed field agarose gel electrophoresis can separate very large DNA molecules



- Agarose gel electrophoresis with a constant electric field can resolve DNA fragments up to ≈20 kb in length. Larger DNAs, ranging from 20 kb to 10 megabases (Mb) in length, can be separated by size with pulsed-field gel electrophoresis.
- This technique depends on the unique behaviour of large DNAs in an electric field that is turned on and off (pulsed) at short intervals. When an electric field is applied to large DNA molecules in a gel, the molecules migrate in the direction of the field and also stretch out lengthwise. If the current then is stopped, the molecules begin to “relax” into random coils. The time required for relaxation is directly proportional to the length of a molecule.
- The electric field then is reapplied at 90° or 180° to the first direction. Longer molecules relax less than shorter ones during the time the current is turned off. Since the molecules must relax into a random coil before moving off in a new direction, longer molecules start moving in the direction imposed by the new field more slowly than shorter ones. Repeated alternation of the field direction gradually forces large DNA molecules of different size farther and farther apart.
- Pulsed-field gel electrophoresis is very important for purifying long DNA molecules up to ≈10^7 base pairs in length. Very large chromosomes must be digested into fragments of 10^7 base pairs or less before they can be analyzed. Such large restriction fragments can be generated with restriction enzymes that cut at rarely occurring 8-bp restriction sites.

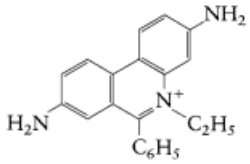
DNA fragments can be detected by autoradiography or fluorescent dyes



Detection of DNA fragments in gels:

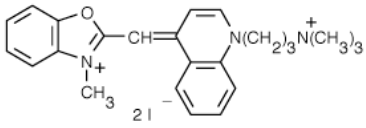
1. Radioactive label + photographic film or a phosphorimager
2. Fluorescent dye
   a. Ethidium bromide

Ethidium



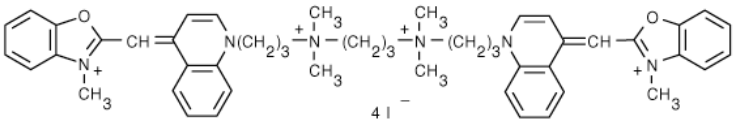
b. SYBR or monomeric cyanine dyes

Yo-pro-1



c. Dimeric cyanine dyes

Yoyo-1



http://www.probes.com/handbook/sections/0804.html

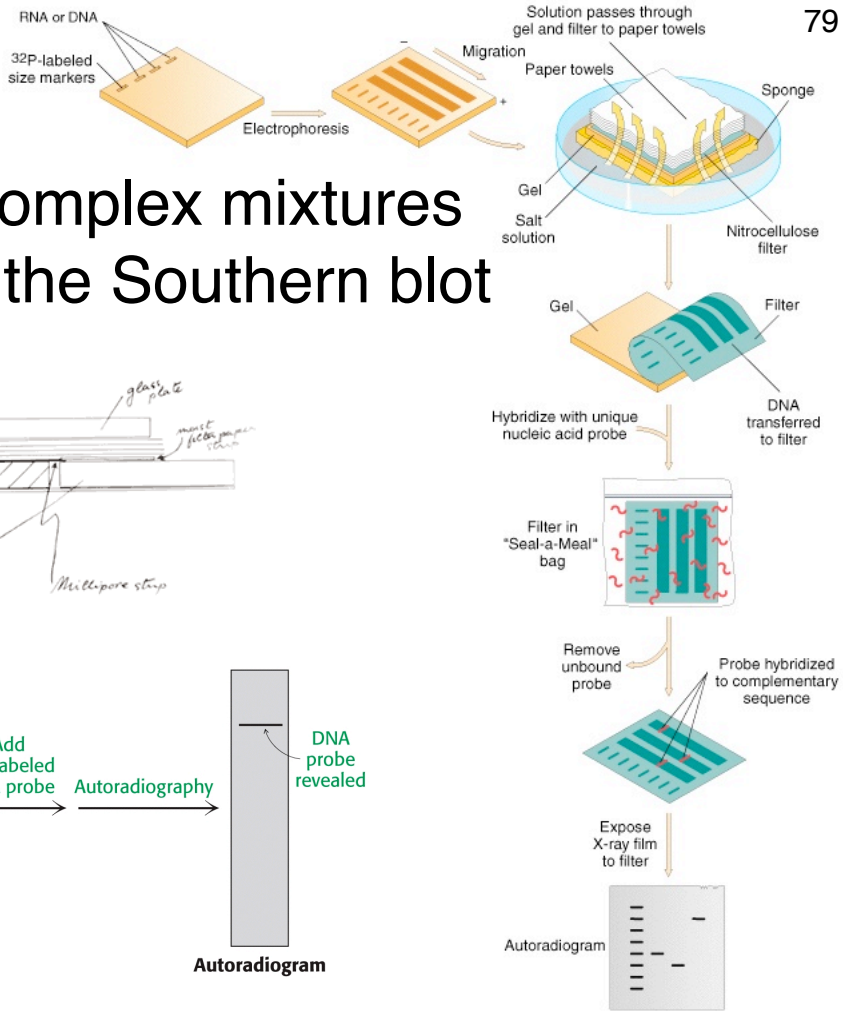
- Whereas agarose gel electrophoresis is typically used to resolve DNA molecules that differ by at least ~100 base paris or so, the resolving power of polyacrylamide gel electrophoresis is so great that single-stranded DNA molecules up to about 500 nucleotides long can be separated if they differ in length by only 1 nucleotide. This high resolution is critical to the DNA-sequencing procedures described later.
- DNA molecules composed of up to ≈2000 nucleotides can be separated electrophoretically on *polyacrylamide gels,* and molecules from 500 nucleotides to 20 kb on *agarose gels.*
- 3 gels labeled A to C:

  (A) a polyacrylamide gel is used to fractionate single-stranded DNA. In the size range 50 to 200 nucleotides, DNA molecules that differ in size by only a single nucleotide can be separated from each other.

  (B) an agarose gel with medium-sized pores is used to separate double-stranded DNA molecules. This method is most useful in the size range 300 to 10,000 nucleotide pairs.

  (C) the technique of pulsed-field agarose gel electrophoresis has been used to separate 16 different chromosomes that range in size from 220,000 to 2.5 million nucleotide pairs. DNA molecules as large as 107 nucleotide pairs can be separated in this way.
- Radioactively labeled DNA can be visualized by autoradiography of the gel. In this case, the gel is laid against a sheet of photographic film in the dark, exposing the film at the positions where labeled DNA is present. When the film is developed, a photographic image of the DNA is observed. Radiolabeled DNA bands also can be detected by laying the gel against a phosphorimager screen, which counts β-particles released by labeled molecules in the gel.
- If the DNA is not radiolabeled, the gel is incubated in a solution containing the fluorescent dye ethidium: This planar molecule binds to DNA by intercalating between the base pairs. Binding concentrates ethidium in the DNA and also increases its intrinsic fluorescence. As a result, when the gel is illuminated with ultraviolet light, the regions of the gel containing DNA fluoresce much more brightly than the regions of the gel without DNA.

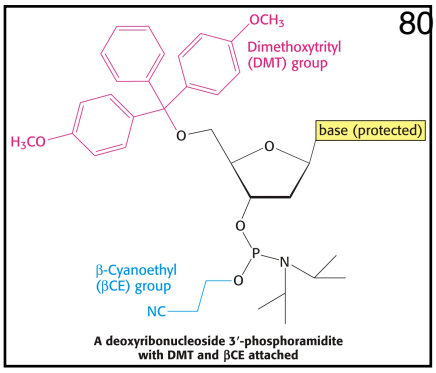
Analyzing very complex mixtures of genomic DNA: the Southern blot

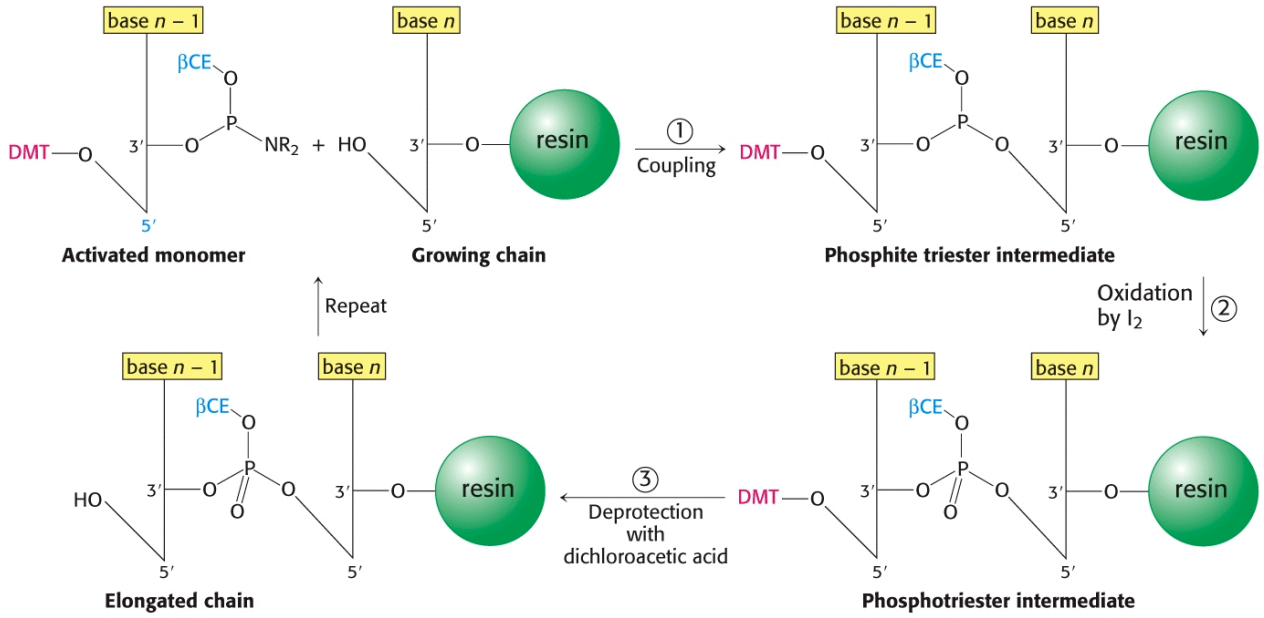


- "Gels used for electrophoresis of nucleic acids and proteins are permeable. This obvious fact didn't dawn on me until I tried to dissolve some agarose by floating it on a solution of sodium perchlorate and noticed a bead of liquid form on the top. I reasoned that if DNA molecules were carried through with the flow it would be possible to capture them on a nitrocellulose membrane, using the setup shown in the sketch. The big thrill came when, at the first attempt, I saw genes lit up as bands after hybridizing with radiolabeled probes. The *Journal of Molecular Biology* [initially] rejected the first manuscript1 as a 'methods paper' and the sketch is what I sent to people who had heard of the method and wanted to get on and use it." -- Ed Southern, Oxford University
- Restriction-enzyme digestion of genomic DNA can result in so many fragments that a stained electrophoretic gel shows a smear of DNA. A probe can identify one fragment in this mixture, with the use of a technique developed by E. M. Southern called Southern blotting.
- After DNA fragments are fractionated on the gel, an absorbent membrane is laid over the gel and the DNA bands are transferred (“blotted”) onto the membrane by capillary action. When transferred to the membrane, the DNA bands stay in the same relative positions as on the gel.
- The membrane is bathed in a labeled probe, and an autoradiogram is used to reveal the presence of any bands on the gel that are homologous to the probe. If appropriate, those bands can be cut out of the gel and further processed.
- The gel can be calibrated for DNA fragment size by running a standard “ladder” of fragments of known size on the same gel. Hence the sizes of any interesting fragments in the experimental sample can be inferred.
- E. Southern, "Detection of specific sequences among DNA fragments separated by gel-electrophoresis," *J Mol Biol*, 98:503, 1975. (Cited in > 30,666 papers) [see http://www.the-scientist.com/article/display/14216/]

Oligonucleotides can be made by chemical synthesis

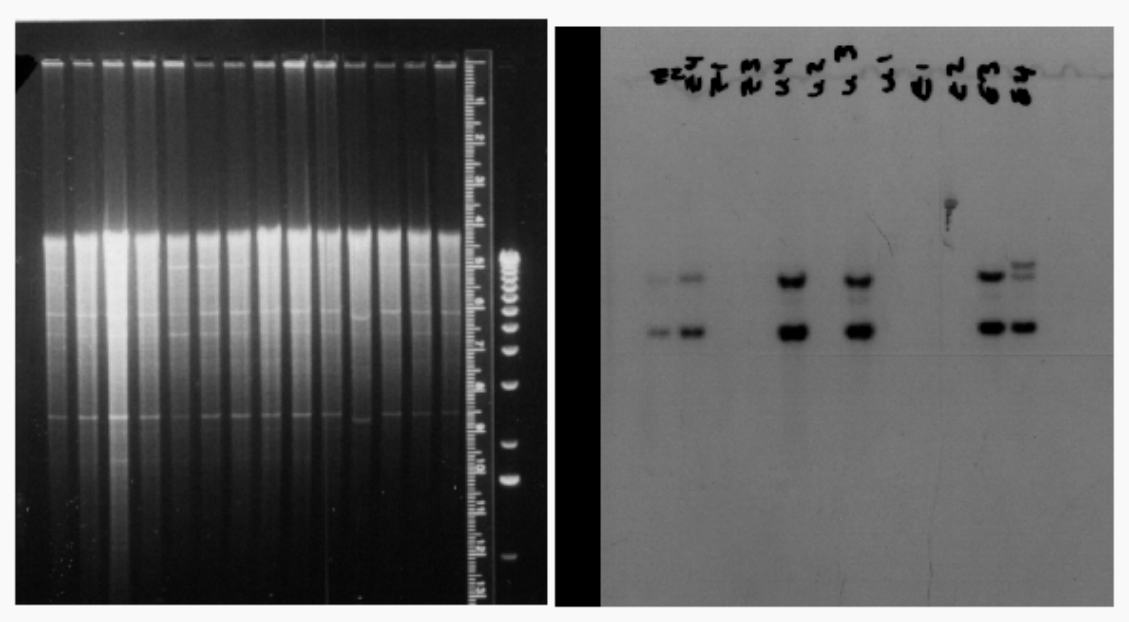
- http://www.invitrogen.com/
- https://www.sigma-genosys.com/
- http://www.idtdna.com/





- DNA strands can be synthesized by the sequential addition of activated monomers to a growing chain that is linked to an insoluble support.
- In step 1, the 3'-phosphorus atom of this incoming unit becomes joined to the 5'-oxygen atom of the growing chain to form a *phosphite triester*. The 5'-OH group and the 3'-phosphoryl group of the activated monomer are rendered unreactive by attachment of protecting groups.
- In step 2, the phosphite triester (in which P is trivalent) is oxidized by iodine to form a *phosphotriester* (in which P is pentavalent).
- In step 3, the DMT protecting group on the 5'OH of the growing chain is removed by the addition of dichloroacetic acid, which leaves other protecting groups intact. The DNA chain is now elongated by one unit and ready for another cycle of addition.
- Each cycle takes only about 10 minutes and elongates more than 98% of the chains.
- At the end of the synthesis, NH_3 is added to remove all protecting groups and release the oligonucleotide from the solid support.
- The sample can be purified by high-pressure liquid chromatography or by electrophoresis on polyacrylamide gels.
- Groups other than nucleotides can be attached to the DNA. An oligonucleotide labeled at one end with ^{32}P or a fluorescent tag can be used to search for a complementary sequence in a very long DNA molecule or even in a genome consisting of many chromosomes.

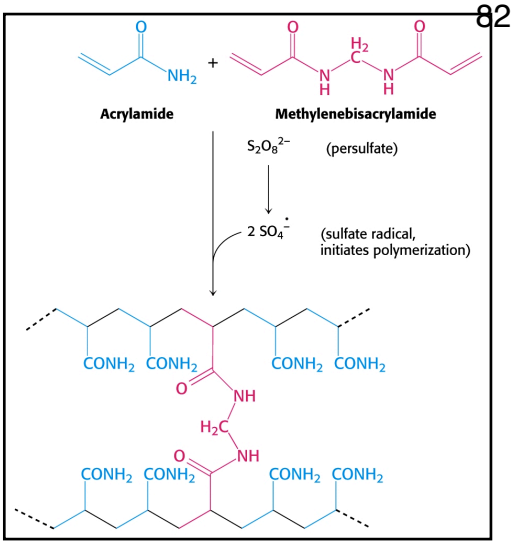
Southern blot example

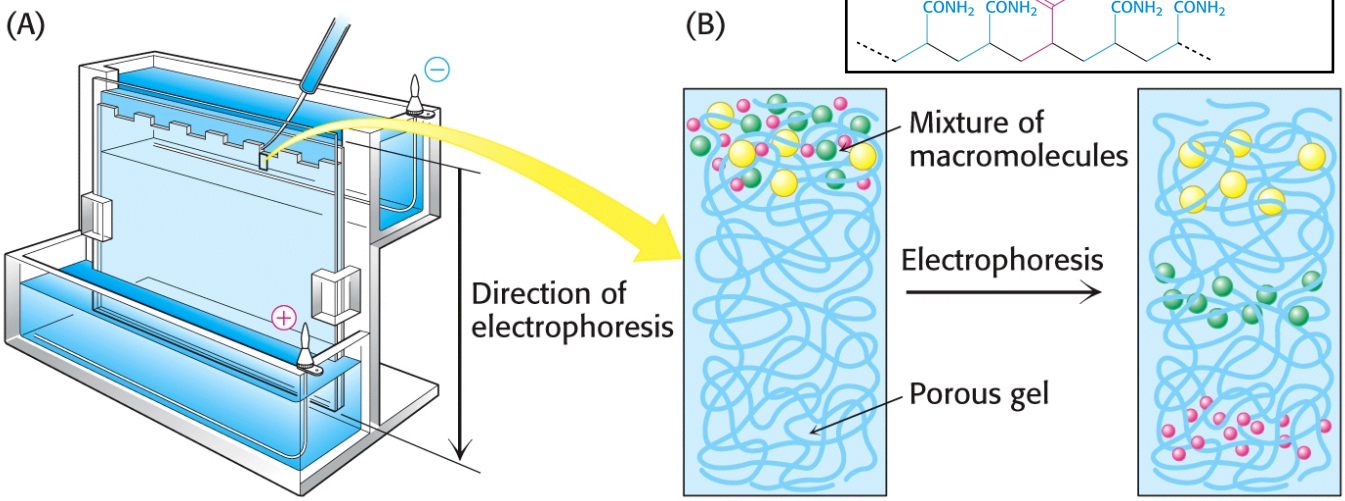


http://www.bio.davidson.edu/courses/Bio111/RealSouthernblot.html

- The figure on the left shows a photograph of a 0.7% agarose gel that has 14 different samples loaded on it (plus molecular weight marker in the far right lane and a glowing ruler used for analysis of the results). The samples are genomic DNA isolated from different strains of the unicellular green alga *Chlamydomonas reinhardtii*. Each sample of DNA has been digested with the same restriction enzyme (EcoRI). Notice that the DNA does not appear as a series of discrete bands but rather as a smear. [Why is that?]
- This DNA was transferred to nitrocellulose and then probed with a radioactive fragment of DNA that is also present in a subset of the DNA isolated from the different strains. The figure on the right is a copy of the X-ray film and reveals which strains contain the target DNA and which ones do not.
- *Question: Question about southern blot. Is the DNA probe single strand or double strand? How about those used in FISH and SKY?*
- *Answer: If a probe is synthetic, it will definitely be single stranded. However, in some cases the probe might be prepared by PCR and so it would be double stranded. The same could be true for all techniques, including southern, FISH and SKY.*
- *Question: Is Eastern blotting or lectin blotting an important part of identifying post translational mods in lipid and carbohydrates or proteins or is it better to just use a combination of various methods followed by a more well known blotting technique?*
- *Answer: It is my understanding that Eastern blotting is not well-defined, and the name has been used over the years for a variety of different applications. None of these applications is as universal as the southern/northern/western approaches which are well-defined and well established. You'd have to somewhat egotistical to give a technique that you've invented the name 'Eastern'.*

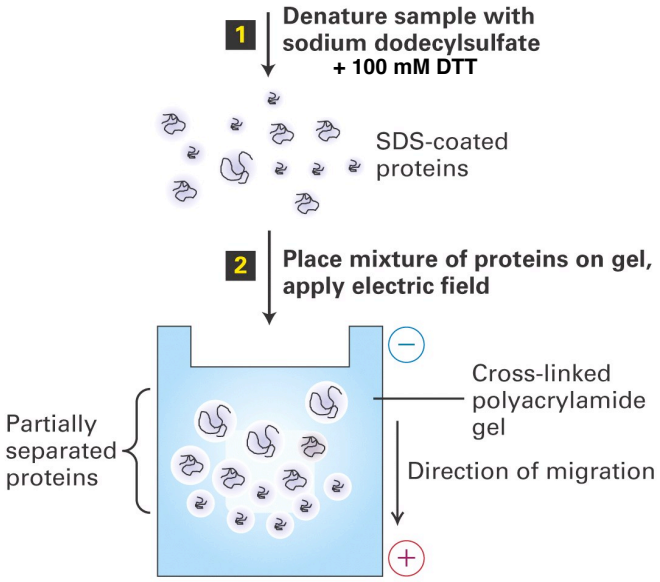
Polyacrylamide gel electrophoresis

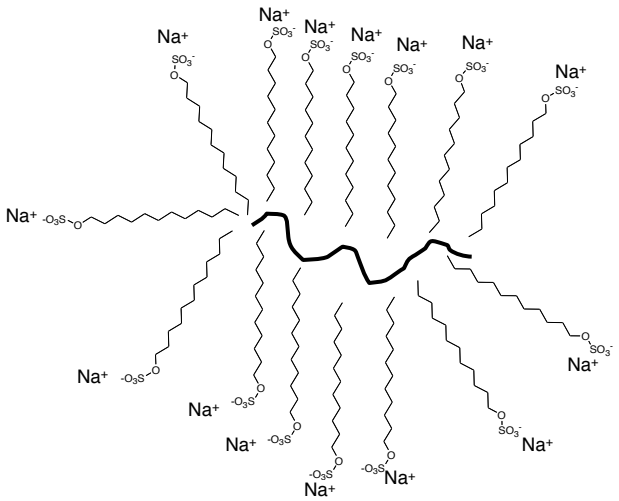




- Electrophoretic separations are nearly always carried out in gels (or on solid supports such as paper) because the gel serves as a molecular sieve that enhances separation. Molecules that are small compared with the pores in the gel readily move through the gel, whereas molecules much larger than the pores are almost immobile. Intermediate-size molecules move through the gel with various degrees of facility.
- Protein and DNA polyacrylamide gel electrophoresis is normally performed in a thin, vertical slabs. The direction of flow is from top to bottom. Polyacrylamide gels, formed by the polymerization of acrylamide and cross-linked by methylenebisacrylamide, are choice supporting media for electrophoresis because they are chemically inert and are readily formed.
- The fundamental difference between electrophoresis and gel filtration in that in PAGE all of the molecules, regardless of size, are forced to move through the same matrix. In gel filtration larger molecules 'see' a matrix that is different than what the smaller molecules 'see'.
- The degree of crosslinking is proportional to the concentration of bisacrylamide. The more crosslinking, the finer the mesh, the higher the effective viscosity of the gel. This leads to better separation of smaller proteins.
- SDS-PAGE gel is synthesized from a solution containing: Acrylamide, Bis-acrylamide, TEMED (N,N,N´,N´-tetramethylenediamine), APS (ammonium persulfate, $(NH_4)_2S_2O_8$), SDS, Tris buffer
- SDS-sample buffer contains: Glycerol, SDS, Tris, DTT, Bromophenol blue
- **What is the purpose of each of these components?**

Sodium dodecylsulfate polyacrylamide gel (SDS-PAGE) electrophoresis

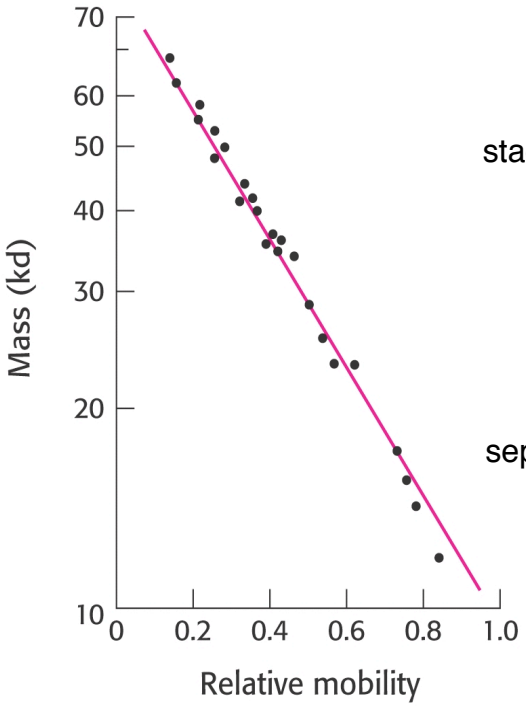


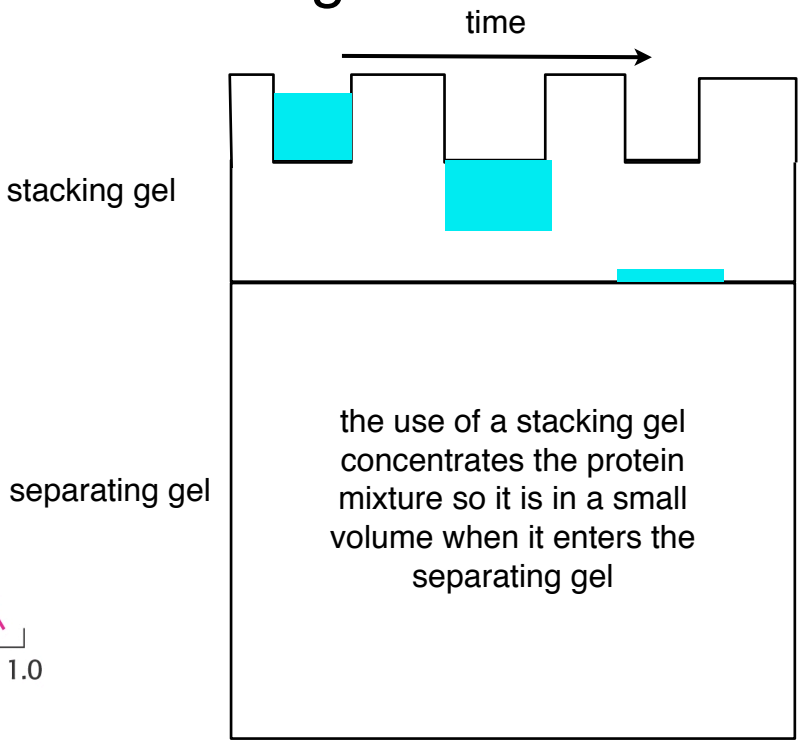


..but isn't v going to constant for all SDS-coated proteins?

- Proteins can be separated largely on the basis of mass by electrophoresis in a polyacrylamide gel under denaturing conditions. The mixture of proteins is first dissolved in a solution of sodium dodecyl sulfate (SDS), an anionic detergent that disrupts nearly all noncovalent interactions in native proteins. This technique is referred to as SDS-polyacrylamide gel electrophoresis (SDS-PAGE).
- Mercaptoethanol (2-thioethanol) or dithiothreitol also is added to reduce disulfide bonds. Anions of SDS bind to main chains at a ratio of about one SDS anion for every two amino acid residues.
- This complex of SDS with a denatured protein has a large net negative charge that is roughly proportional to the mass (actually the number of amino acids) of the protein. The negative charge acquired on binding SDS is usually much greater than the charge on the native protein; this native charge is thus rendered insignificant. The SDS-protein complexes are then subjected to electrophoresis.
- All proteins have roughly the same size to charge ratio so this is not the basis for the separation in electrophoresis.
- The separation is actually due to the sieving mechanism of the gel and small molecules can move faster than larger molecules.

Relative mobilities of proteins in a gel depends on size of the molecules and the amount of crosslinking

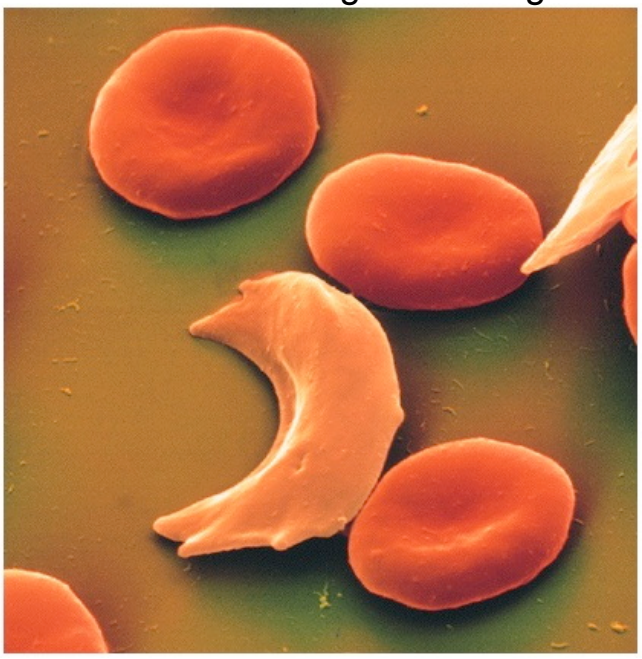


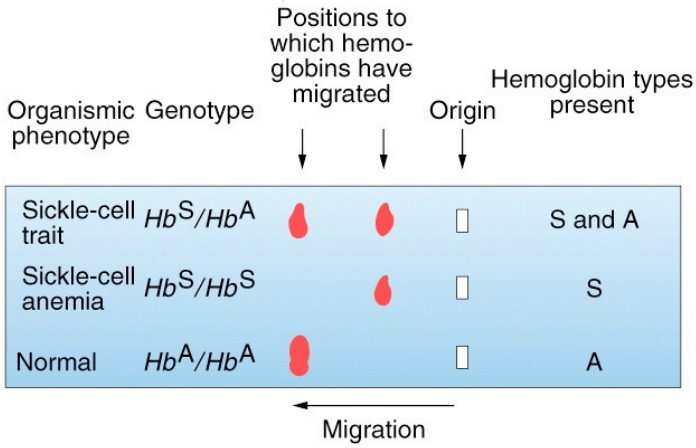


- Typical sample volumes in SDS-PAGE are 10-20 microlitre. This is a relatively large volume that is loaded in well that is much deeper than the width of the protein ‘bands’ that we would hope to resolve. That is, the sample volume is large relative to the resolving power of the gel.
- A solution to this problem is to use a stacking gel. A stacking gel is a very loosely cross linked gel in which all molecules will run quickly. When they reach the edge of the resolving gel, the molecules greatly slow down and effectively ‘pile up’ at the interface.
- This means that the sample entering the resolving gel is in a much smaller effective volume than the sample loaded in the well.

Electrophoresis can be used to diagnose diseases

Red blood cells = bags of hemoglobin





Two copies of *Hb* gene

Hb^A/Hb^A: Normal; red blood cells never sickle.

Hb^S/Hb^S: Severe, often fatal anemia; abnormal hemoglobin causes red blood cells to have sickle shape.

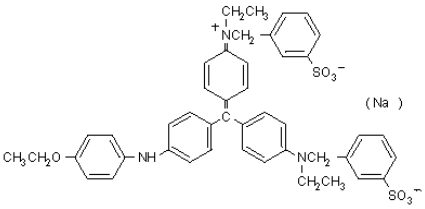
HB^A/Hb^S: No anemia; red blood cells sickle only under low oxygen concentrations.

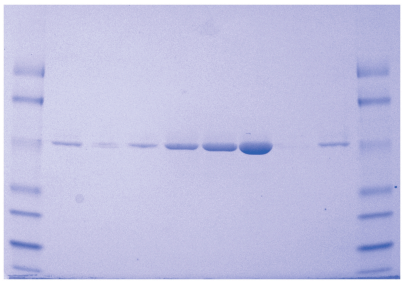
Hb^A and Hb^S differ by a single glutamic acid to valine mutation in the protein

- HbA is the hemoglobin from normal adults and hemoglobin S (HbS) is the mutated protein that causes sickle-cell anemia; the disease in which red blood cells take on a sickle-cell shape.
- The difference between these two proteins is substitution of valine for glutamic acid at position 6. This mutation is all that is needed to produce the defective hemoglobin. Unless patients with HbS receive medical attention, this single error in one amino acid in one protein will hasten their death.
- Because the A and S forms of hemoglobin have different charges, so they can be separated by electrophoresis. We see that homozygous normal people have one type of hemoglobin (A) and anemics have type S, which moves more slowly in the electric field. The heterozygotes have both types, A and S.
- Why does HBA move further down the gel than HBS when they are identical in size?

Staining of proteins on SDS-PAGE gels

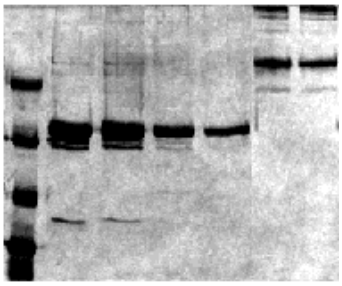
1.Coomassie



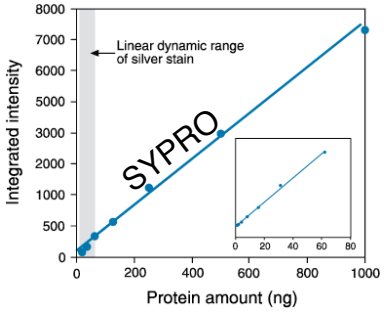


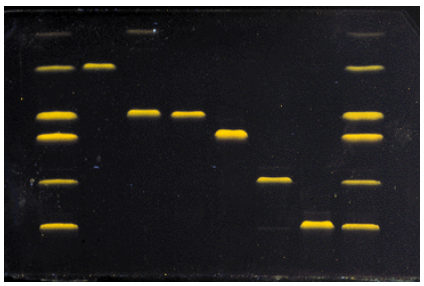
2.Silver staining

$$Ag^{+} \xrightarrow{\text{protein}} Ag$$



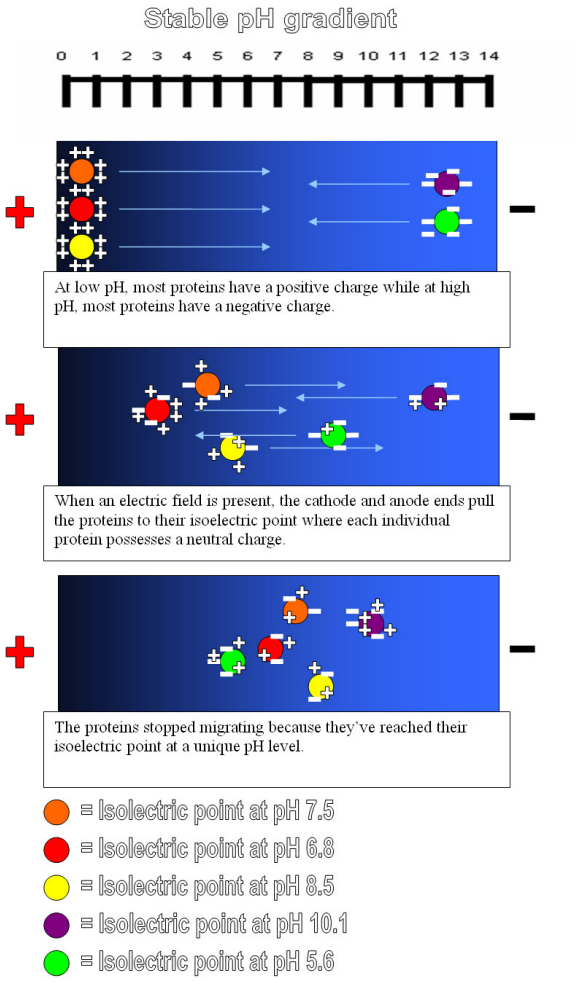
3.Fluorescent staining

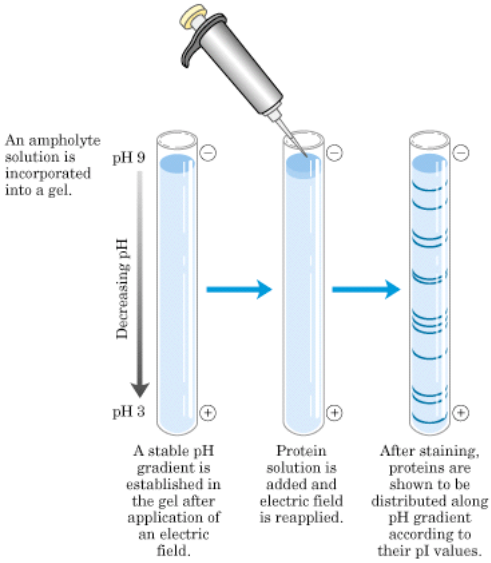




- As little as 0.1 µg (~2 pmol) of a protein gives a distinct band on an SDS-PAGE gel when stained with Coomassie blue, and even less (~0.02 µg) can be detected with a silver stain. Proteins that differ in mass by about 2% (e.g., 40 and 41 kd, arising from a difference of about 10 residues) can often be distinguished. Fluorescent staining can detect < 1 ng of protein which is comparable with optimized silver staining protocols.
- Coomassie brilliant blue: relatively insensitive but quick and easy. Recall the Bradford assay that used the same principle.
- Silver Staining: much more sensitive but more steps. Recall the Lowry assay in which proteins were used to reduce Cu^{2+} to Cu^{1+}.
- Fluorescent Staining: Molecular probes sells SYPRO orange stain. Will require specialized (but relatively low cost) imaging equipment. SYPRO is likely to be very similar to NanoOrange (see the lecture on protein detection) in terms of mechanism of fluorescence enhancement. It likely interacts with the protein/SDS complex and increases its fluorescence in this relatively hydrophobic environment.

Isoelectric Focusing

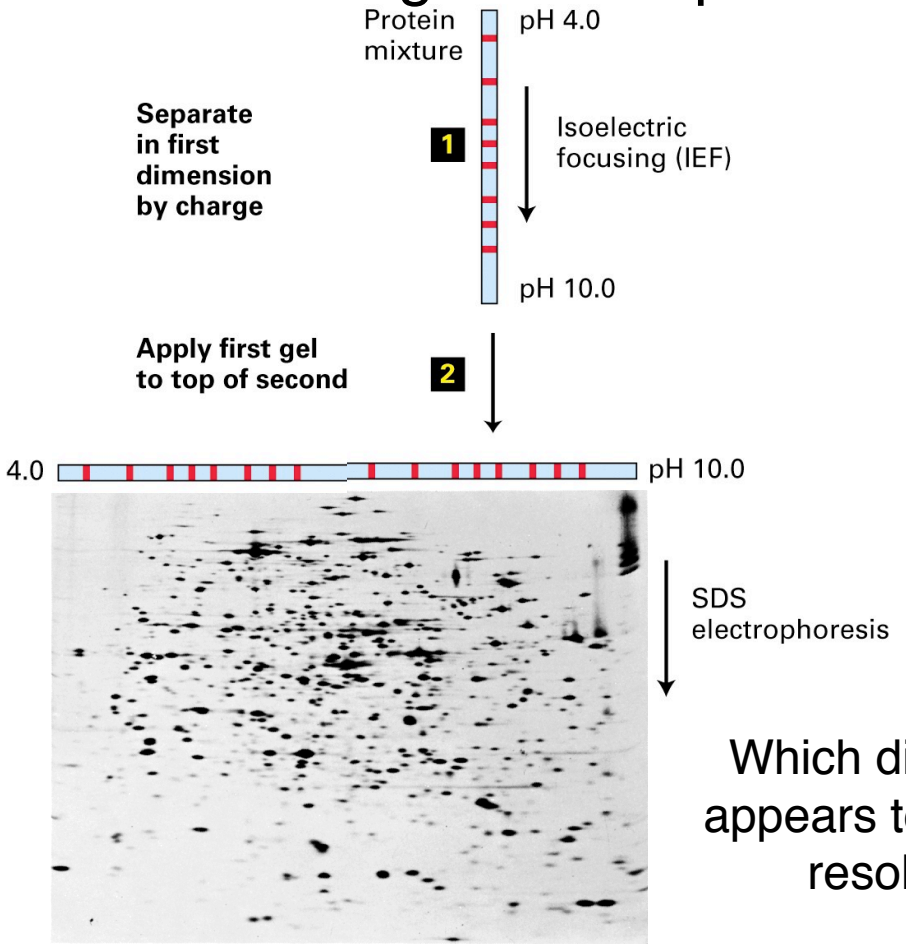




Note: no SDS because you don't want to mask the charges that make proteins different

- Proteins can also be separated electrophoretically on the basis of their relative contents of acidic and basic residues.
- The *isoelectric point* (pI) of a protein is the pH at which its net charge is zero. At this pH, its electrophoretic mobility is zero because *z* in the equation provided earlier in the notes is equal to zero. For example, the pI of cytochrome *c,* a highly basic electron-transport protein, is 10.6, whereas that of serum albumin, an acidic protein in blood, is 4.8.
- Suppose that a mixture of proteins undergoes electrophoresis in a pH gradient in a gel in the absence of SDS. Each protein will move until it reaches a position in the gel at which the pH is equal to the pI of the protein. This method of separating proteins according to their isoelectric point is called *isoelectric focusing*.
- The pH gradient in the gel is formed first by subjecting a mixture of *polyampholytes* (small multicharged polymers) having many pI values to electrophoresis. Isoelectric focusing can readily resolve proteins that differ in pI by as little as 0.01, which means that proteins differing by one net charge can be separated.
- Images: http://commons.wikimedia.org/wiki/File:Isoelectric_focusing_contribute2.jpg, http://www.biochem.arizona.edu/classes/bioc462/462a/462a.html

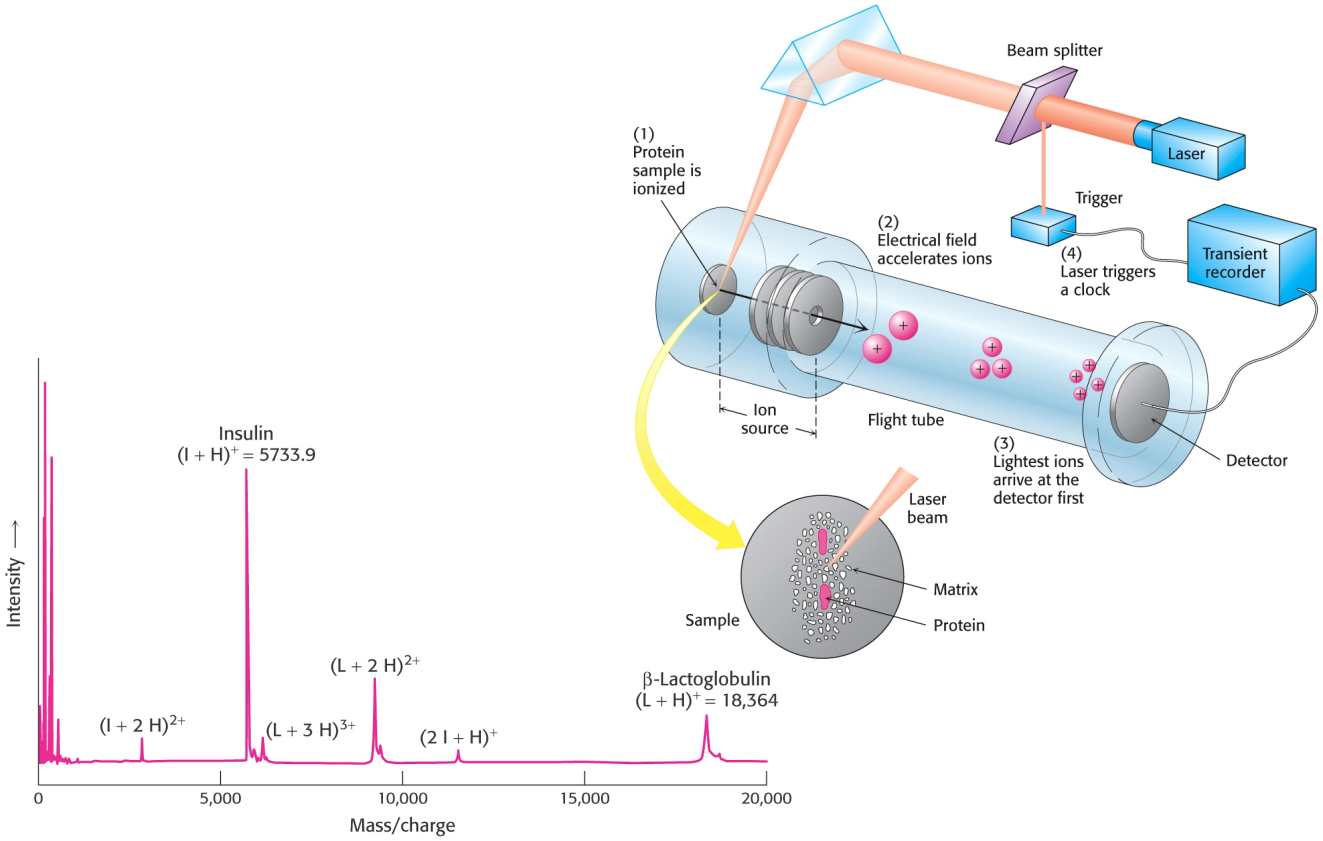
2-dimensional gel electrophoresis



Which dimension appears to be more resolved?

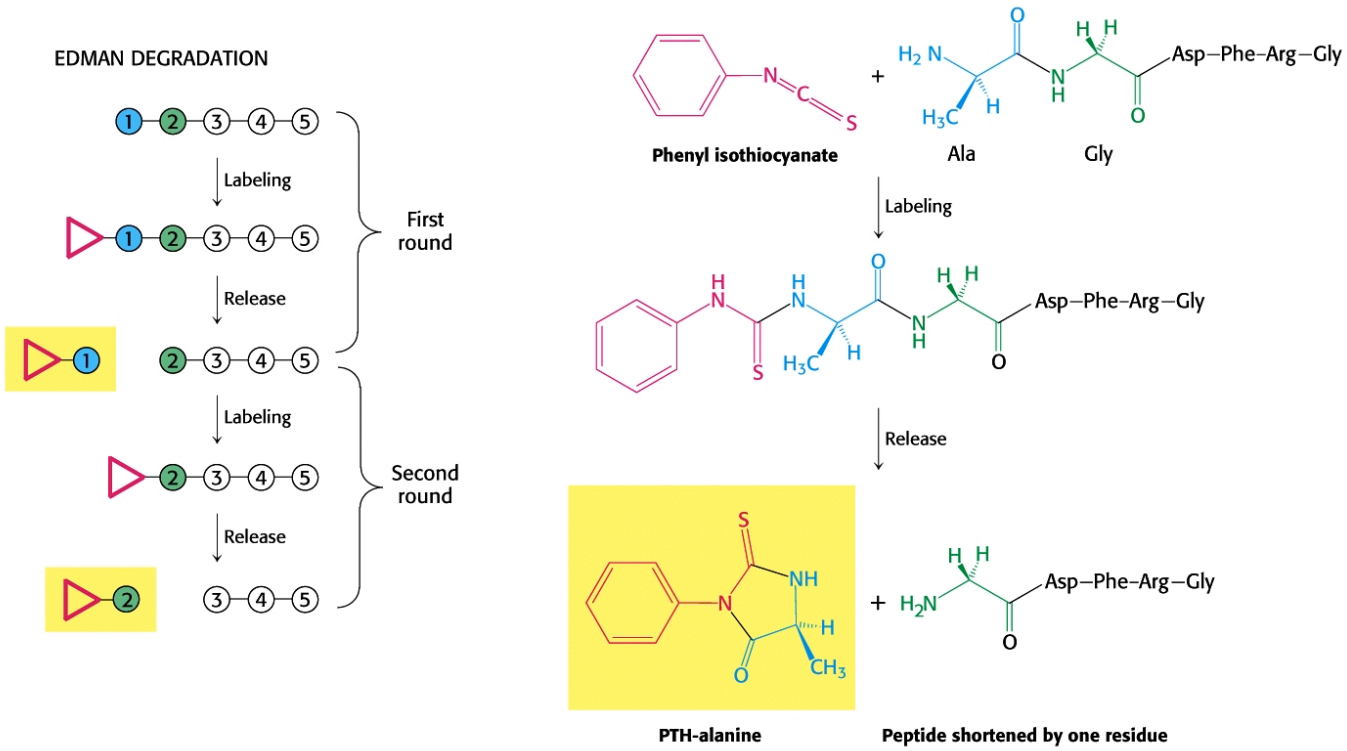
- Isoelectric focusing can be combined with SDS-PAGE to obtain very high resolution separations.
- A single sample is first subjected to isoelectric focusing. This single-lane gel is then placed horizontally on top of an SDS-polyacrylamide slab. The proteins are thus spread across the top of the polyacrylamide gel according to how far they migrated during isoelectric focusing. They then undergo electrophoresis again in a perpendicular direction (vertically) to yield a 2-dimensional pattern of spots. In such a gel, proteins have been separated in the horizontal direction on the basis of isoelectric point and in the vertical direction on the basis of mass.
- It is remarkable that more than a thousand different proteins in the bacterium *Escherichia coli* can be resolved in a single experiment by two-dimensional electrophoresis. Proteins isolated from cells under different physiological conditions can be subjected to two-dimensional electrophoresis, followed by an examination of the intensity of the signals. In this way, particular proteins can be seen to increase or decrease in concentration in response to the physiological state.
- But how can we tell what protein is being regulated? A former drawback to the power of the two-dimensional gel is that, although many proteins are displayed, they are not identified. It is now possible to identify proteins by coupling two-dimensional gel electrophoresis with mass spectrometric techniques.
- *Question: I don't really recall going over the details of orthogonal chromatography (the technique used now a days instead of 2D gels).*
- *Answer: We didn't go over this in class, but the idea is pretty similar to things we've seen elsewhere. Basically, instead of running a gel in two dimensions, you can do steps of chromatography where the separation is based on different principles.*

Determining the identity of an unknown protein on a gel: Mass spectroscopy



- Mass spectrometry has been an established analytical technique in organic chemistry for many years. Until recently, however, the very low volatility of proteins made mass spectrometry useless for the investigation of these molecules. This difficulty has been circumvented by the introduction of techniques for effectively dispersing proteins and other macromolecules into the gas phase. These methods are called *matrix-assisted laser desorption-ionization (MALDI)* and *electrospray spectrometry*.
- In this MALDI technique, protein ions are generated and then accelerated through an electrical field. They travel through the flight tube, with the smallest traveling fastest and arriving at the detector first. Thus, the *time of flight (TOF)* in the electrical field is a measure of the mass (or, more precisely, the mass/charge ratio).
- Mass spectrometry has permitted the development of *peptide mass fingerprinting*. This technique for identifying peptides has greatly enhanced the utility of two-dimensional gels. Two-dimensional electrophoresis is performed and the sample of interest is extracted and cleaved *specifically* by chemical or enzymatic means. The masses of the protein fragments are then determined with the use of mass spectrometry. Finally, the peptide masses, or *fingerprint*, are matched against the fingerprint found in databases of proteins that have been "electronically cleaved" by a computer simulating the same fragmentation technique used for the experimental sample. Mass spectrometry has now provided name tags for many of the proteins in 2-dimensional gels.

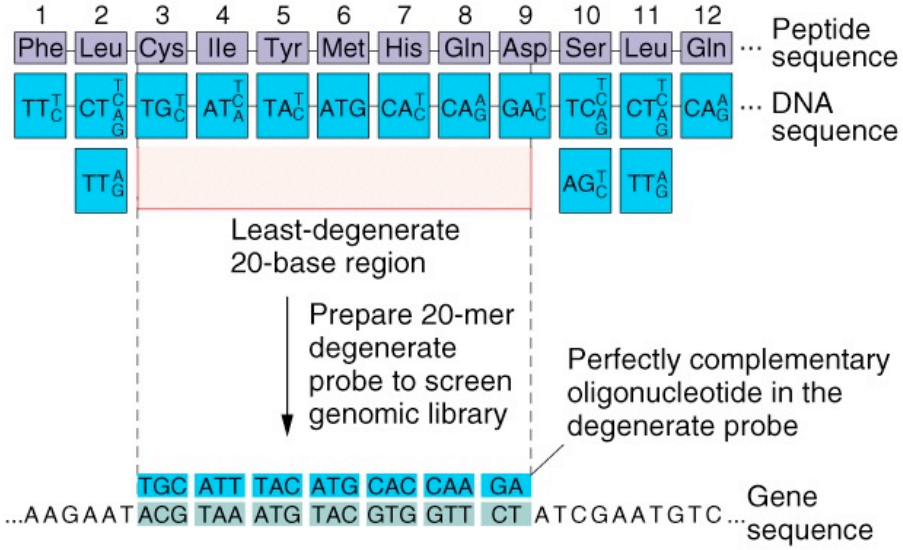
Determining the identity of an unknown protein on a gel: Edman Degradation



- The *Edman degradation* sequentially removes one residue at a time from the amino end of a peptide. *Phenyl isothiocyanate* reacts with the uncharged terminal amino group of the peptide to form a phenylthiocarbamoyl derivative. Then, under mildly acidic conditions, a cyclic derivative of the terminal amino acid is liberated, which leaves an intact peptide shortened by one amino acid. The cyclic compound is a phenylthiohydantoin (PTH)-amino acid, which can be identified by chromatographic procedures. The sample could be spiked with a mix of standards corresponding to each of the PTH version of the 20 common amino acids in order to unambiguously identify the unknown amino acid.
- The Edman procedure can then be repeated again and again on the ever shortening peptide, yielding another PTH-amino acid, which can again be identified by chromatography.
- The development of automated sequencers has markedly decreased the time required to determine protein sequences. One cycle of the Edman degradation—the cleavage of an amino acid from a peptide and its identification—is carried out in less than 1 hour. By repeated degradations, the amino acid sequence of some 50 residues in a protein can be determined.
- High-pressure liquid chromatography provides a sensitive means of distinguishing the various amino acids. Using modern techniques it is feasible to analyze the sequence of a protein sample eluted from a single band of an SDS-polyacrylamide gel.

With a little bit of protein sequence, you can design a probe for a gene

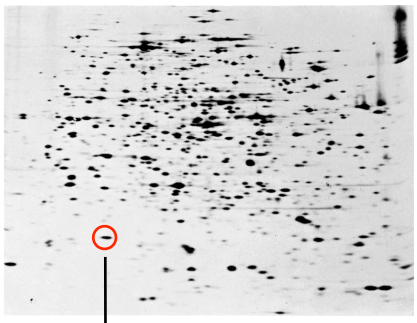
1. A known gene in a related organism
2. The sequence of the protein



How many different probes are in this degenerate 20mer?

- A source of DNA for a probe might be a homologous gene from a related organism. For example, if a certain gene has been cloned in the ascomycete fungus *Neurospora,* then it is very likely that this gene can be used as a probe to find the homologous gene in the related fungus *Podospora.* This method depends on the evolutionary conservation of DNA sequences through time.
- Even though the probe DNA and the DNA of the desired clone might not be identical, they are often similar enough to promote hybridization.
- Probe DNA can be synthesized if the protein product of the gene of interest is known and an amino acid sequence has been obtained. However, because of the redundancy of the code several possible DNA sequences could have encoded the protein in question.
- To get around this problem, a short stretch of amino acids with minimal redundancy is selected and a “cocktail” of oligonucleotides is used as a probe. The correct strand within this cocktail would find the gene of interest.

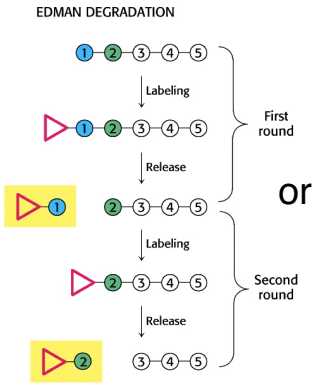
With a little bit of protein sequence you can go after the whole gene using molecular biology

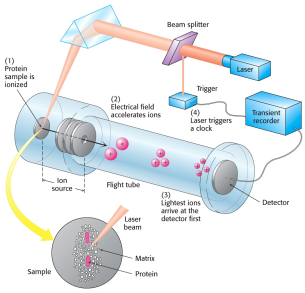


Structure/function studies on recombinant protein

Isolate ('fish out') gene for protein of interest

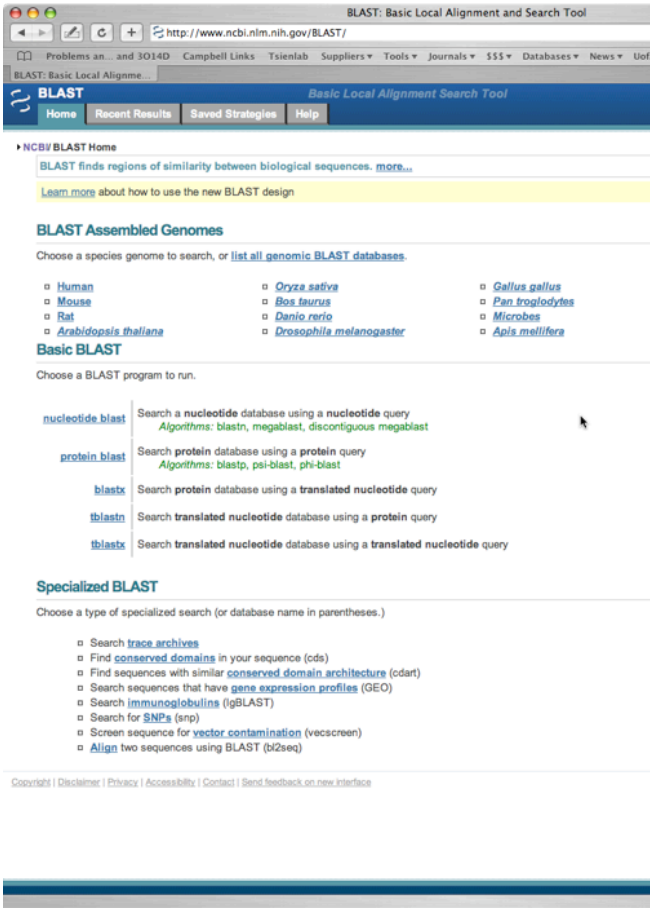
DNA probe + genomic library





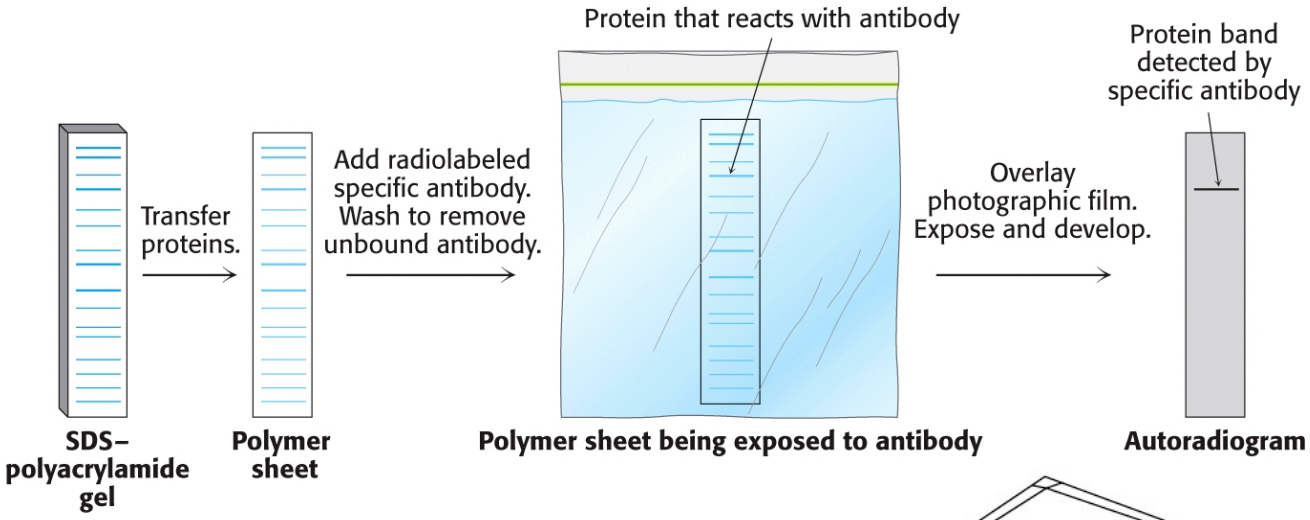
- Many proteins were sequenced by Edman degradation of peptides derived from specific cleavages.
- Nowadays, a complementary experimental approach based on recombinant DNA technology is often more efficient. Long stretches of DNA can be cloned and sequenced, and the nucleotide sequence directly reveals the amino acid sequence of the protein encoded by the gene.
- For example, to investigate an unknown spot on a gel, a *probe for a gene of interest can be prepared if only a small part of the amino acid sequence of the protein encoded by the gene is known*. This small portion of the gene could be used as one primer in a PCR reaction. It could potentially be used following reverse transcription of mRNA in conjunction with a polyT primer in order to PCR amplify the whole gene.
- However, a problem arises because a given peptide sequence can be encoded by a number of oligonucleotides but this can be overcome.
- Even with the use of the DNA base sequence to determine primary structure, there is still a need to work with isolated proteins. The amino acid sequence deduced by reading the DNA sequence is that of the *nascent* protein, the direct product of the translational machinery. Many proteins will have postranslational modifications. Chemical analyses of proteins in their final form are needed to delineate the nature of these changes, which are critical for the biological activities of most proteins. *Thus, genomic and proteomic analyses are complementary approaches to elucidating the structural basis of protein function*

www.ncbi.nlm.nih.gov/



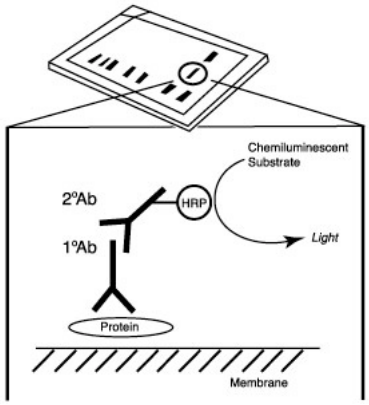
- The NCBI is basically a one-stop shop for all of your biotechnology needs.
- This site includes the powerful ‘PubMed’ database, a variety of sequence databases, the PubChem database, and a variety of other resources. One under-appreciated resource is the large number of textbooks that are freely available online at this site.
- For an overview, see http://www.ncbi.nlm.nih.gov/sites/gquery?itool=toolbar or just click ‘All Databases’ on the menu bar on the top left.
- ‘BLAST’ is the tool for searching sequence databases.
- Depending on the particular ‘flavor’ of BLAST you pick, you can search nucleotide or protein databases. The protein database can be restricted to the subset for which structures are known.
- These tools allow you to take a known sequence, perhaps one that you have identified experimentally, and search large databases to find sequences of DNA or protein that are perfect or close matches to the submitted sequences.
- Standard protein BLAST is designed for protein searches.
- Position-Specific Iterated (PSI)-BLAST is the most sensitive BLAST program, making it useful for finding very distantly related proteins .
- If I do a protein-protein BLAST search and limit database to Protein Data Bank (pdb), I get only those proteins whose structures have been determined by x-ray crystallography or NMR.

The Western blot



- needless to say, enzyme reporters (e.g., HRP plus luminol-type substrate) have largely replaced radioactivity for Western blot detection.

- typically would use an unlabelled primary Ab plus a labeled secondary Ab

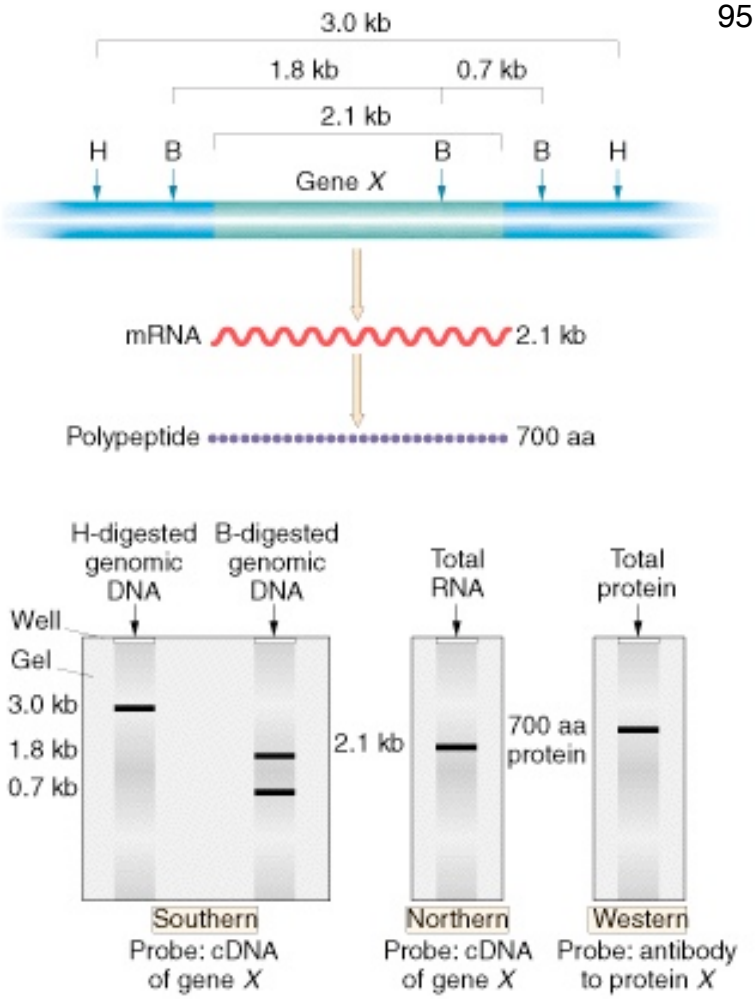


www.cellsignal.com

- Often it is necessary to detect small quantities of a particular protein in the presence of many other proteins, such as a viral protein in the blood. Very small quantities of a protein of interest in a cell or in body fluid can be detected by an immunoassay technique called *Western blotting*.
- A sample is subjected to electrophoresis on an SDS-polyacrylamide gel. Blotting (or more typically electroblotting) transfers the resolved proteins on the gel to the surface of a polymer sheet to make them more accessible for reaction. An antibody that is specific for the protein of interest is added to the sheet and reacts with the antigen. The antibody-antigen complex on the sheet then can be detected by rinsing the sheet with a second antibody specific for the first (e.g., goat antibody that recognizes mouse antibody). A radioactive label on the second antibody produces a dark band on x-ray film (an autoradiogram).
- The term Western blotting as a synonym for immunoblotting arose because the comparable technique for detecting specific DNA sequences is known as Southern blotting, after Ed Southern who devised it, which in turn provoked the name Northern for blots of size-separated RNA, and Western for blots of size-separated proteins. Western blots have many applications in basic research and clinical diagnosis. They are often used to test sera for the presence of antibodies to specific proteins, for example to detect antibodies to different constituents of the human immunodeficiency virus, HIV.
- *Question: Why is it that a second antibody, which is radio-labelled, is added to bind to the antibody which is bound to your protein of interest? Why couldn't you just radiolabel the antibody that binds to the protein?*
- *Answer: You could also just radiolabel the primary antibody. However, there are two advantages of using a labeled secondary antibody. The first is that one secondary can work for many different primary antibodies so you don't have to do the labelling reaction as often (or purchase so many different radioactive antibodies). The second is that several secondary antibodies will bind to each primary antibody so your signal will be amplified.*
- *Question: If the proteins are essentially unfolded upon SDS-PAGE, how can we detect it with a specific Ab during Western blotting, considered that the recognition site might be destroyed. Could the protein fold back to its natural state after blotting?*
- *Answer: The most important thing to keep in mind is the commercially available antibodies have been validated for different techniques, including Westerns, immunohistochemistry, and immunoprecipitation. So there are probably some antibodies that don't recognize their target proteins once they've been blotted (due to a folded epitope being destroyed), but these antibodies would not be sold for the purpose of Westerns. The antibodies that do work for Westerns probably recognize a linear, unfolded, epitope. Or maybe the target protein is able to partially fold back to it's natural state on the membrane.*

Comparison of Southern, Northern, and Western blots

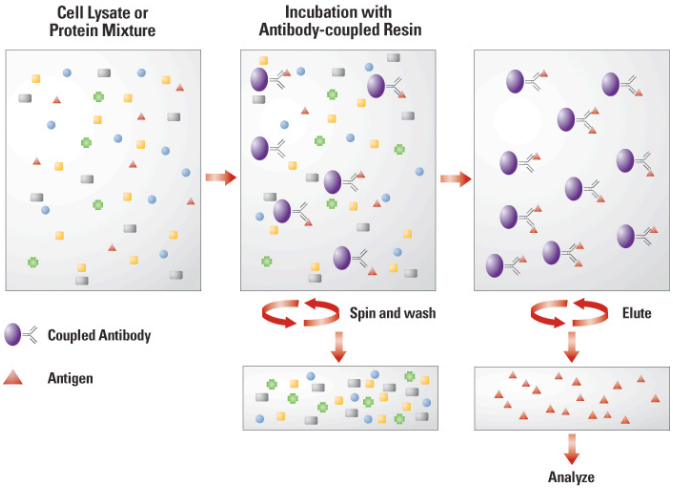
For a given gene, the molecular weight of a band on Northern and a Western should correlate. There does not have to be any correlation with a Southern. Why?



- We have just learned about Southern blots and we have previously discussed Western blots.
- The Southern blotting technique can be extended to detect a specific *RNA* molecule from a mixture of RNAs fractionated on a gel. This technique is called Northern blotting to contrast it with the Southern technique for *DNA* analysis. In Northern blotting, the density of a band is proportional to the amount of mRNA present. The Northern blot is widely used to look at changes in mRNA levels upon different treatments of cells or tissue. For example, you would expect to see differences in the levels of some mRNA molecules between healthy cells and cancer cells. By figuring out which genes are being more actively transcribed in cancer cells, you might be able to identify a good target for a cancer therapy (e.g., by developing an inhibitor of the protein product).
- The fractionated RNA is blotted onto a membrane and probed in the same way as in Southern blotting. RNA is extracted from the appropriate cell sample and then electrophoresed, blotted, and probed with the cloned gene in question. A positive signal shows the presence of the transcript.
- Hence we see that cloned DNA finds widespread application as a probe for detecting either a specific clone or a specific DNA fragment or a specific RNA. In all these cases, the ability of nucleic acids with complementary nucleotide sequences to find and bind to each other in solution is being exploited.
- Western, Southern, and Northern blots are very powerful experimental tools in detecting and sizing specific macromolecules in molecular genetics.
- Note that the size of the bands should correlate on a Northern and a Western but not necessarily with a Southern!
- *Question: You mentioned that Northern blot is correlated to Western blot because the size of mRNA should correspond to the mass of protein. However, eukaryotic mRNA contains UTR and polyA tail and prokaryotic mRNA even encode multiple gene on a single RNA strand. In such case, the size relationship between particular mRNA and protein upon blotting doesn't make sense to me.*
- *Answer: You are correct that the UTR regions means that the correlation is not perfect. However, a longer protein will tend have a longer mRNA and a shorter protein will tend to have a shorter mRNA. So they are correlated, just not perfectly.*

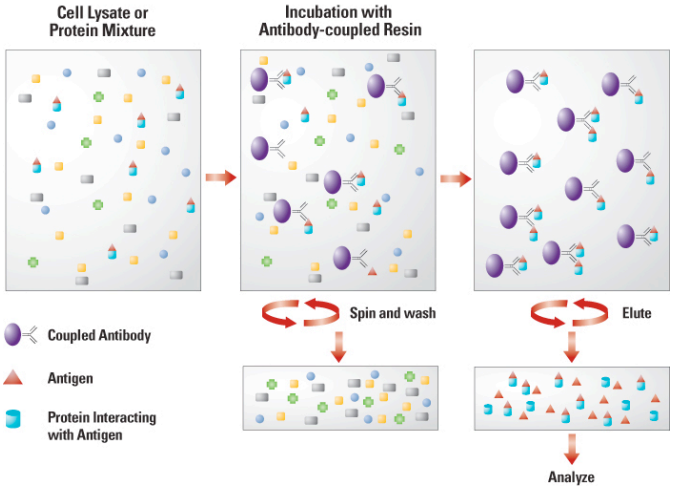
The Western blot is typically used to analyze proteins isolated by (co-)*immunoprecipitation*

immunoprecipitation (IP)



In an IP experiment, a protein is isolated from a cell lysate by capture with an antibody, which is then captured by a bead coated with *Protein A*. The isolated proteins are typically analyzed by Western blot

co-immunoprecipitation (co-IP)



A co-IP experiment is much like an IP experiment, except the researcher is interested in *the proteins that interact with the target protein*. It is these interacting proteins that would be probed in the Western blot.

- co-IP is a widely used technique for the discovery and validation of protein-protein interactions in real tissues.
- Note that Protein A is a protein from the bacteria *Staphylococcus aureus* that has a strong interaction with the Fc region of antibodies from many mammalian species. This protein is widely used for affinity purification of antibodies and basically any technique that involves capture of a wide variety of antibodies. Since it binds to the Fc region, it is indifferent to the particular specificity of the antibody.

An important variation on co-IP is chromatin IP (ChIP)

ChIP is used to study protein-DNA interactions. For example, it could be used to determine which sites in the genome a particular transcription factor was binding to.

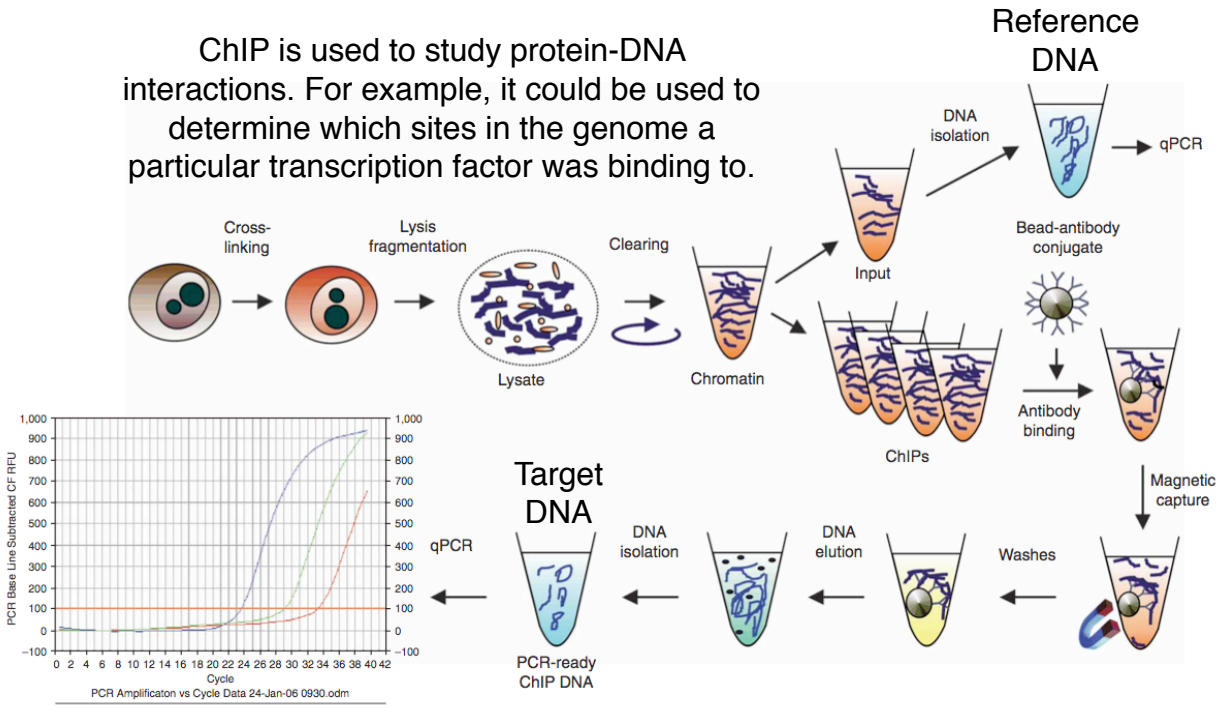


Figure 1 | Outline of the micro chromatin immunoprecipitation (μChIP) assay. qPCR, quantitative PCR.

- The first step is to cross-link the DNA to the proteins that are attached to them. This is typically done by treating with formaldehyde. What sort of linkages would you expect to get?
- The genomic DNA is then fragmented into small pieces using the non-specific approach of sonication. Fragments are typically 400-500 bp in size.
- A bead with antibodies specific for the target protein of interest is used to capture the protein-DNA complexes of interest. These are isolated from the rest of the proteins and DNA by several washing steps.
- The DNA is isolated by reversing the crosslinks. This can be done by heating in the presence of a high concentration of an amine. For example, Tris buffer has a free amine and it is widely used in these sorts of experiments.
- Finally, the DNA is analyzed by RT-PCR (also known as qPCR) and the relative concentrations of specific sequences to control sequences is compared for the target and reference DNA samples.
- If a specific sequence is enriched in the target DNA, it is likely a binding site for the protein that the antibody was specific for.
- Note that you do need to know 'what you are looking for' in a case like this, since you would need to add PCR primers that are specific for a particular region of DNA.
- An alternative approach would be sequence all of the target DNA. In this approach you could find previously unknown binding sites.

- Figure from: J. A. Dahyl, P. Collas "A rapid micro chromatin immunoprecipitation assay (lChIP)", *Nature Protocols*, 3, 1032-1045 (2008)

Gel Electrophoresis

- Gel electrophoresis separates DNA (and RNA) molecules according to their size.
- The mobility of either a double- or single-stranded DNA molecule during gel electrophoresis is inversely proportional to the logarithm of its length in nucleotides.
- The size of a DNA molecule of unknown length can be determined by comparison to the electrophoretic migration of molecules of known length.
- DNA molecules from 1 to 2000 nucleotides long are usually separated by electrophoresis in polyacrylamide gels; molecules from 500 nucleotides to 20 kb, by electrophoresis in agarose gels; and molecules from 20 to 10,000 kb, in pulsed-field agarose gels.
- Gel electrophoresis separates proteins based on their rate of movement in an electric field.
- SDS polyacrylamide gel electrophoresis (SDS-PAGE) can resolve polypeptide chains differing in molecular weight by 10 percent or less.
- Isoelectric focusing (IEF) separates proteins based on their pI
- SDS-PAGE and IEF can be combined in 2-dimensional electrophoresis. This technique can separate thousands of proteins into well-resolved spots.
- There is enough protein in one spot of a gel to analyze by Edman degradation or by MALDI mass spectrometry. These techniques can be used to unambiguously identify a protein.
- Even if only a small region of the protein is known from Edman degradation, degenerate DNA probes can be designed and used to isolate the complete gene from a library of genomic DNA.
- In a Western blot, proteins separated on a gel are transferred to a membrane. The membrane is then probed with an antibody specific for the protein of interest. This antibody can be radioactive or linked to an enzyme.
- Antibodies are powerful reagents used to detect, quantify, and isolate proteins.