8

Toolbox

"In theory, there is no difference between theory and practice. But in practice, there is."

Yogi Berra

"The workman who would perfect his craft must first sharpen his tools"- Confucius.

The pace of discovery in biochemistry is astounding. It is hard to believe that the first demonstration that an enzyme was a protein was made only in 1926. The speed with which the field has grown is in no small part due to the development of the tools and techniques with which to study life at the molecular level.

Basic Techniques

Introduction

The environment of a cell is very complex, making it difficult to study individual reactions, enzymes, or pathways in situ. The traditional approach used by biochemists for the study of these things is to isolate molecules, enzymes, DNAs, RNAs, and other items of interest so they can be analyzed independently of the millions of other processes occurring simultaneously. Today, these approaches are used side by side with newer methods that allow us to understand events inside cells on a larger scale- for example, determining all the genes that are being expressed at a given time in specific cells. In this section we take a brief look at some commonly used methods used to study biological molecules and their interactions.

YouTube Lectures by Kevin HERE & HERE

Breaking cells open

To separate compounds from cellular environments, one must first break open (lyse) the cells. Cells are broken open, in buffered solu-

tions, to obtain a *lysate*. There are several ways of accomplishing this.

Osmotic shock and enzymes

One way to lyse cells is by lowering the ionic strength of the medium the cells are in. This can cause cells to swell and burst. Mild surfactants may be used to disrupt membranes. Most bacteria, yeast, and plant tissues are resistant to osmotic shocks, because of the presence of cell walls, and stronger disruption techniques are usually required. Enzymes may be useful in helping to degrade the cell walls. Lysozyme, for example, is very useful for breaking down bacterial walls. Other enzymes commonly employed include cellulase (plants), proteases, mannases, and others.

Figure 8.1 - A sonicator in use

Wikipedia

Mechanical disruption

Mechanical agitation may be employed in the form of beads that are shaken with a mixture of cells. In this method, cells are bombarded with tiny, glass beads that break the cells open. Sonication (20-50 kHz sound waves) provides an alternative type of agitation that can be effective. The method is noisy, however, and generates heat that can be problematic for heat-sensitive compounds.

Pressure disruption

Another means of disrupting cells involves using a “cell bomb”. In this method, cells are placed under very high pressure (up to 25,000 psi) and then the pressure is rapidly released. The rapid pressure change causes dissolved gases in cells to be released as bubbles which, in turn, break open cells.

Cryopulverization

Cryopulverization is often employed

for samples having a tough extracellular matrix, such as connective tissue, seed, and cartilage. In this technique, tissues are frozen using liquid nitrogen and then impact pulverization (typically, grinding, using a mortar and pestle or a powerful electric grinder) is performed. The powder so obtained is then suspended in the appropriate buffer.

Figure 8.2 - A high-speed centrifuge can be used to obtain different cell fractions from a crude lysate

Wikipedia

Whatever method is employed to create a lysate, crude fractions obtained from it must be further processed *via* fractionation.

Fractionation

Fractionation of samples, as the name suggests, is a process of separating out the components or fractions of the lysate. Fractionation typically begins with centrifugation of the lysate. Using low-speed centrifugation, one can remove cell debris, leaving a supernatant containing the contents of the cell. By using successively higher centrifugation speeds (and resulting g forces) it is possible to separate out different cellular components, like nuclei, mitochondria, etc., from the cytoplasm. These may then be separately lysed to release molecules that are specific to the particular cellular compartment. The soluble fraction of any lysate can, then, be further separated into its constituents using various methods.

Column chromatography

One powerful method used for this purpose is chromatography. We will consider several chromatographic approaches. Chromatography is used to separate out the components of a mixture based on differences in their size, charge or other characteristics. During chromatography, the mobile phase (buffer or other solvent) moves through the stationary phase (usually a solid matrix) carrying the components of the mixture. Separation of the components is achieved, because the different components move at different rates, for reasons that vary, depending on the type of chromatography used. We will consider several different kinds of chromatography to illustrate this process.

1. Ion exchange chromatography
2. Gel exclusion chromatography
3. Affinity chromatography
4. HPLC

These variations on chromatography are performed with the stationary phase held within so-called columns (Figure 8.3). These are

Figure 8.3 -- An ion exchange column apparatus

Wikipedia

tubes containing the stationary phase (also called the “support” or solid phase).

Supports are composed of tiny beads suspended in buffer (Figure 8.4) and are designed to exploit the chemistry or size differences of the components of the samples and thus provide a means of separation. Columns are “packed” or filled with the support, and a buffer or solvent carries the mixture of compounds to be separated through the support. Molecules in the sample interact differentially with the support and, consequently, travel through it at different speeds, thus enabling separation.

Figure 8.4 - Ion exchange beads

Wikipedia

Ion exchange chromatography

In ion exchange chromatography, the support consists of tiny beads to which are attached chemicals possessing a charge. Before use, the beads are equilibrated in a solution containing an appropriate counter-ion to the charged molecule on the bead. Figure 8.5 shows the repeating unit of polystyrolsulfonate, a compound used as a cation exchange resin. As you can see, this molecule is negatively charged, and thus the beads would be equilibrated in a buffer containing a positively charged ion, say sodium. In the suspension, the negatively charged polystyrolsulfonate is unable to leave the beads, due to its covalent attachment, but the counter-ions (sodium) can be “exchanged” for molecules of the same charge.

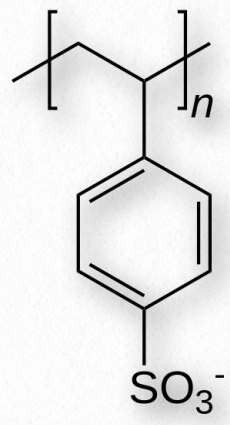


Figure 8.5 - Polystyrolsulfonate, a cation exchange resin

Wikipedia

Exchanges

Thus, a cation exchange column will have positively charged counter-ions and nega-

tively charged molecules covalently attached to the beads. Positively charged compounds from a cell lysate passed through the column will exchange with the counter-ions and "stick" to the negatively charged compounds covalently attached to the beads. Molecules in the sample that are neutral in charge or negatively charged will pass quickly through the column. At this point, only positively charged molecules from the original sample would be bound to the column. These may then be washed off, or eluted, by using buffers containing high concentrations of salt. Under these conditions, the interaction between the positively charged molecules and the polystyrosulfonate would be disrupted, allowing the molecules that were bound to the column to be recovered.

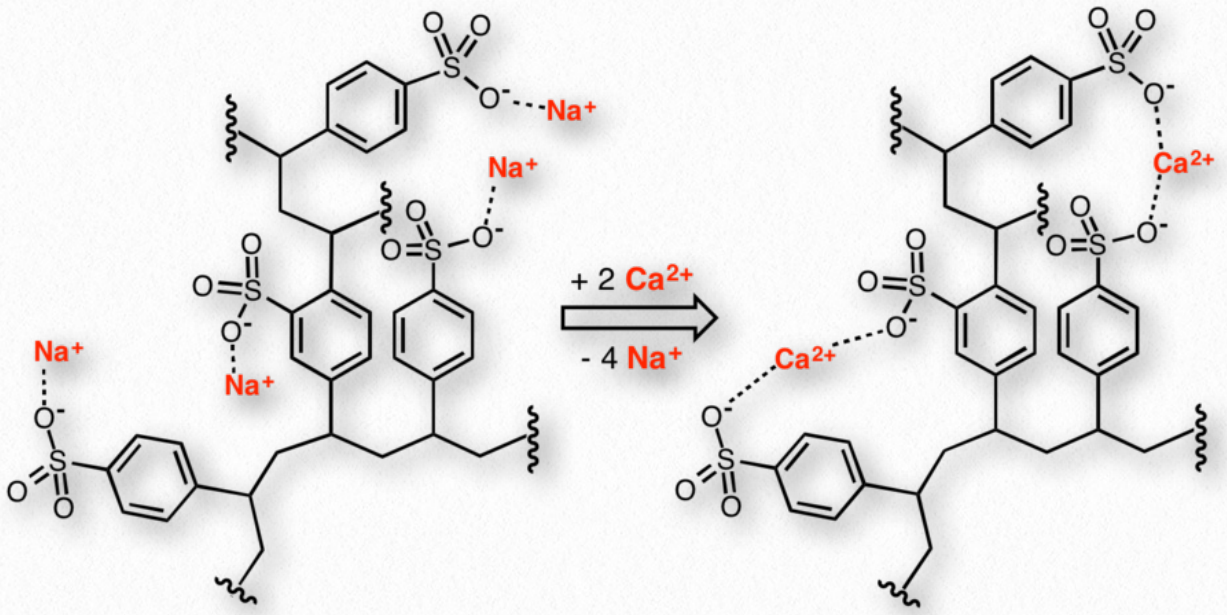


Figure 8.6 - Removal of calcium ions by an ion exchanger
Wikipedia

YouTube Lectures by Kevin HERE & HERE

Anion exchange

On the other hand, in anion exchange chromatography, the chemicals attached to the beads are positively charged and the counterions are negatively charged (chloride, for example). Negatively charged molecules in the cell lysate will "stick" and other molecules will pass through quickly. To remove the molecules "stuck" to a column, one simply needs to add a high concentration of counter-ions to release them.

Uses

Ion exchange resins are useful for separating charged from uncharged, or oppositely charged, biomolecules in solution. The resins have a variety of other applications, including water purification and softening. Figure 8.6 shows use of a polystyrolsulfonate polymer in removing calcium for water softening.

Interactive Learning Module HERE

Size exclusion chromatography

Size exclusion chromatography (also called molecular exclusion chromatography, gel exclusion chromatography, or gel filtration chromatography) is a low resolution separa-

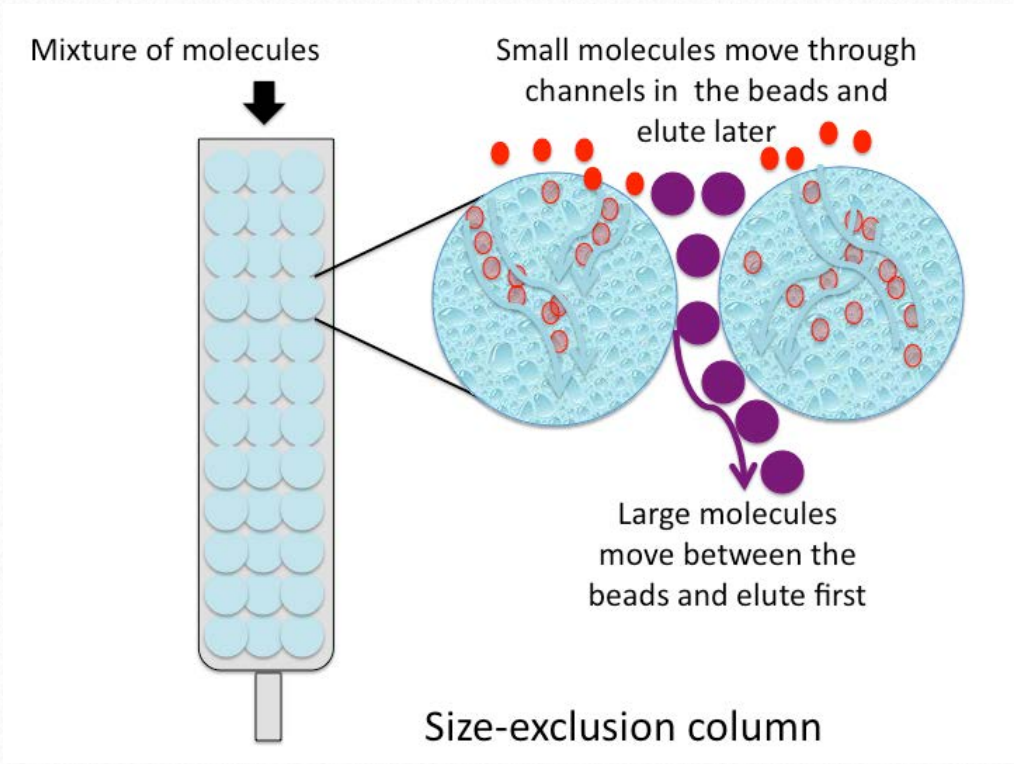


Figure 8.7 - Separation of molecules by size in a size-exclusion (aka gel filtration) column

tion method that employs beads with tiny "tunnels" in them that each have a precise opening. The size of the opening is referred to as an "exclusion limit," which means that molecules above a certain molecular weight will not be able to pass through the tunnels. Molecules with physical sizes larger than the exclusion limit do not enter the tunnels and pass through the column relatively quickly, in the spaces outside the beads. Smaller molecules, which can enter the tunnels, do so, and thus, have a longer path that they take in passing through the column and elute last (Figure 8.7).

Figure 8.8 shows a profile of a group of proteins separated by size exclusion chromatography using beads with an exclusion limit of about 30,000 Daltons. Proteins 30,000 in molecular weight or larger elute in the void volume (left) while smaller proteins elute later (middle and right).

Affinity chromatography

Affinity chromatography is a very powerful and selective technique that ex-

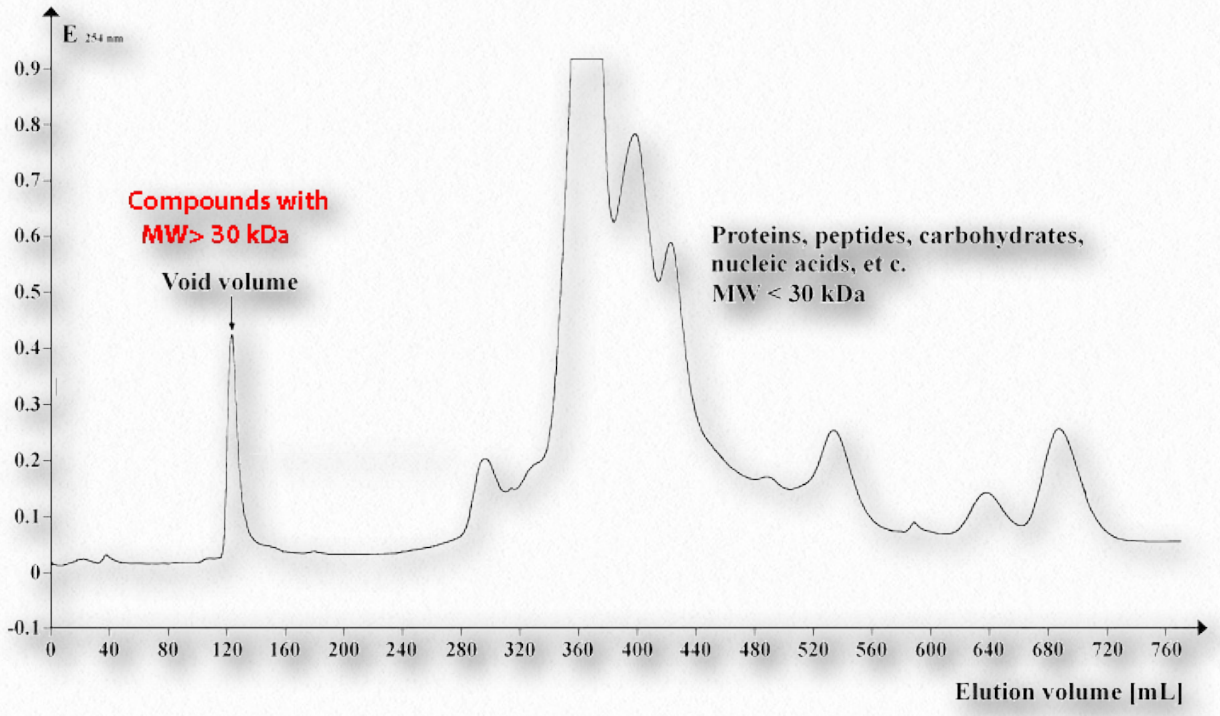


Figure 8.8 - Result of a size exclusion separation

Wikipedia

ploits the binding affinities of sample molecules (typically proteins) for molecules covalently linked to the support beads. In contrast to ion-exchange chromatography, where all molecules of a given charge would bind to the column, affinity chromatography exploits the specific binding of a protein or proteins to a ligand that is immobilized on the beads in the column.

Interactive Learning Module HERE

For example, if one wanted to separate all of the proteins in a cell lysate that bind to ATP from proteins that do not bind ATP, one could use a column that has ATP attached to the support beads and pass the sample through the column. All proteins that bind ATP will "stick" to the column, whereas those that do not bind ATP will pass quickly through it. The bound proteins may then be released from the column by adding a solution of ATP that will displace the bound proteins by competing, for the proteins, with the ATP attached to the column matrix.

Histidine tagging

Histidine tagging (His-tagging) is a special kind of affinity chromatography and is a

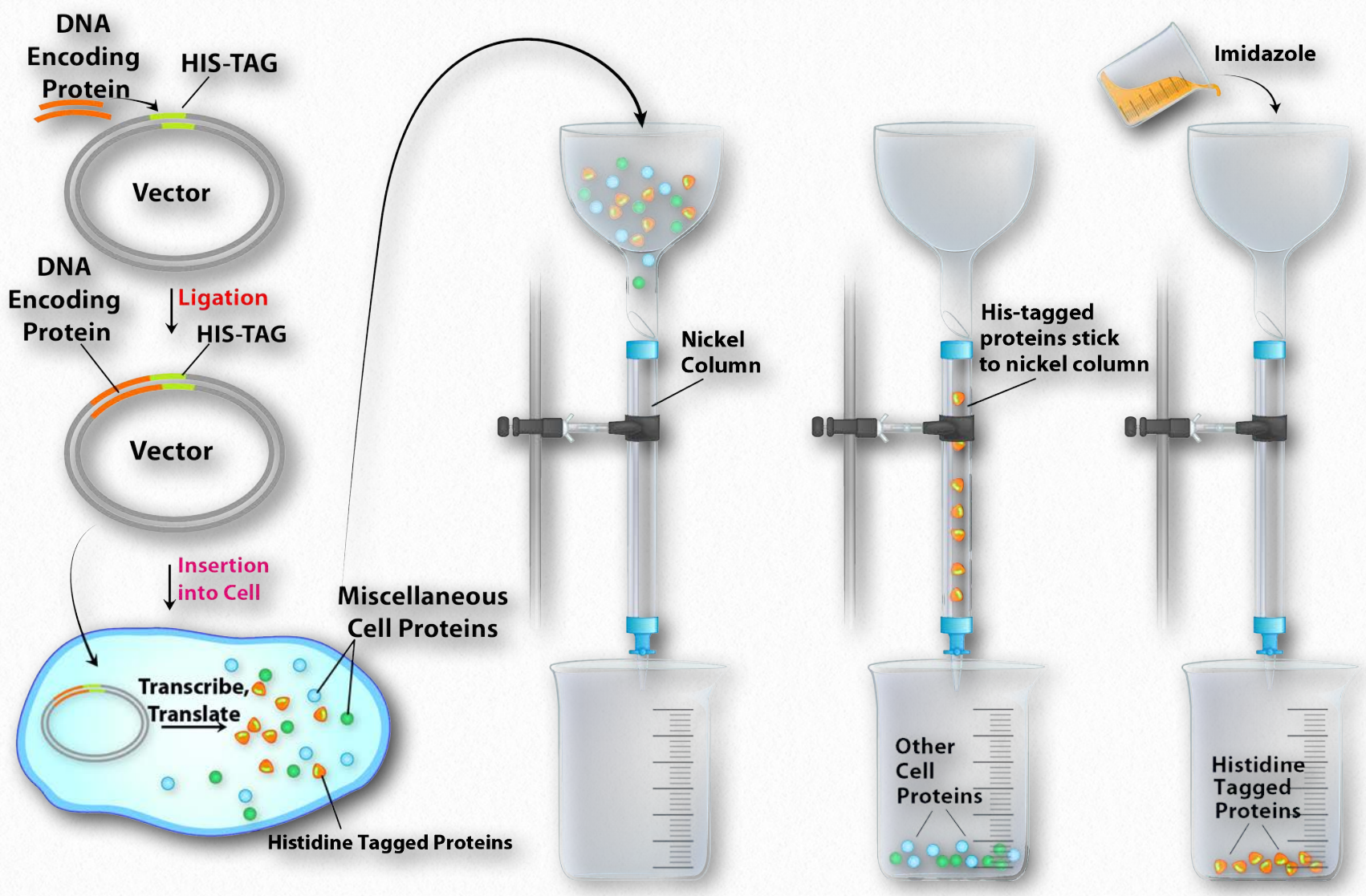


Figure 8.9 - Affinity chromatographic purification of a protein by histidine tagging

Image by Aleia Kim

powerful tool for isolating a recombinant protein from a cell lysate. His-tagging relies on altering the DNA coding region for a protein to add a series of at least six histidine residues to the amino or carboxyl terminal of the encoded protein. This "His-Tag" is useful in purifying the tagged protein because histidine side chains can bind to nickel or cobalt ions. Separation of His-tagged proteins from a cell lysate is relatively easy (Figure 8.9).

Figure 8.10 - HPLC: Pumps on left/Column in center/Detector on the right

Wikipedia

Passing the crude cell lysate through a column with nickel or cobalt attached to beads allows the His-tagged proteins to "stick," while the remaining cell proteins all pass quickly through. The His-tagged proteins are then eluted by addition of imidazole to the column. Imidazole, which resembles the side chain of histidine, competes with the His-tagged proteins and displaces them from the column. Although non-tagged proteins in the lysate may also contain histidine as part of their sequence, they will not bind to the column as strongly as the His-tagged protein and will, thus, be displaced at lower imidazole concentrations than needed to elute the His-tagged protein. Surprisingly, many His-tagged proteins appear to function normally despite the added histidines, but if needed, the histidine tags may be cleaved from the purified protein by treatment with a protease that excises the added histidines, allowing the recovery of the desired protein with its native sequence.

HPLC

High performance liquid chromatography (HPLC) is a powerful tool for separating a variety of molecules based on their differential polarities (Figure 8.10). A more efficient form of column chromatography, it employs columns with tightly packed supports and very tiny beads such that flow of solvents/buffers through the columns requires high pressures. The supports used may be polar (normal phase separation) or non-polar (reverse phase separation). In normal phase separations, non-polar molecules elute first followed by the more polar compounds. This order is switched in reverse phase chromatography. Of the two, reverse phase is much more commonly em-

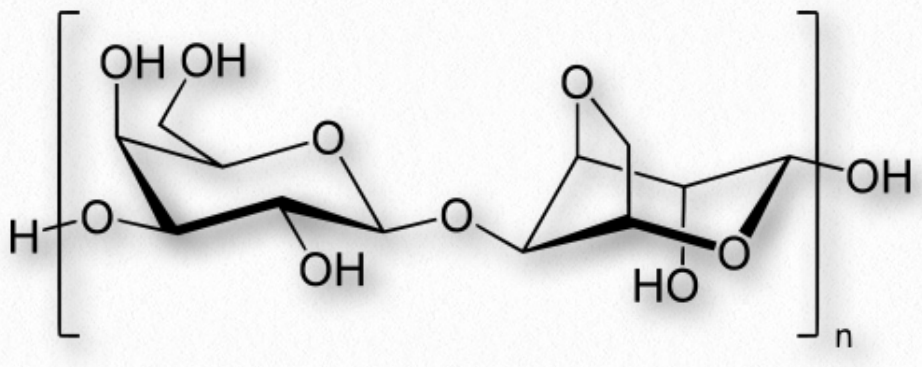


Figure 8.11 - Structure of the agarose polysaccharide

Wikipedia

ployed due to more reproducible chromatographic profiles (separations) that it typically produces.

Gel electrophoresis

Electrophoresis uses an electric field applied across a gel matrix to separate large molecules such as DNA, RNA, and proteins by charge and size. Samples are loaded into the wells of a gel matrix that can separate molecules by size and an electrical field is applied across the gel. This field causes negatively charged molecules to move towards the positive electrode. The gel matrix, itself, acts as a sieve, through which the smallest molecules pass rapidly, while longer molecules are slower-moving. For DNA and RNA, sorting molecules by size in this way is trivial, because of the uniform negative charge on the phosphate backbone. For proteins, which vary in their charges, a clever trick must be employed to make them mimic nucleic acids - see polyacrylamide gel electrophoresis (PAGE) below. Different kinds of gels have different pore sizes. Like sieves with finer or coarser meshes, some gels do a better job of separating smaller molecules while others work better for larger ones.

Gel electrophoresis may be used as a preparative technique (that is, when purifying proteins or nucleic acids), but most often it is used as an analytical tool.

Agarose gel electrophoresis

Agarose gel electrophoresis is a technique used to separate nucleic acids primarily by size. Agarose is a polysaccharide obtained

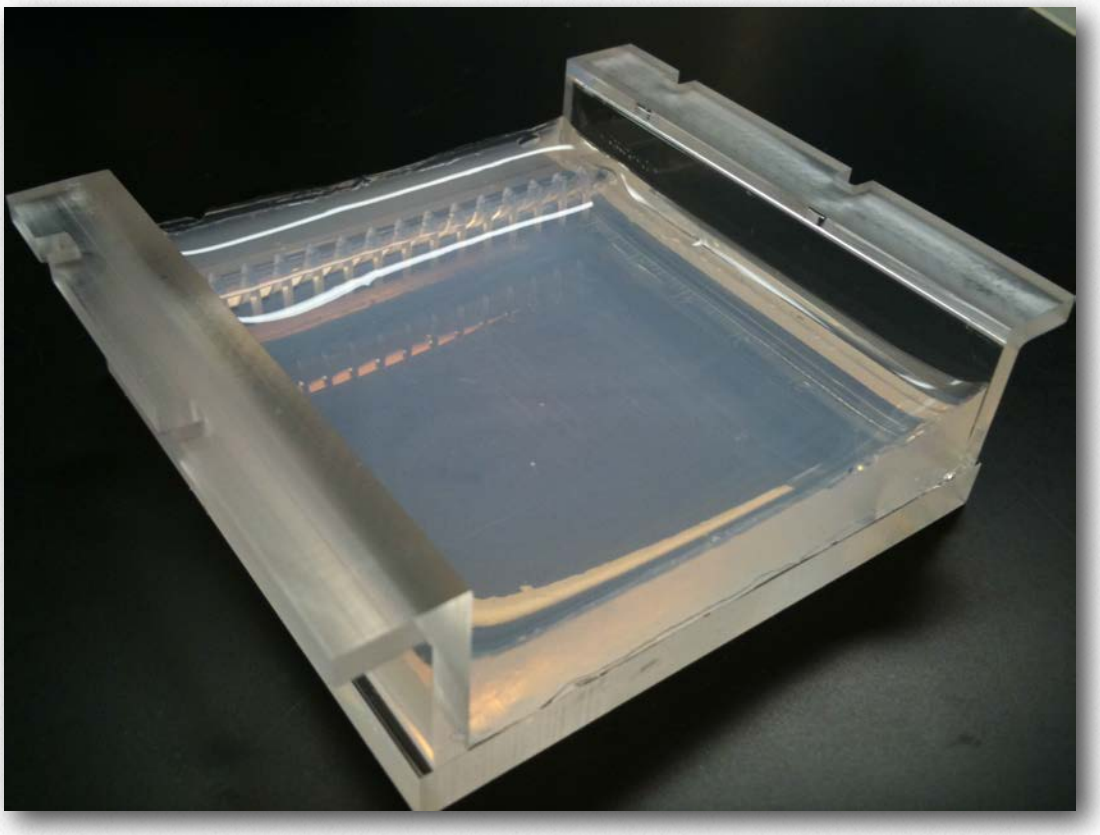


Figure 8.12 - Agarose gel electrophoresis separation of DNA - orange bands are DNA fragments

Wikipedia

from seaweeds (Figure 8.11). It can be dissolved in boiling buffer and poured into a tray, where it sets up as it cools (Figure 8.12) to form a slab. Agarose gels are poured with a comb in place to make wells into which DNA or RNA samples are placed after the gel has solidified. The gel is immersed in a buffer and a current is applied across the slab. Double-stranded DNA has a uniform negative charge that is independent of the sequence composition of the molecule. Therefore, if DNA fragments are placed in an electric field they will migrate from the cathode (-) towards the anode (+). The rate of migration is directly dependent on the ability of each DNA molecule to worm or wiggle its way through the sieving gel. The agarose matrix provides openings for macromolecules to move through. The largest macromolecules have the most difficult time navigating through the gel, whereas the smallest macromolecules slip through it the fastest.

YouTube Lectures by Kevin HERE & HERE

Because electrophoresis uses an electric current as a force to drive the molecules through the matrix, the molecules being separated must be charged. Since the size to charge ratio for DNA and RNA is constant for all sizes of these nucleic acids, the molecules simply sort on the basis of their size - the smallest move fastest and the largest move slowest.

Interactive Learning Module HERE

All fragments of a given size will migrate the same distance on the gel, forming the so-called “bands” on the gel. Visualization of the DNA fragments in the gel is made possible by addition of a dye, such as ethidium bromide, which intercalates between the bases and fluoresces when viewed under ultraviolet light (Figure 8.13) By running reference DNAs of known sizes alongside the samples, it is possible to determine the sizes of the DNA fragments in the sample. It is useful to note that, by convention, DNA fragments are not described by their molecular

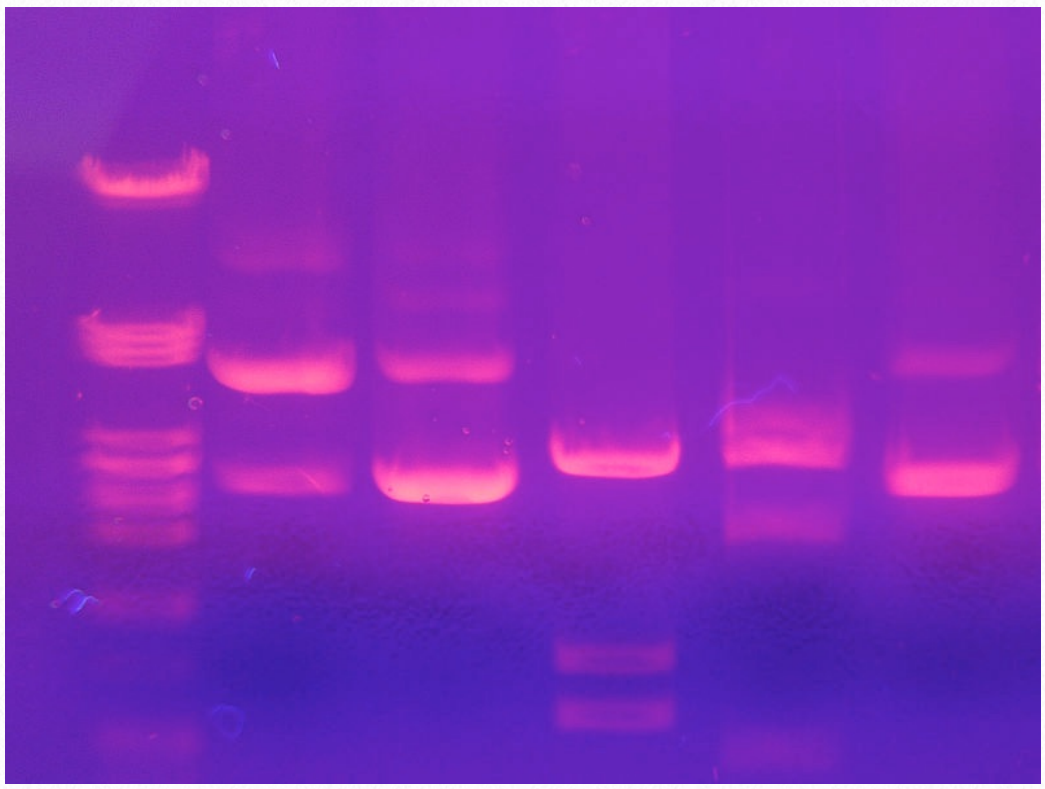


Figure 8.13 - DNA bands visualized with ethidium bromide staining

Wikipedia

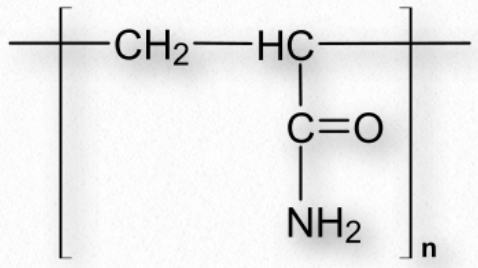


Figure 8.14 - Acrylamide monomer

Wikipedia

weights (unlike proteins), but by their length in base-pairs(bp) or kilobases (kb).

Polyacrylamide gel electrophoresis (PAGE)

Like DNA and RNA, proteins are large macromolecules, but unlike nucleic acids, proteins are not necessarily negatively charged. The charge on each protein depends on its unique amino acid sequence. Thus, the proteins in a mixture will not necessarily all move towards the anode.

Additionally, whereas double-stranded DNA is rod-shaped, most proteins are globular (folded). Further, proteins are considerably smaller than nucleic acids, so the openings of the matrix of the agarose gel are simply too large to effectively provide separation. Consequently, unaltered (native) proteins are not very good prospects for electrophoresis on agarose gels. To separate proteins by mass using electrophoresis, one must make several modifications.

Gel matrix

First, a matrix made by polymerizing and cross-linking acrylamide units is employed. A monomeric acrylamide (Figure 8.14) is polymerized and the polymers are cross-linked using N,N'-Methylene-bisacrylamide (Figure 8.15) to create a mesh-like structure. One can adjust the size of the openings of the matrix/mesh readily by changing the percentage of acrylamide in the reaction. Higher percentages of acrylamide give smaller openings and are more effective for separating smaller molecules, whereas lower percentages of acrylamide are used when resolving mixtures of larger molecules. (Note: polyacrylamide gels are also used to separate small nucleic acid fragments, with some acrylamide gels capable of separating pieces of DNA that differ in length by just one nucleotide.)

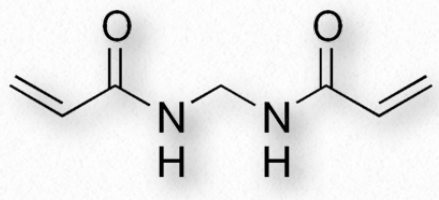


Figure 8.15 - N,N'-Methylenebisacrylamide - acrylamide crosslinking reagent

Wikipedia

Charge alteration by SDS

A second consideration is that proteins must be physically altered to "present" themselves to the matrix like the negatively charged rods of DNA. This is accomplished by treating the proteins with the anionic detergent, SDS (sodium dodecyl sulfate).

SDS denatures the proteins so they assume a rod-like shape and the SDS molecules coat the proteins such that the exterior surface is

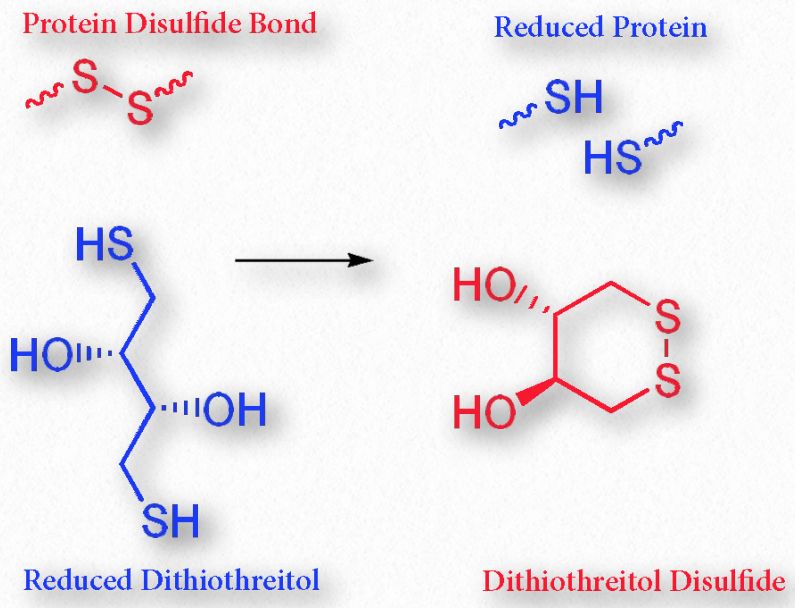


Figure 8.16 - Reduction of disulfide bonds by dithiothreitol

Wikipedia

loaded with negative charges, masking the original charges on the proteins and making the charge on the proteins more proportional to their mass, like the backbone of DNA.

Since proteins typically have disulfide bonds that prevent them from completely unfolding in detergent, samples are boiled with mercaptoethanol to break the disulfide bonds and ensure the proteins are as rod-like as possible in the SDS. Reagents like mercaptoethanol (and also dithiothreitol) are sulfhydryl-containing reagents that become oxidized as they reduce disulfide bonds in other molecules (see Figure 8.16)

Stacking gel

A third consideration is that a "stacking gel" may be employed at the top of a polyacrylamide gel to provide a way of compressing the samples into a tight band before they enter the main polyacrylamide gel (called the resolving gel).

Just like DNA fragments in agarose gel electrophoresis get sorted on the basis of size (largest move slowest and smallest move fastest), the proteins migrate through the gel matrix at velocities inversely related to their size. Upon completion of the electrophoresis, proteins may be visualized by staining with compounds that bind to proteins, like Coomassie Brilliant Blue (Figure 8.17) or silver nitrate.

Non-denaturing gel electrophoresis

The SDS_PAGE technique described above is the commonest method used for electrophoretic separation of proteins. In some situations, however, proteins may be resolved on so-called "native" gels, in the absence of SDS.

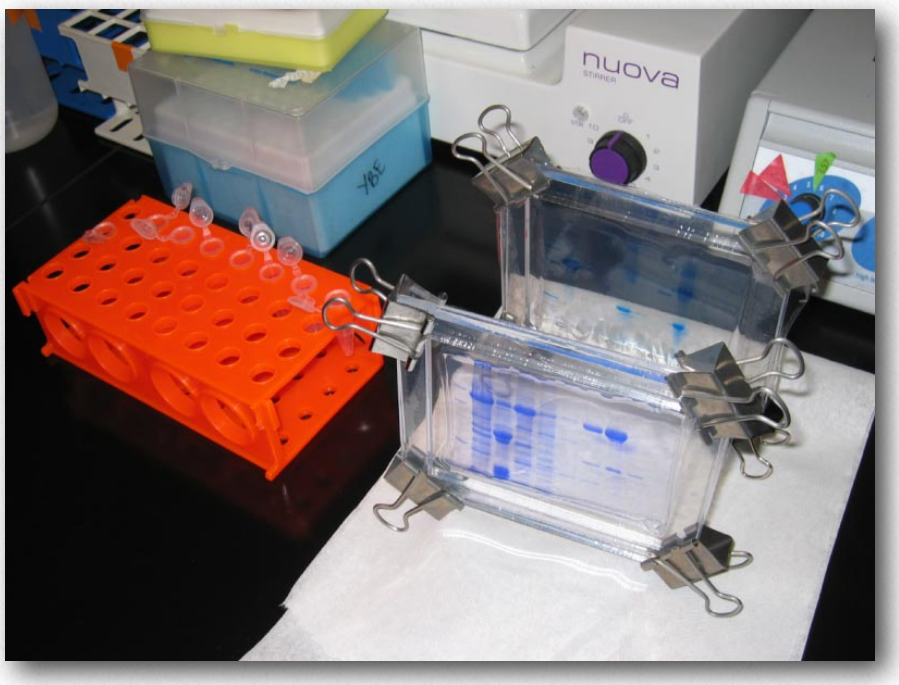


Figure 8.17 - Two SDS-PAGE gels - Proteins are the blue bands (stained with Coomassie Blue)

Wikipedia

Under these conditions, the movement of proteins through the gel will be affected not simply by their mass, but by their charge at the pH of the gel, as well. Proteins complexed with other molecules may move as single entity, allowing the isolation of the binding partners of proteins of interest.

Isoelectric focusing

Proteins vary considerably in their charges and, consequently, in their pI values (pH at which their charge is zero). This can be exploited to separate proteins in a mixture. Separating proteins by isoelectric focusing requires establishment of a pH gradient in a tube containing an acrylamide gel matrix. The pore size of the gel is adjusted to be large, to reduce the effect of sieving based on size. Molecules to be separated are applied to the gel containing the pH gradient and an electric field is applied.

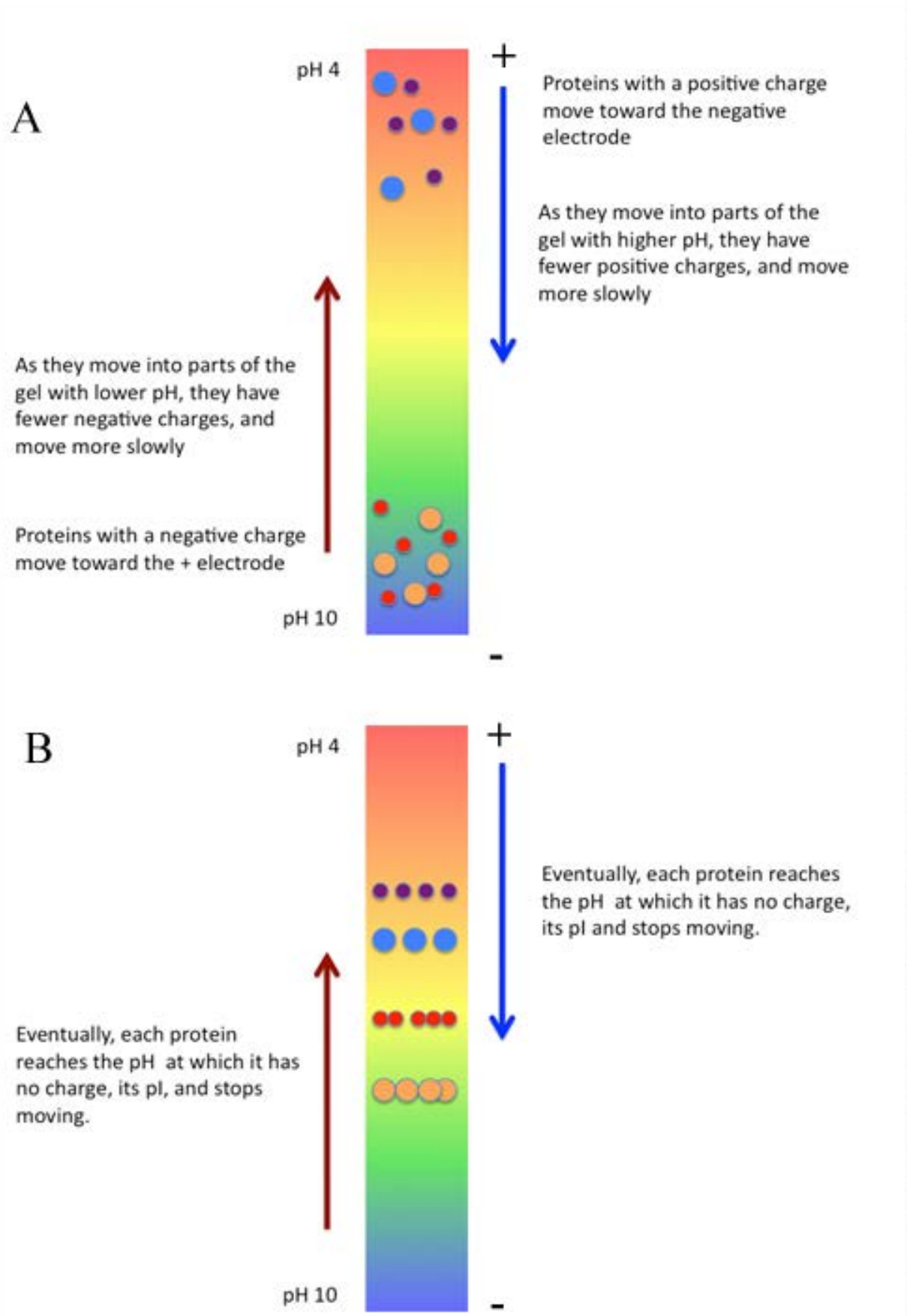


Figure 8.18 - Isoelectric focusing: A. At the start of the run; B. at the end of the run

Image by Indira Rajagopal

Under these conditions, proteins will move according to their charge.

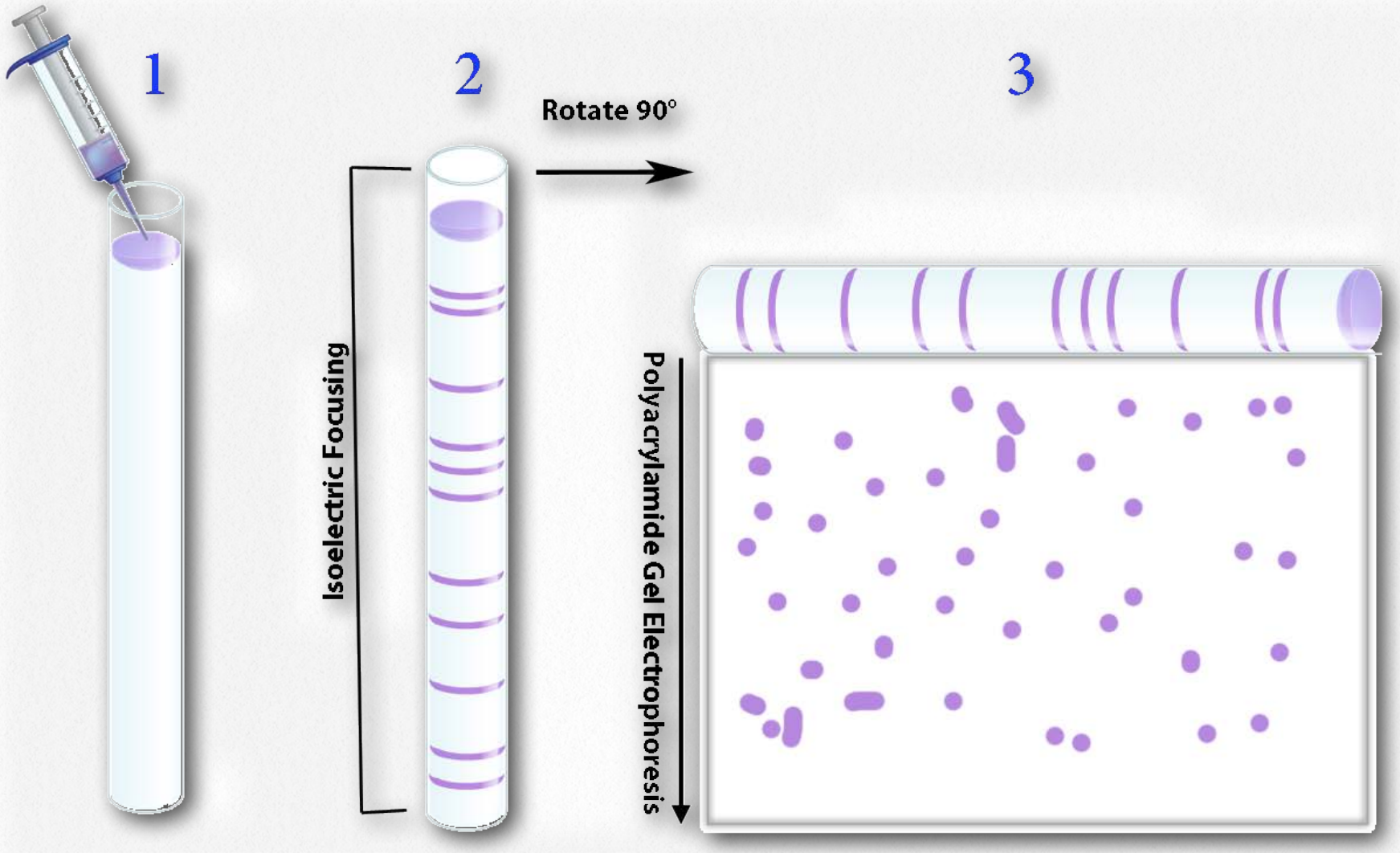


Figure 8.19 - Scheme for performing 2-D gel analysis

Image by Aleia Kim

Positively charged molecules, for example, move towards the negative electrode, but since they are traveling through a pH gradient, as they pass through it, they reach a region where their charge is zero and, at that point, they stop moving. They are at that point attracted to neither the positive nor the negative electrode and are thus "focused" at their pI (Figure 8.18). Using isoelectric focusing, it is possible to separate proteins whose pI values differ by as little as 0.01 units.

YouTube Lectures by Kevin HERE & HERE

2D gel electrophoresis

Both SDS-PAGE and isoelectric focusing are powerful techniques, but a clever combination of the two is a powerful tool of proteomics - the science of studying all of the proteins of a cell/tissue simultaneously. In 2-D gel electrophoresis, a lysate is first prepared from the cells of interest. The proteins in the lysate are separated first by their pI, through isoelectric focusing and then by size by SDS-PAGE.

The mixture of proteins is first applied to a

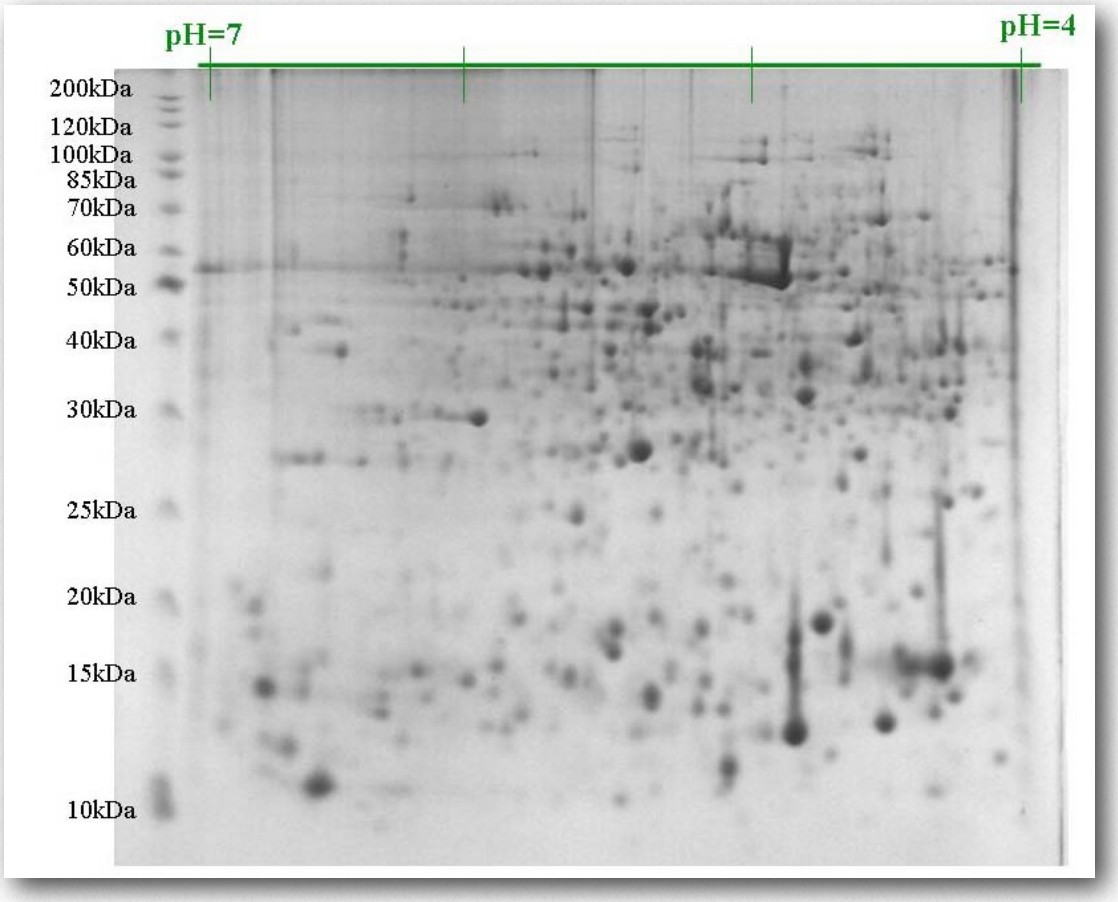


Figure 8.20 - Result of 2-D gel electrophoresis separation

Wikipedia

tube or strip (Figure 8.19, Step 1) where isoelectric focusing is performed to separate the proteins by their pI values (Step 2). Next, as shown in the figure, the gel containing the proteins separated by their pIs is turned on its side and applied along the top of a polyacrylamide slab for SDS-PAGE to separate on the basis of size (Step 3). The proteins in the isoelectric focusing matrix are electrophoresed into the polyacrylamide gel and separated on the basis of size. The product of this analysis is a 2-D gel as shown in Figure 8.20.

The power of 2-D gel electrophoresis is that virtually every protein in a cell can be separated and appear on the gel as a spot defined by its unique size and pI. In the figure, spots in the upper left correspond to large positively charged proteins, whereas those in the lower right are small negatively charged ones. Every spot on a 2-D gel can be eluted and identified by using high throughput mass spectrometry. This is particularly powerful when one compares protein profiles between different tissues or between control and treated samples of the same tissue.

Interactive Learning Module HERE

Protein profiles comparison

Comparison of 2-D gels of proteins from non-cancerous tissue and proteins from a cancerous tissue of the same type provides a quick identification of proteins whose level of expression differs between the two. Information such as this can be useful in designing treatments or in understanding the mechanism(s) by which the cancer develops.

Detection, identification and quantitation of specific nucleic acids and proteins

While gel electrophoresis can be used to resolve molecules in a mixture, by itself, the

technique does not permit the detection and identification of specific nucleic acid sequences or proteins. For example, the 2-D gel shown above clearly separates a large number of proteins in a sample into individual spots. However, if we wanted to know whether a specific protein was present, we could not tell by simply looking at the gel. Likewise, in an agarose gel, while bands of DNA could be assigned a size, one could not distinguish between two DNAs of different sequence if they were both the same length in base-pairs. One way to detect the presence of a particular nucleic acid or protein is dependent on transferring the separated molecules from the gels onto a membrane made of nitrocellulose or nylon to create a "blot" and probing for the molecule(s) of interest using reagents that specifically bind to those molecules. The next section will discuss how this can be done for nucleic acids as well as for proteins.

Southern and northern blots

The Southern blot is named for its inventor, Oxford professor, Edwin Southern, who came up with a protocol for transferring DNA fragments from a gel onto a nitrocellulose sheet and detecting a specific DNA sequence among the bands on the blot.

As shown in Figure 8.21, the method works as follows. A mixture of DNA molecules (often DNA that has been cut into smaller fragments using restriction endonucleases) is loaded on an agarose gel, as usual. After the gel run is complete, the DNA bands are transferred from the gel onto a membrane. This can be achieved by capillary transfer, where the gel is placed in contact with a piece of membrane and buffer is pulled through the gel by wicking it up into a stack of absorbent paper placed above the mem-

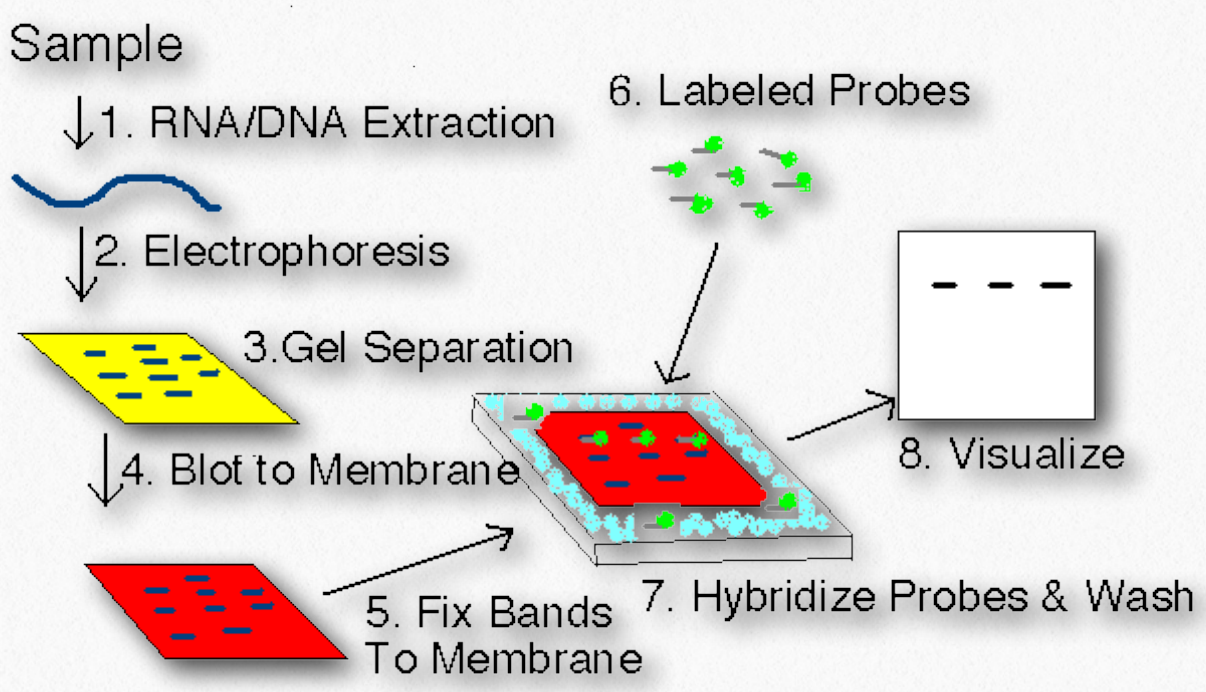


Figure 8.21 - Northern or Southern blot scheme. Southern blotting adds strand denaturation between steps 4 and 5.
Wikipedia

brane. As the buffer moves, it carries with it the DNA fragments. The DNA binds to the membrane leaving a “print” of DNA fragments that exactly mirrors their positions in the gel. The blotting membrane may be treated with UV light, heat, or chemicals to firmly attach the DNA to the membrane.

Next, a probe, or visualizing agent specific for the molecule of interest is added to the membrane. In Figure 8.21, this is called a labeled probe. The probes in a Southern blot are pieces of DNA designed to be complementary to the desired target sequence. If the sequence of interest is present on the blot, the probe, which is complementary to it, can base-pair (hybridize) with it. The blot is then washed to remove all unbound probe. Probes are labeled with radioactivity or with other chemical reagents that allow them to be easily detected when bound to the blot, so it is possible to visually determine whether the probe has bound to any of the DNA bands on the blot. Given that the Southern blot relies on specific base-pairing between the probe and the target sequence, it is easy to adapt the technique to detect specific RNA molecules, as well. The modification of this method to detect RNAs was jokingly named a “northern” blot.

Western blots

Proteins cannot, for obvious reasons, be detected through base-pairing with a DNA probe, but protein blots, made by transferring proteins, separated on a gel, onto a membrane, can be probed using specific antibodies against a particular protein of interest.

Protein detection usually employs two antibodies, the first of which is not labeled. The label is on the second antibody, which is designed to recognize only the first antibody in a piggyback fashion. The first antibody specifically binds to the protein of interest on the blot and the second antibody recognizes and binds the first antibody.

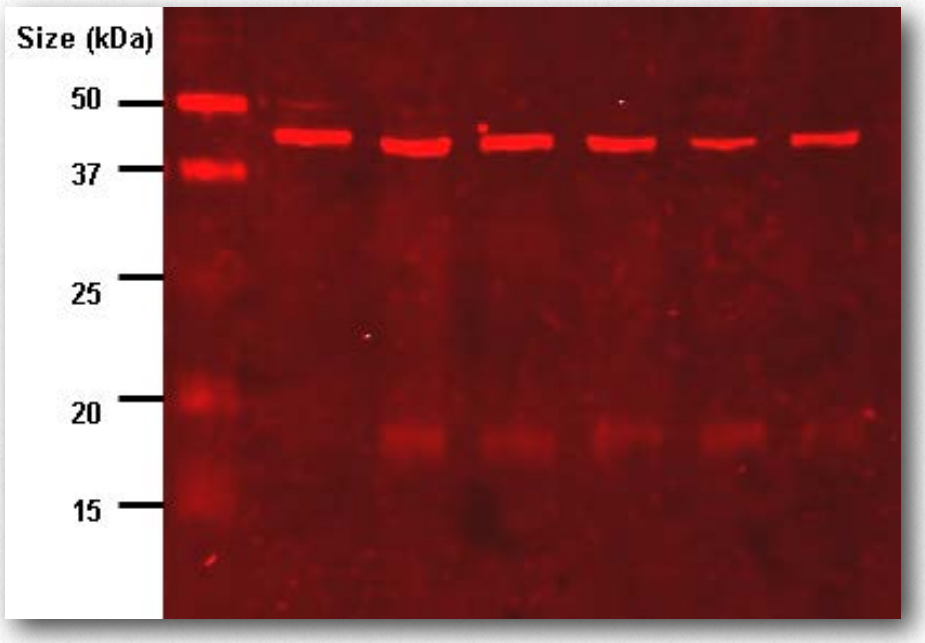


Figure 8.22 - Result of a western blot analysis

Wikipedia

The second antibody commonly carries an enzyme or reagent which can cause a reaction to produce a color upon further treatment. In the end, if the molecule of interest is in the original mixture, it will “light” up and reveal itself on the blot. This variation on the blotting theme was dubbed a western blot (Figure 8.22).

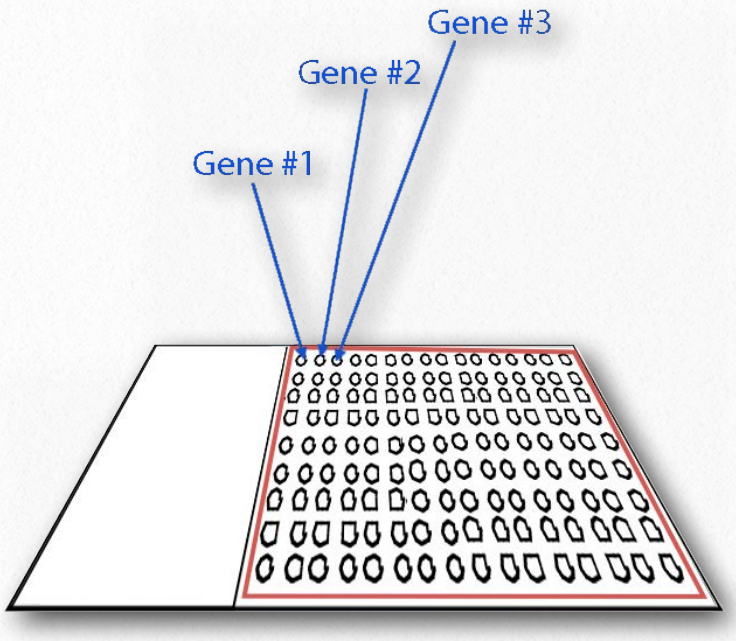


Figure 8.23 - Microarray design
Image by Taralyn Tan

In each of the blots described above, binding of the probe to the target molecule allows one to determine whether the target sequence or protein was in the sample. Although blots are designed to be used for detection, rather than for precise quantitation, it is possible to obtain estimates of the abundance of the target molecule from densitometry measurements of signal intensity.

Microarrays

2-D gels are a way of surveying a broad spectrum of protein molecules simultaneously. One approach to doing something similar with DNA or RNA involves what are called microarrays. Microarrays are especially useful for monitoring the expressions of thousands of genes, simultaneously. Where a northern blot would allow the identification of a single mRNA from a mixture of mRNAs, a microarray experiment can allow the simultaneous identification of thousands of mRNAs that may be made by a cell at a given time. It is also possible to perform quantitation much more reliably than with a blot.

Microarrays employ a glass slide, or chip, to which are attached short sequences of single-stranded DNA, arranged in a grid, or matrix (Figure 8.23) Each position in the grid corresponds to a unique gene. That is, the DNA sequence at this spot is part of the sequence of a specific gene. Each spot on the grid has multiple identical copies of the same sequence.

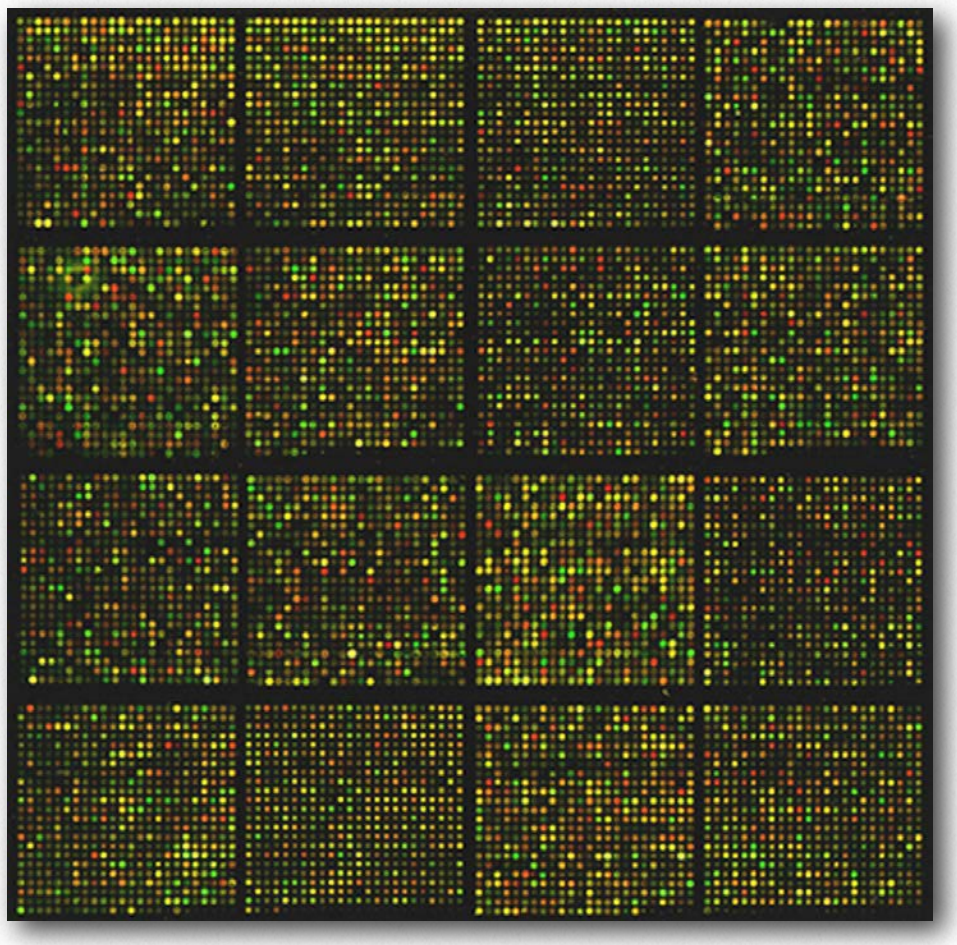


Figure 8.24 - Large scale microarray analysis of mouse transcriptome
Wikipedia

The gene sequence immobilized at each position in the grid is recorded.

Adding samples

To the slide are added a mixture of sample molecules, some of which will recognize and bind specifically to the sequences on the slide. Binding between the sample molecules and the sequences attached to the slide occurs by base pairing, in the case of DNA microarrays. The slide is then washed to remove sample molecules that are not specifically bound to the sequences in the grid.

Sample molecules are tagged with a fluorescent dye, allowing the spots where they bind to be identified. The grid is analyzed spot by spot for binding of the sample molecules to the immobilized sequences. The more sample molecules are bound at a spot, the greater the intensity of dye fluorescence that will be observed. Information from this analysis can give information about the presence/absence/abundance of molecules in the sample that bind to the sequences in the grid.

YouTube Lectures by Kevin
HERE & HERE

Transcriptomics

For example, consider a matrix containing all of the known gene sequences in a genome. To make such a matrix for analysis, one would need to make copies of every gene, either by chemical synthesis or by using the polymerase chain reaction. The strands of the resulting DNAs would then be separated to obtain single-stranded sequences that could be attached to the chip. Each box of the grid would contain sequence from one gene. With this grid, one could analyze the transcriptome - all of the mRNAs being made in selected cells at a given time. For a simple analysis, one could take a tissue (say liver) and extract all the mRNAs from it. This mRNA population represents all the genes that were being expressed in the liver cells at the time the RNA was extracted. These RNAs should be able to hybridize (base-pair) with their corresponding genes on the microarray. Genes that were not being expressed would have no mRNAs to bind to their corresponding genes on the grid.

In practice, the mRNAs are not used directly, but are copied into single-stranded DNA copies called cDNAs. The cDNAs are tagged with a fluorescent dye and added to the microarray under conditions that allow base pairing so that the cDNAs can find and base pair with complementary sequences on the matrix (Figure 8.26). The matrix is then washed to remove unhybridized cDNAs. The presence/absence/abundance of each mRNA is then readily determined by measuring the amount of dye at each box of the grid.

Interactive Learning Module
HERE

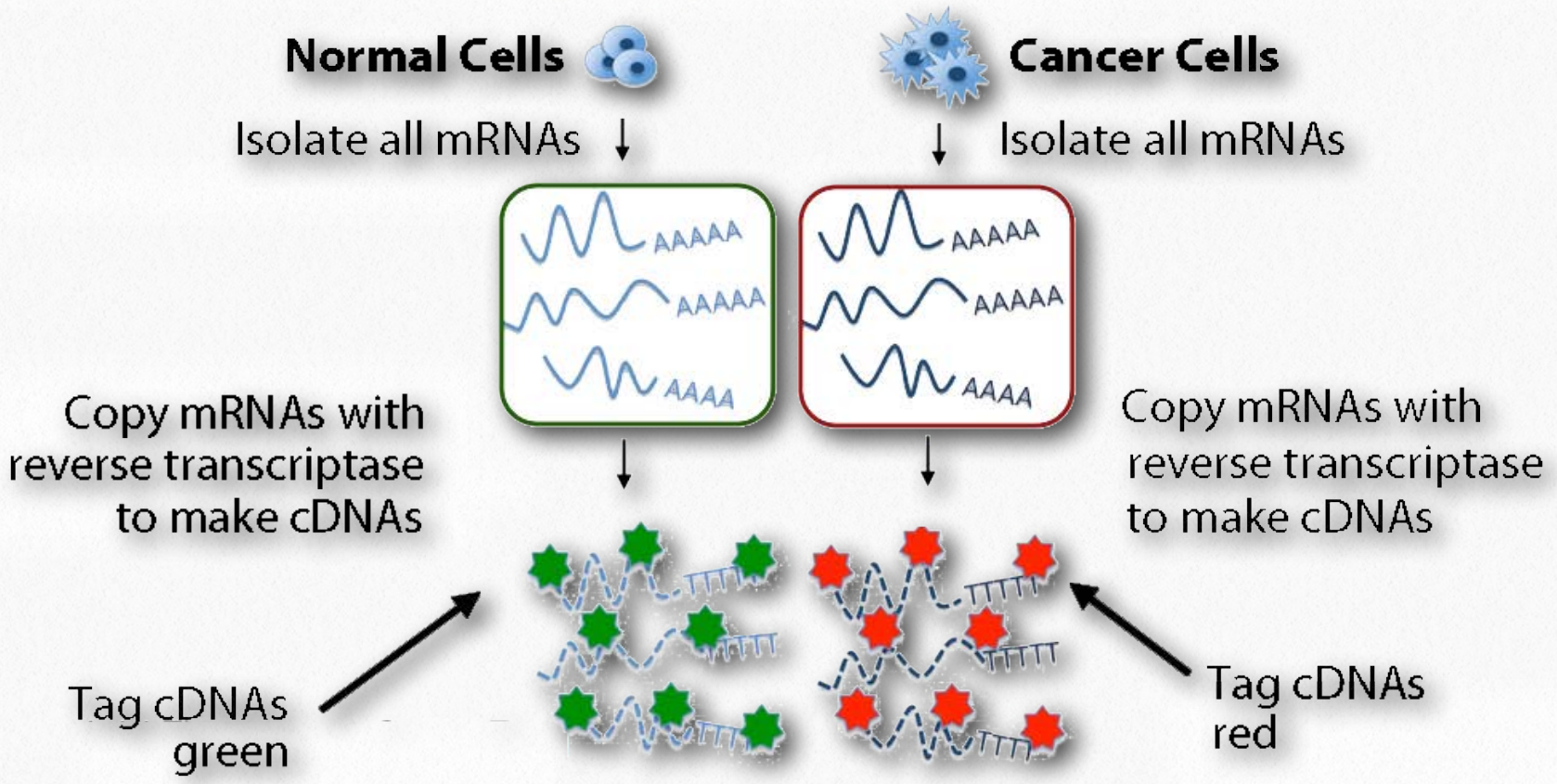


Figure 8.25 - Copying and labeling of transcriptome

Image by Taralyn Tan

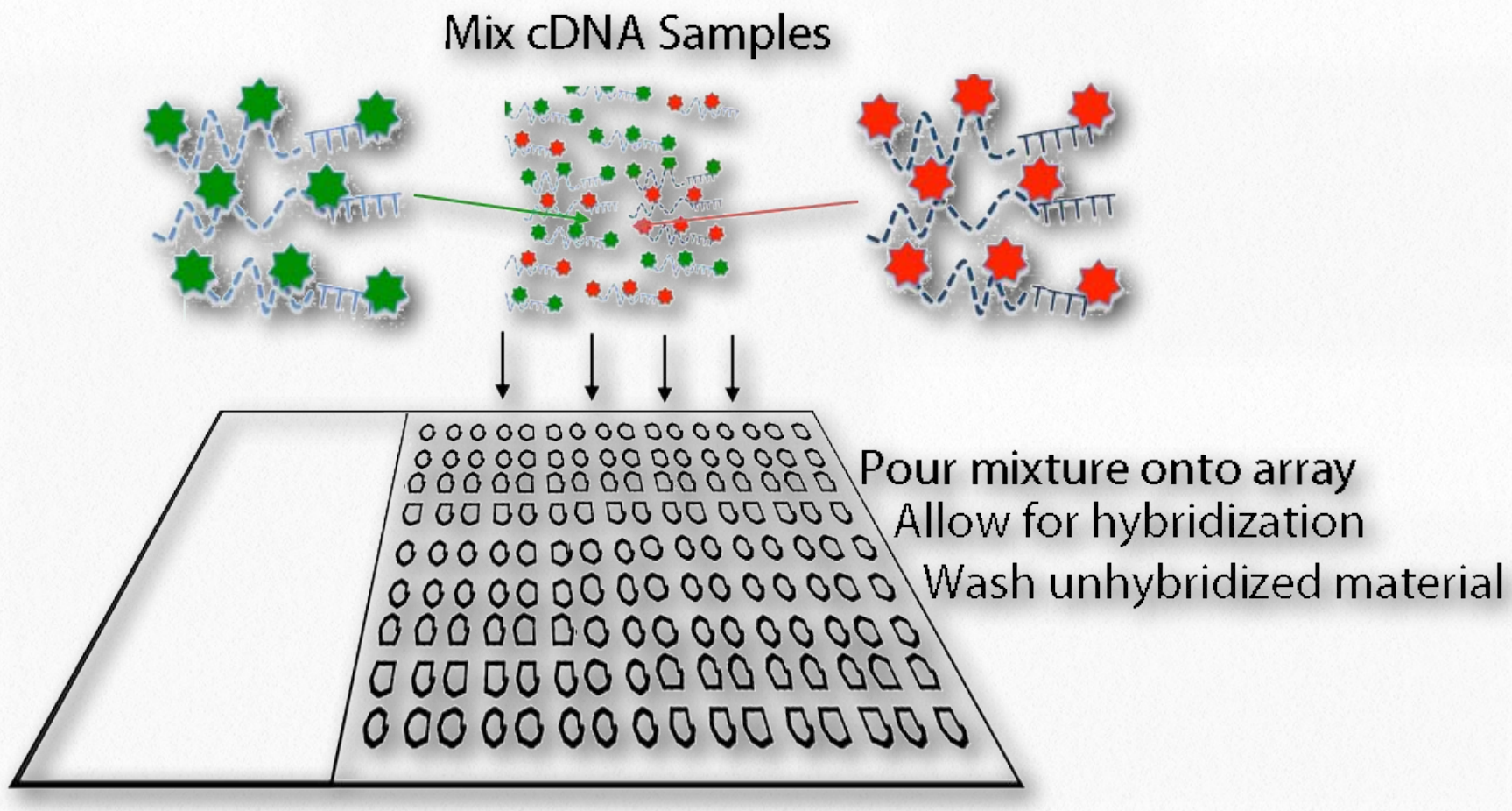


Figure 8.26 - Add labeled cDNAs to microarray plate

Image by Taralyn Tan

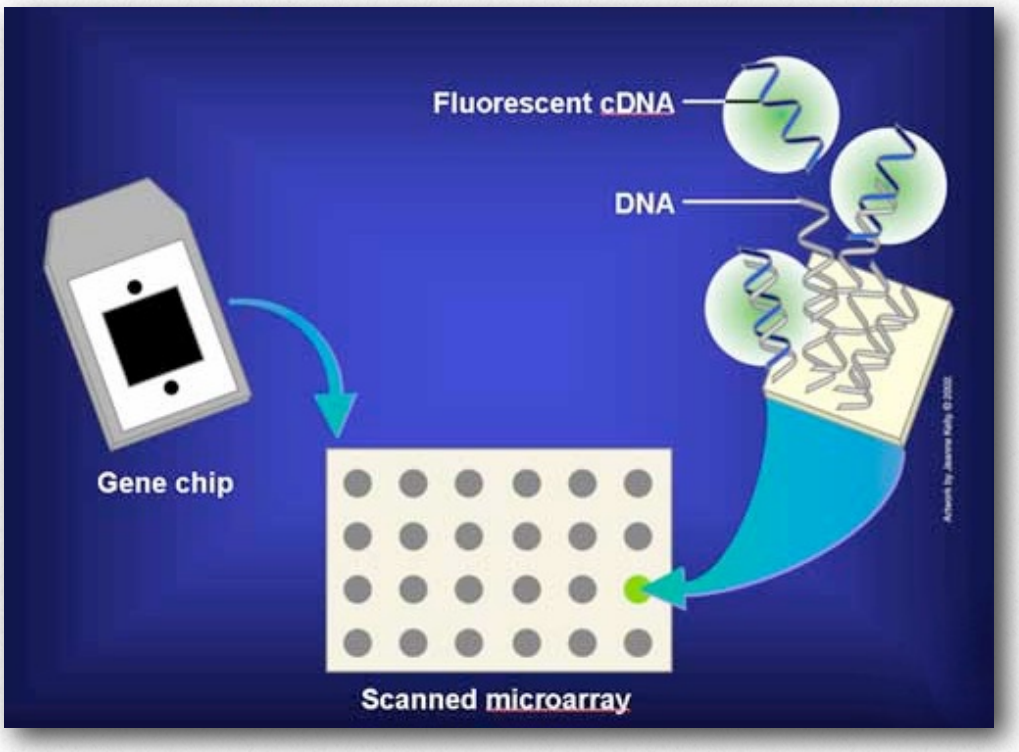


Figure 8.27 - Binding of a fluorescent cDNA copy of a specific mRNA to DNA immobilized on one spot in a microarray

Wikipedia

In Figure 8.27, a fluorescent cDNA has bound to the spot on the far right in the third row of the grid. This means that the sequence of the cDNA was complementary to the sequence of the gene sequence immobilized at that spot. Because the identity of the genes at each position on the grid is known, we then know that the sample contained mRNA that corresponded to that particular gene. In other words, that gene was being expressed in the cells from which the mRNAs were obtained.

A more powerful analysis could be performed with two sets of mRNAs simultaneously. . One set of cDNAs could come from a cancerous tissue and the other from a non-cancerous tissue, for example. The cDNAs derived from each sample is marked with a different color (say green for normal and red for cancerous) (Figure 8.25). The cDNAs are mixed and then added to the matrix and complementary sequences are once again allowed to form

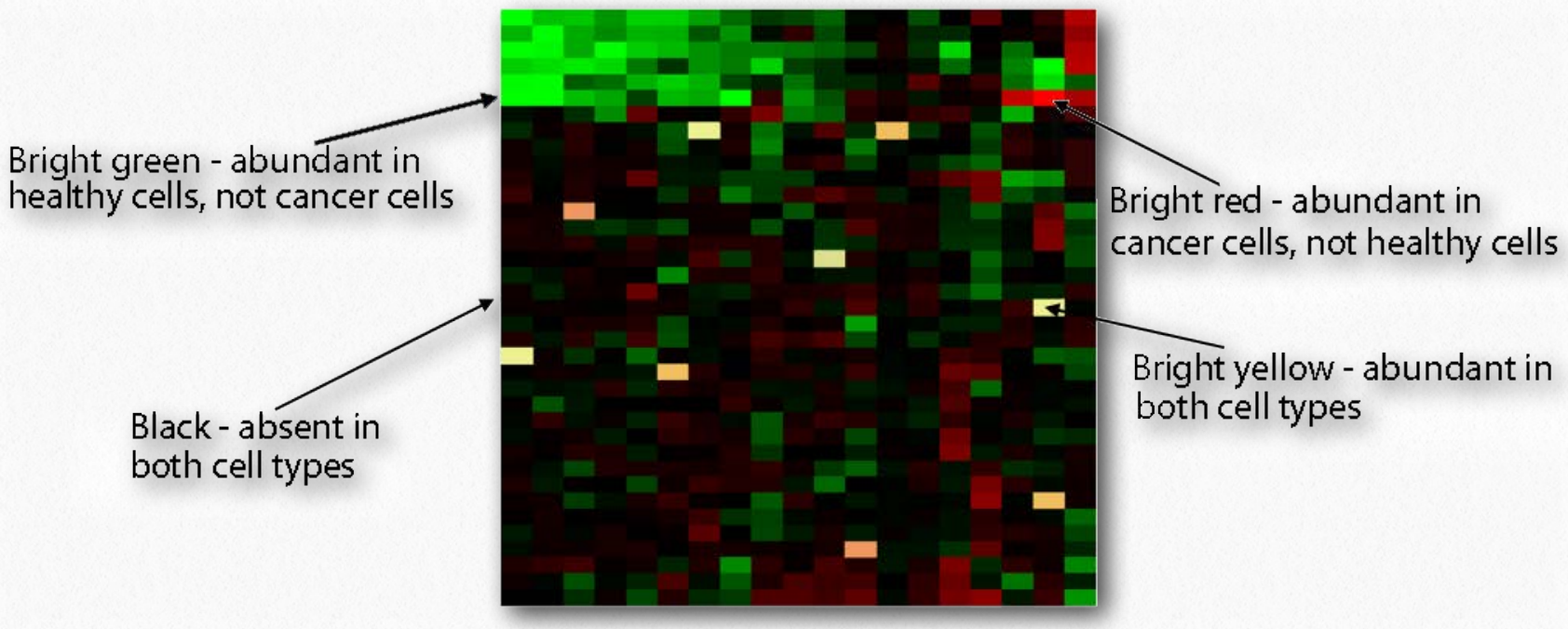


Figure 8.28 - Microarray analysis comparing gene expression in normal and cancer cells

Wikipedia

duplexes (Figure 8.27). Unhybridized cDNAs are washed away and then the plate is analyzed. Red grid boxes correspond to an mRNA present in the cancerous tissue, but not in the non-cancerous tissue. Green grid boxes correspond to an mRNA present in the non-cancerous tissue, but not in the cancerous tissue. Yellow would correspond to mRNAs present in equal abundance in the two tissues (Figure 8.28). The intensity of each spot also gives information about the relative amounts of each mRNA in each tissue.

Protein microarrays

The same principle used for nucleic acid microarrays can be adapted for analyzing other molecules. For example, polypeptides could be bonded to the glass slide instead of DNA to create a protein chip. Protein chips are useful for studying the interactions of proteins with other molecules as well as for diagnostics.

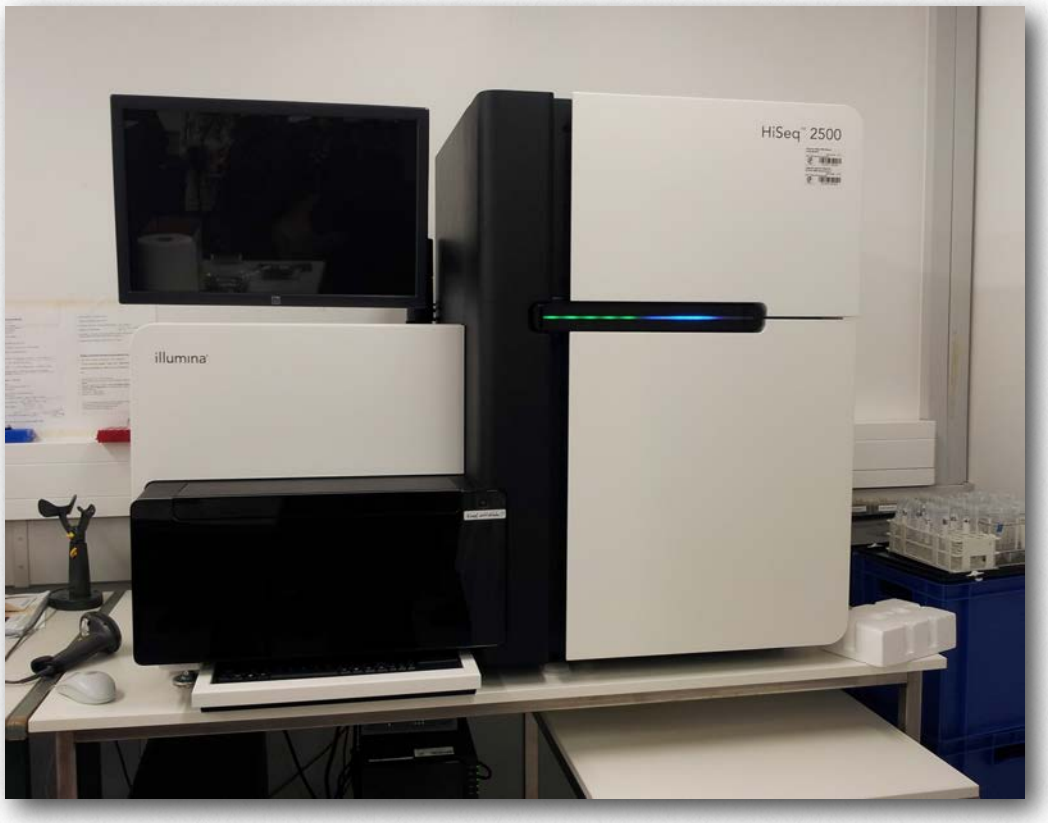


Figure 8.29 - Automated high throughput sequencer
Wikipedia

RNA-Seq

Like microarrays, a newer method called RNA-Seq, is a tool for simultaneously detecting and quantitating all of the transcripts in a given sample. This method relies on recently developed sequencing technologies called next-generation sequencing, or deep sequencing. These techniques allow for rapid, parallel sequencing of millions of DNA fragments and can, thus, be used not only for genomic DNA, but also to sequence all of the reverse-transcribed RNAs from a given sample.

To determine all the protein-coding genes that were being expressed in a particular set of cells under specific physiological conditions, all of the mRNA would first be extracted and reverse-transcribed into cDNA. This step is similar to the preparation of samples for microarrays. However, at this point, the cDNAs are fragmented into smaller pieces, and have small sequencing adapters attached at either end. The fragments are

then subjected to high-throughput sequencing, to obtain short sequences from all of the fragments. These data are aligned against the genome sequence and used to measure the level of expression of different genes. RNA-Seq offers some advantages over microarrays. With microarrays, an RNA can only be detected if the gene sequence corresponding to it is present on the grid. In RNA-Seq every RNA present in the sample is sequenced, so detection of RNAs is not limited by the probes on a chip. RNA-Seq is more sensitive than microarrays and offers a much larger range over which gene expression can be measured accurately.

YouTube Lectures by Kevin HERE & HERE

Isolating genes

Earlier in this chapter, we discussed methods such as column chromatography that are used to purify proteins of interest. Using combinations of these methods, it is possible to isolate a protein to a high degree of purity, thus enabling us to study the protein's activity and properties. This problem is harder to solve for nucleic acids. Genomic DNA can be readily obtained from cells, but is too complex to be analyzed as a whole. Individual genes are the units of DNA that correspond to proteins, and thus, it makes more sense to isolate specific genes for study. Methods to isolate genes were not available till the 1970s, when the discovery of restriction enzymes and the invention of molecular cloning provided, for the first time, ways to obtain large quantities of specific DNA fragments, for study. Although, for purposes of obtaining large amounts of a specific DNA fragment, molecular cloning has been largely replaced by direct amplification using the polymerase chain reaction described later, cloned DNAs are still very useful for a variety of reasons. The development of molecular cloning was dependent on the discovery of restriction endonucleases, described below.

Restriction enzymes

Restriction enzymes, or restriction endonucleases, are enzymes made by bacteria. These enzymes protect bacteria by degrading foreign DNA molecules that are carried into their cells by, for example, an invading bacteriophage. Each restriction enzyme recognizes a specific sequence, usually of four or six nucleotides in the DNA. These sequences, when they occur in the bacterium's own DNA, are chemically modified by methylation, so that they are not recognized and degraded. Where these sequences occur in foreign DNA, they are cut by the restriction enzyme.

The utility and importance of restriction enzymes lies in their ability to recognize specific sequences in DNA and cut near or (usually) at the site they recognize. Over 3000 such enzymes are known. Sequences recognized by these enzymes are typically 4-8 base pairs long and the most commonly used enzymes

Figure 8.30 -A restriction enzyme bound to its recognition sequence on DNA

Wikipedia

which is the same as the bottom strand (read in the same 5' to 3' direction).

While all restriction enzymes must recognize and bind to particular DNA sequences, the exact spot at which they cut the DNA varies. Some enzymes leave a staggered sequence after cutting that has an overhang at the 5' end of one strand of the duplex; some leave a staggered sequence after cutting that has an overhang at the 3' end; and some cut both strands in the same place, leaving no overhanging sequence - called blunt end cutters.

Consider cutting a DNA sequence that contains the Hind III recognition site, which is recognize sequences described as palindromic.

Palindrome

In molecular biology, the term palindrome means that the sequence of the recognition site when read in the 5' to 3' direction for the top strand is exactly the same as that of the bottom strand. Consider the sequence recognized by the restriction enzyme known as Hind III (pronounced hin-dee-three). It is

5' -A-A-G-C-T-T-3'
3' -T-T-C-G-A-A-5'

On the top strand, the recognition sequence is

5' AAGCTT 3'

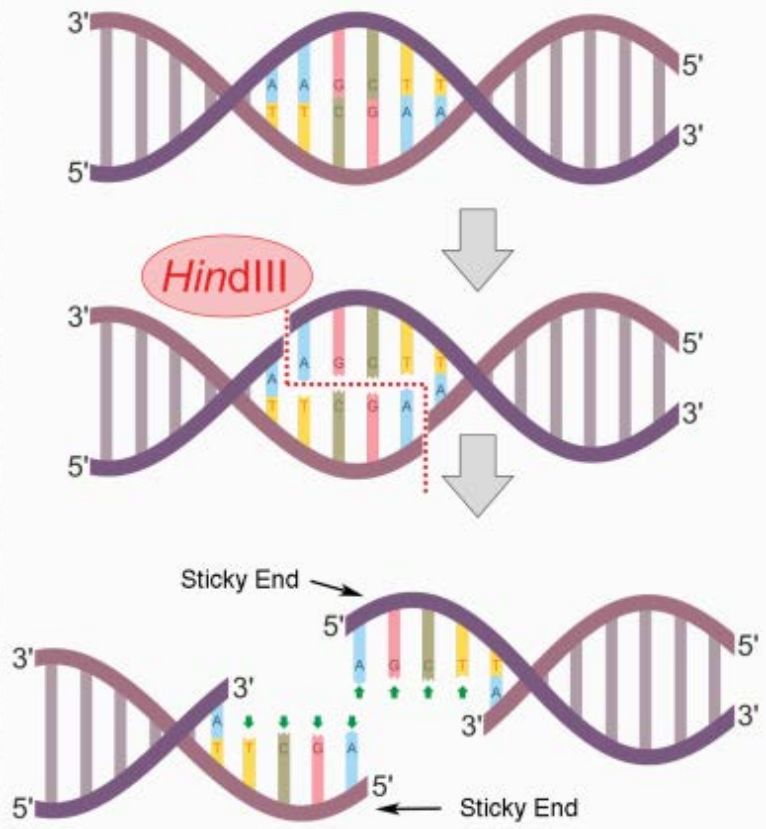


Figure 8.31 - Result of cutting DNA with Hind III

Wikipedia

5’ -A-A-G-C-T-T-3’
3’ -T-T-C-G-A-A-5’

Embedded within a DNA sequence, the Hind III sequence would look like this (Ns correspond to any base and represent all of the DNA around the recognition site).

5’ -N-N-N-A-A-G-C-T-T-N-N-N-3’
3’ -N-N-N-T-T-C-G-A-A-N-N-N-5’

After cutting with Hind III, it would look as follows:

5’ -N-N-N-A 3‘ 5’A-G-C-T-T-N-N-N-N-3’
3’ -N-N-N-T-T-C-G-A-5‘
3’ A-N-N-N-N-5’

where gaps have been inserted to illustrate where cutting has occurred. Hind III cuts between the two ‘A’ containing nucleotides near the 5’ end of the recognition sequence and thus leaves 5’ overhangs (Figure 8.31).

and cuts between the A and the G near the 3’ end of the recognition sequence.

5’ -N-N-N-C-T-G-C-A 3‘ 5’G-N-N-N-N-N 3’
3’ -N-N-N-G 5‘ 3’ A-C-G-T-C-N-N-N-N-N 5’

As you can see, cutting a DNA with Pst I leaves 3’ overhangs of the recognition sequence. The ends left after cutting by a restric-

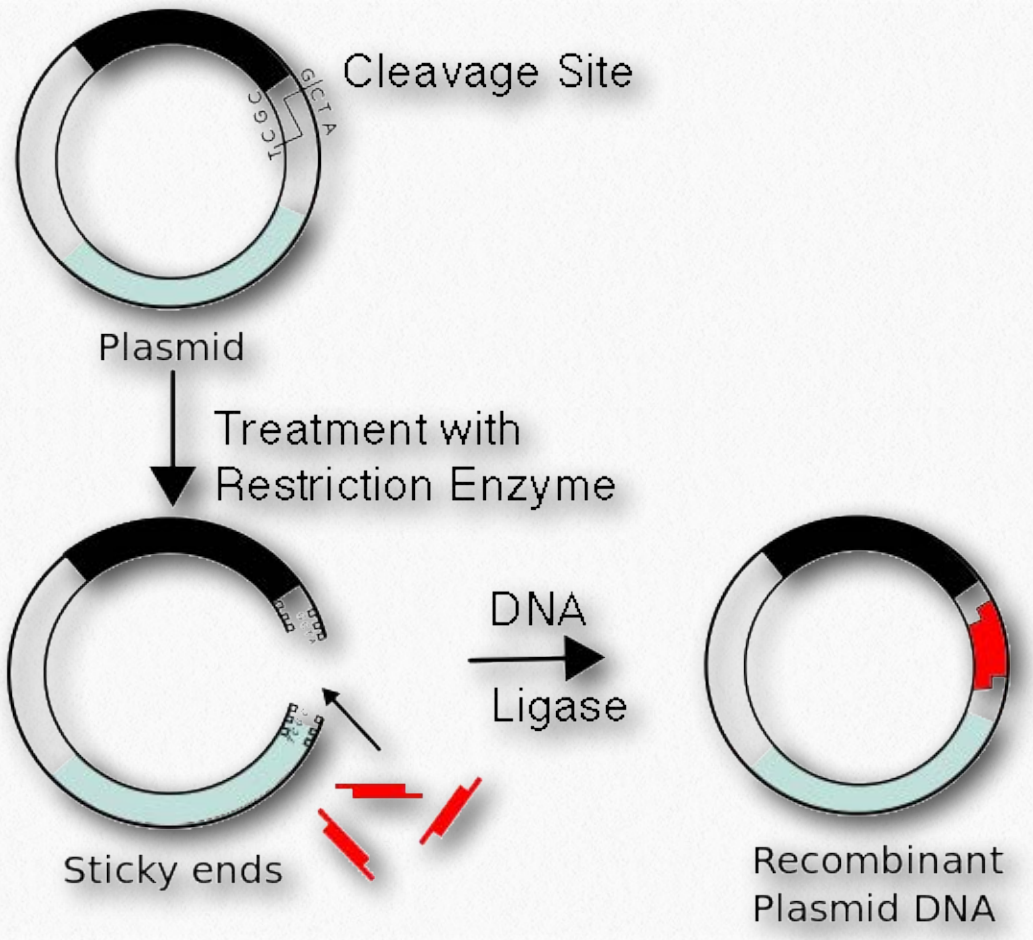


Figure 8.32 - Recombinant DNA construction

Wikipedia

The restriction enzyme Pst I, on the other hand, recognizes the following sequence

5’ -N-N-N-C-T-G-C-A-G-N-N-N-N-3’
3’ -N-N-N-G-A-C-G-T-C-N-N-N-N-5’

tion enzyme that overhang either at the 5’ end or the 3’ end are referred to as being “sticky” because they can form proper base pairs and more readily be joined to a similarly “sticky end”. This means that you can take two unrelated pieces of DNA, cut them with the same restriction enzyme so that they have

compatible sticky ends, and then "paste" them together using DNA ligase to form a new hybrid molecule, or recombinant.

Making recombinant DNAs

Joining together of DNA fragments from different sources creates recombinant DNA. The ability to cut and paste DNA might seem like purely a technical feat, but one key application that arose out of this is molecular cloning. In molecular cloning a gene of interest can be inserted into a vector, usually a plasmid, by cutting both the vector and the gene (called the insert) with the same enzyme to generate sticky ends and joining the two pieces together to generate a recombinant (Figure 8.32). A plasmid is a type of autonomously replicating, extrachromosomal DNA. It is quite simple to extract plasmids from the cells, engineer them to contain the gene of interest and re-introduce the recombinant plasmid into the bacteria. The idea was that when the plasmid DNA was replicated, the extra inserted gene would also be copied. Thus, by growing up a lot of the bacteria carrying the plasmid, many copies of the gene of interest could be obtained, to provide sufficient amounts of the gene to use in experiments. While we now have easier methods to accomplish this goal, cloned DNAs remain very useful. For example, it is possible to clone a gene that encodes a protein of interest so that it can be expressed at high levels in the cells into which the recombinant plasmid is introduced.

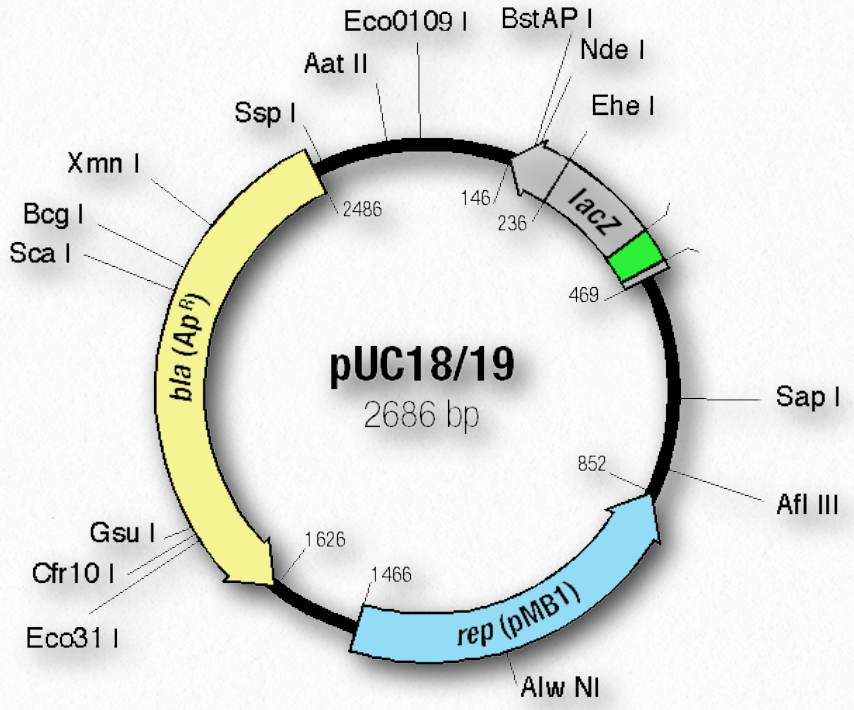


Figure 8.33 - Restriction site map for the pUC 18/19 plasmids, a classic plasmid vector. Genes identified by arrows. Numbers correspond to the pUC 18/19 numbering convention

Whatever the purpose for which the recombinant plasmid is made, it typically carries an antibiotic resistance gene (or genes), called a *selectable marker*. Cells that take up the plasmid will be able to grow in the presence of the antibiotic. If bacterial cells to which the plasmid has been added are plated on agar containing the antibiotic, the cells which took up the plasmid will be able to grow, while the others will not.

Expression cloning

As mentioned above, a gene of interest may be inserted into a vector and the recombinant plasmid be placed into a cell where the gene can be expressed. For instance, one might desire to clone the gene coding for human growth hormone or insulin or other medically important proteins and have a bacterium or yeast make large quantities of it very cheaply. Remember that these are human proteins, and thus it is not feasible to extract the proteins in any quantity from human subjects.

Expression vectors

To clone a gene so that it can be expressed, one needs to set up the proper conditions in order for the human protein to be made in the bacterial cells. This typically involves the use of specially designed plasmids. These plasmids have been engineered to 1) replicate in high numbers; 2) carry markers that allow researchers to identify cells carrying them (antibiotic resistance, for example) and 3) contain sequences (such as a promoter and Shine Dalgarno sequence) necessary for expression of the desired protein, with convenient sites for insertion of the gene of interest in the appropriate place relative to the control sequences. A plasmid which has all of these features is referred to as an expression vector. In addition to plasmids that can be used for expression in bacterial cells, expression vectors are also available that allow protein expression in a variety of eukaryotic cells.

Many sophisticated variations on such vectors have been created that have made it easy to produce and purify large amounts of any pro-

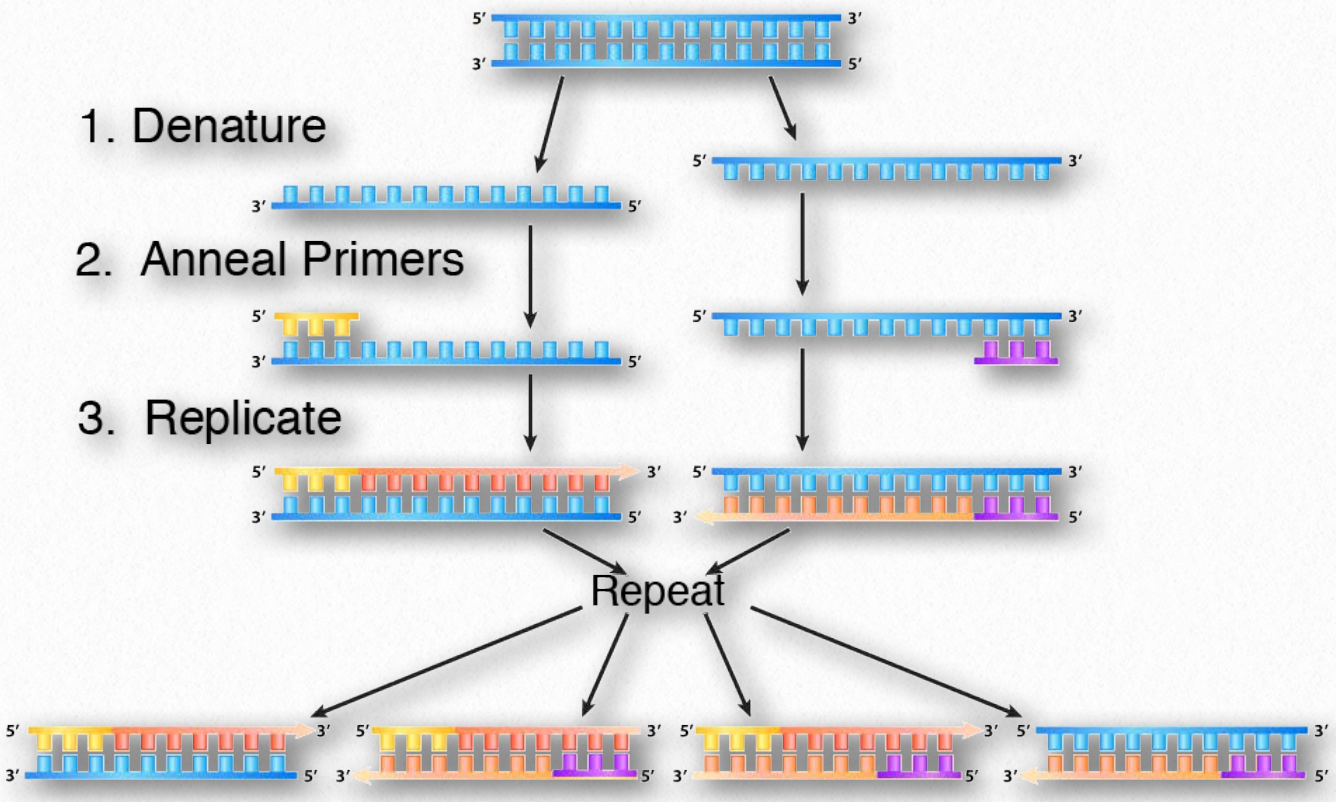


Figure 8.34 - Steps in the polymerase chain reaction

Image by Aleia Kim

tein of interest for which the gene has been cloned. A handy feature in some expression vectors is a sequence encoding an affinity tag either up- or downstream of the gene being expressed. This sequence allows a short affinity tag (such as a run of histidine residues) to be fused onto the encoded protein. The tag can be used to readily purify the protein, as described in the section on affinity chromatography.

Figure 8.35 - PCR tubes with DNA samples ready for reaction

Wikipedia

Polymerase chain reaction (PCR)

As we just saw, molecular cloning was the first method available to isolate a gene of interest and make many copies of it to obtain sufficient amounts of the DNA to study. Today, there is a faster and easier way to obtain large amounts of a DNA sequence of interest -the polymerase chain reaction, or PCR.

PCR allows one to use the power of DNA replication to amplify DNA enormously in a short period of time. As you know, cells replicate their DNA before they divide, and in doing so, double the amount of the cell's DNA. PCR essentially mimics cellular DNA replication in the test tube, repeatedly copying the target DNA over and over, to produce large quantities of the desired DNA.

YouTube Lectures by Kevin HERE & HERE

Selective replication

In contrast to cellular DNA replication, which amplifies all of a cell's DNA during a replication cycle, PCR does targeted amplification to replicate only a segment of DNA bounded by the two primers that determine where DNA polymerase begins replication. Figure 8.34 illustrates the process. Each cycle of PCR involves three steps, denaturing, annealing and extension, each of which occurs at a different temperature.

The starting materials

Since PCR is, basically, replication of DNA in a test-tube, all the usual ingredients needed for DNA replication are required:

A template (the DNA containing the target sequence that is being copied)

Primers (to initiate the synthesis of the new DNA strands)

Thermostable DNA polymerase (to carry out the synthesis). The polymerase needs to be heat stable, because heat is used to separate the template DNA strands in each cycle.

dNTPs (DNA nucleotides to build the new DNA strands).

The template is the DNA that contains the tar-

get you want to amplify (the "target" is the specific region of the DNA you want to amplify). The primers are short synthetic single-stranded DNA molecules whose sequence matches a region flanking the target sequence. It is possible to chemically synthesize DNA molecules of any given base sequence, to use as primers. To make primers of the correct sequence that will bind to the template DNA, it is necessary to know a little bit of the template sequence on either side of the region of DNA to be amplified. DNA polymerases and dNTPs are commercially available from biotechnology supply companies.

Figure 8.36 - A PCR thermocycler system

Wikipedia

First, all of the reagents are mixed together. Primers are present in millions of fold excess over the template. This is important because each newly made DNA strand starts from a primer. The first step of the process involves separating the strands of the target DNA by heating to near boiling.

Interactive Learning Module HERE

Next, the solution is cooled to a temperature that favors complementary DNA sequences finding each other and making base pairs, a process called annealing. Since the primers are present in great excess, the complementary sequences they target are readily found and base-paired to the primers. These primers direct the synthesis of DNA. Only where a primer anneals to a DNA strand will replication occur, since DNA polymerases require a primer to begin synthesis of a new strand.

Extension

In the third step in the process, the DNA polymerase replicates DNA by extension from the 3' end of the primer, making a new DNA strand. At the end of the first cycle, there are twice as many DNA molecules, just as in cellular replication. But in PCR, the process is repeated, usually for between 25 and 30 cycles. At the end of the process, there is a theoretical yield of 2^{30} (over 1 billion times) more DNA than there

was to start. (This enormous amplification power is the reason that PCR is so useful for forensic investigations, where very tiny amounts of DNA may be available at a crime scene.)

The temperature cycles are controlled in a thermocycler, which repeatedly raises and lowers temperatures according to the set program. Since each cycle can be completed in a couple of minutes, the entire amplification can be completed very rapidly. The resulting DNA is analyzed on a gel to ensure that it is of the expected size, and depending on what it is to be used for, may also be sequenced, to be certain that it is the desired fragment.

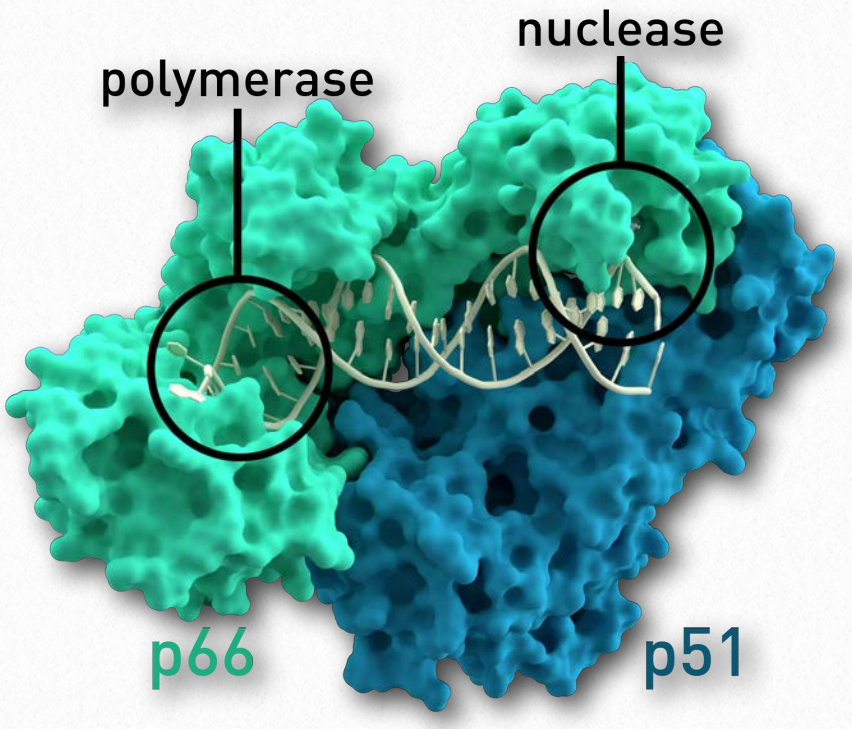


Figure 8.37 - Reverse transcriptase of HIV. The nuclease function is needed for the viral life cycle, but not for lab use.

Wikipedia

Mutagenesis

PCR is frequently used to obtain gene sequences to be cloned into vectors for protein expression, for example. Besides simplicity and speed, PCR also has other advantages. Because primers can be synthesized that differ from the template sequence at any given position, it is possible to use PCR for site-directed mutagenesis. That is, PCR can be used to mutate a gene at a desired position in the sequence. This allows the proteins encoded by the normal and mutant genes to be expressed, purified and compared.

Analysis of gene expression

PCR can also be used to measure gene expression. Where in PCR the amount of amplified product is not determined till the end of all the cycles, a variation called quantitative real-time PCR is used, in which the accumulation of product is measured at each cycle. This is possible because real-time PCR machines have a detector module that can measure the levels of a fluorescent marker in the reaction, with the amount of fluorescence proportional to the amount of amplified product. By following the accumulation of product over the cycles it is possible to calculate the amount of starting template. To measure gene expression, the template used is mRNA reverse-transcribed into cDNA (see below). This coupling of reverse transcription with quantitative real-time PCR is called qRT-PCR.

Reverse transcription

In the central dogma, DNA codes for mRNA, which codes for protein. One known exception to the central dogma is exhibited by retroviruses. These RNA-encoded viruses have a phase in their life cycle in which their genomic RNA is converted back to DNA by a virally-encoded enzyme known as reverse transcriptase. The ability to convert RNA to DNA is a method that is desirable in the laboratory for numerous reasons. For example, converting RNAs of interest to cDNA is used in RT-PCR as well as in other applications like microarray analysis.

Process

First, one creates a DNA oligonucleotide to serve as a primer for reverse transcriptase to use on a target RNA. The primer must, of course, be complementary to a segment (near the 3' end) of the RNA to be amplified. The RNA, reverse transcriptase, the primer, and four dNTPs are mixed. With one round of replication, the RNA is converted to a single strand of DNA. Denaturation frees the single stranded cDNA, which can be used as is, or converted to double-stranded cDNA, depending on the application.

Detecting molecular interactions

The study of biochemistry is basically the study of the interactions of cellular molecules. Methods for detecting interactions among biomolecules are, for this reason, very useful to biochemists. We will now discuss a couple of very different methods for detecting these inter-molecular interactions.

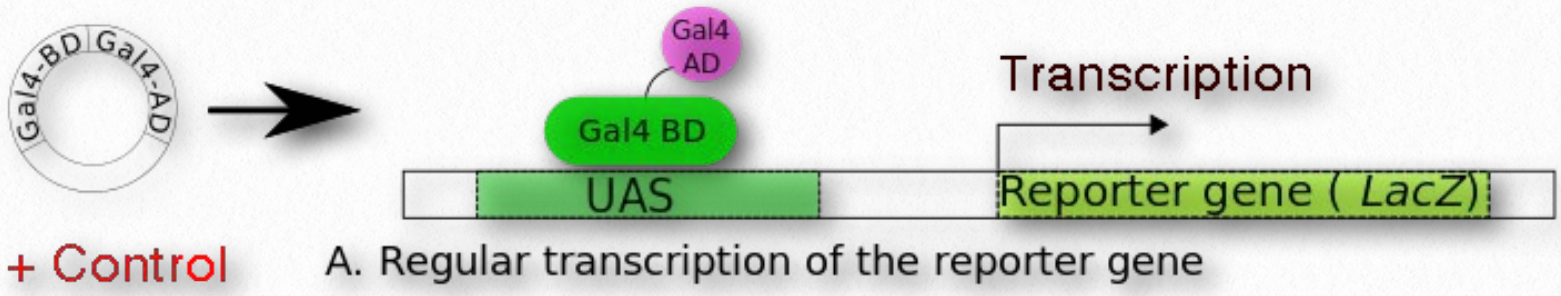
Yeast two-hybrid system (Y2H)

Yeast two-hybrid screening is a sophisticated technique for identifying which protein(s), out of a collection of all of a cell's proteins, interacts with a specific protein of interest. The method relies on the interaction between two proteins to reconstitute a functional transcriptional activator within yeast cells. You may remember that many transcriptional activators are modular proteins that have a domain that binds to DNA and another domain that activates transcription (Figure 8.38).

Figure 8.38 - Structure of a eukaryotic transcriptional activator, showing the DNA-binding domain (DBD) and transcriptional activation domain (TAD).

Wikipedia

If the transcription factor is split, so that the binding domain is attached to one protein, and the activation domain to another protein, a functional transcriptional activator can only be re-created if the two "carrier"proteins come into close proximity - that is, they inter-



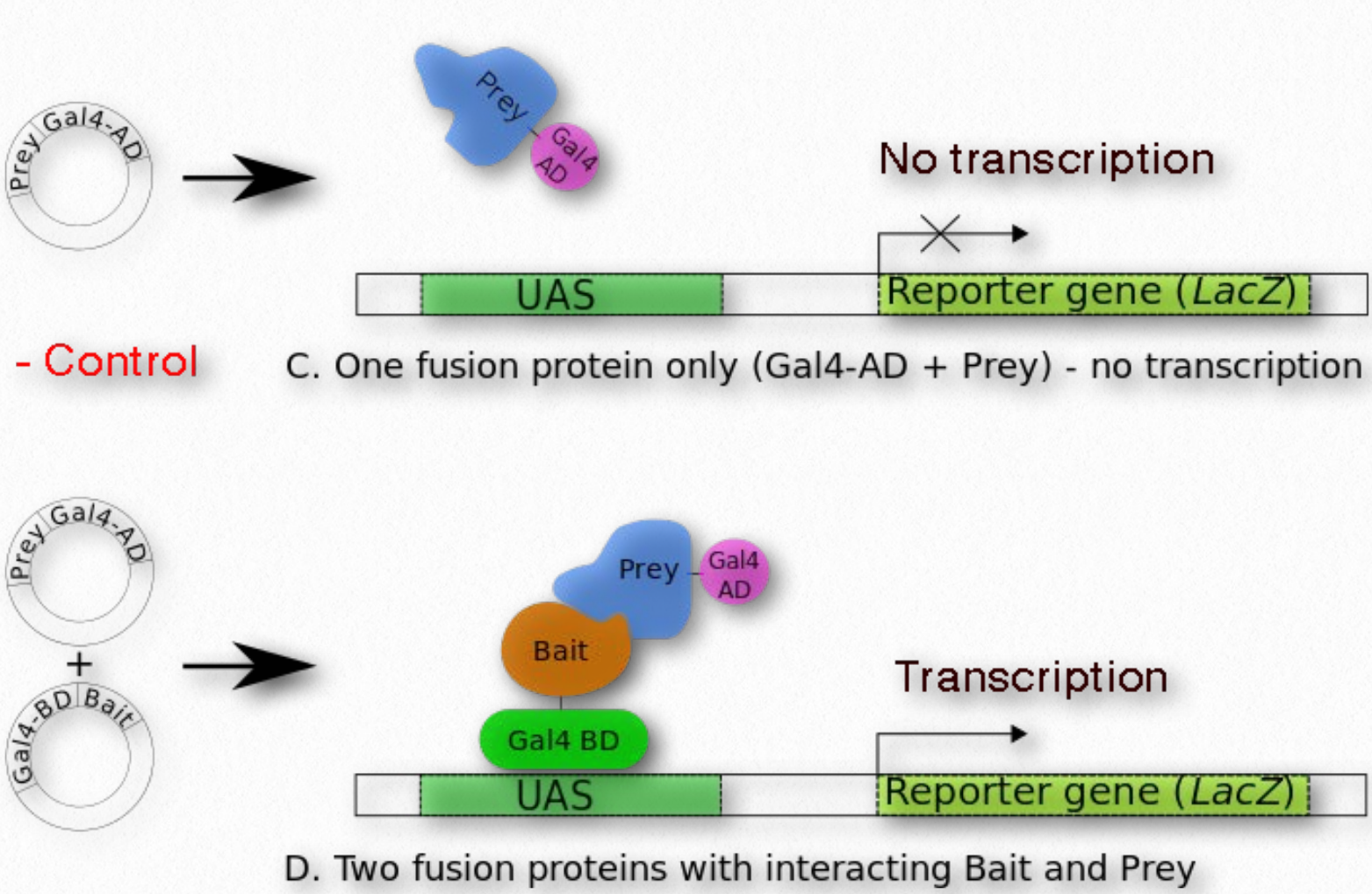


Figure 8.39 - Four scenarios for the yeast two-hybrid system. UAS = Upstream Activator Sequences - acts like a promoter. Scenario A shows that the two transcription factors start out as one protein

Wikipedia

act. The presence of this functional activator can be detected by the expression of a reporter gene.

A simple way to understand this idea is by thinking of a transcriptional activator as a device, like a flashlight, that has two parts, the battery and the lamp, that must be together in order to function. Neither a person who has just a battery nor one who has only the lamp will be able to see in a dark room. But if the two interact by coming close enough to insert the battery in the flashlight, their interaction can be detected by the fact that the flashlight will now be functional as evidenced by the light produced.

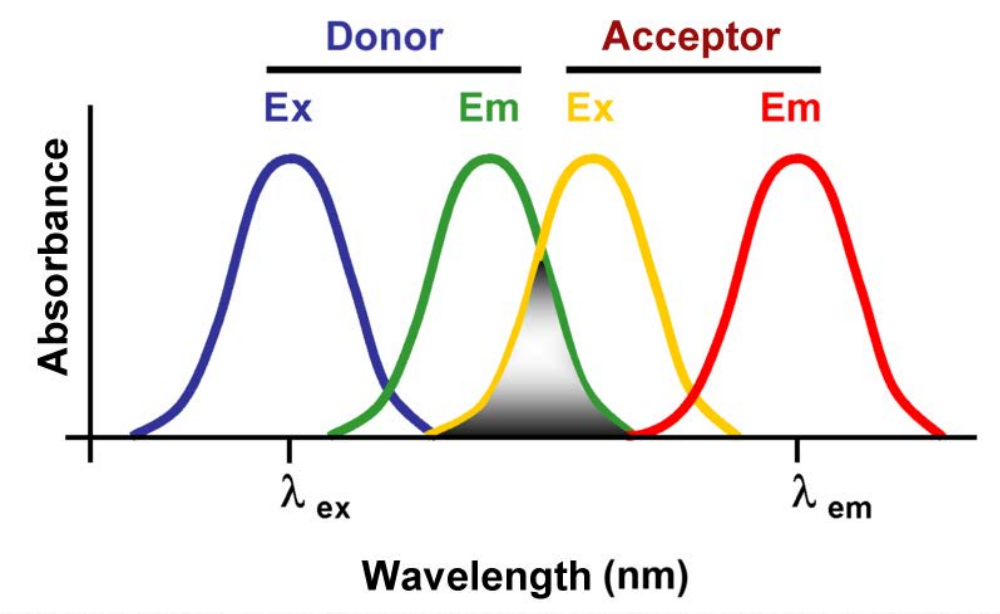


Figure 8.40 - Excitation and emission spectra for donor and acceptor fluorophores in FRET

Wikipedia

It takes two to tango

Figure 8.39 (A) shows the normal yeast transcriptional activator, GAL4, with both the DNA-binding (DBD) and Activation domains (AD). It is able to stimulate transcription of the downstream reporter gene, *lac z*. Panels B and C show constructs that produce the GAL4 DBD and AD, respectively, fused to other proteins, one of which is termed the "bait" and the other as "prey". Neither of these fusion proteins can stimulate transcription of the *lac z* gene. When constructs encoding both the bait and prey are in the same yeast cell, if the bait protein interacts with the prey, the DBD and AD of the GAL4 will be brought together to reconstitute a functional GAL4. The presence of functional GAL4 is readily detectable because it will stimulate expression of the l*ac z* reporter gene. If the bait and prey proteins do not interact, then there will be no lac z expression.

When interaction is detected through expression of the reporter gene, the specific prey protein can then be identified.

YouTube Lectures by Kevin HERE & HERE

The yeast two-hybrid system allows for simultaneous screening of many prey proteins, by constructing large collections of fusion constructs, with each potential protein partner of the bait protein fused to the GAL4 activation domain.

FRET

Another method for detecting molecular interactions is Fluorescence resonance energy transfer (FRET) - also called Förster reso-

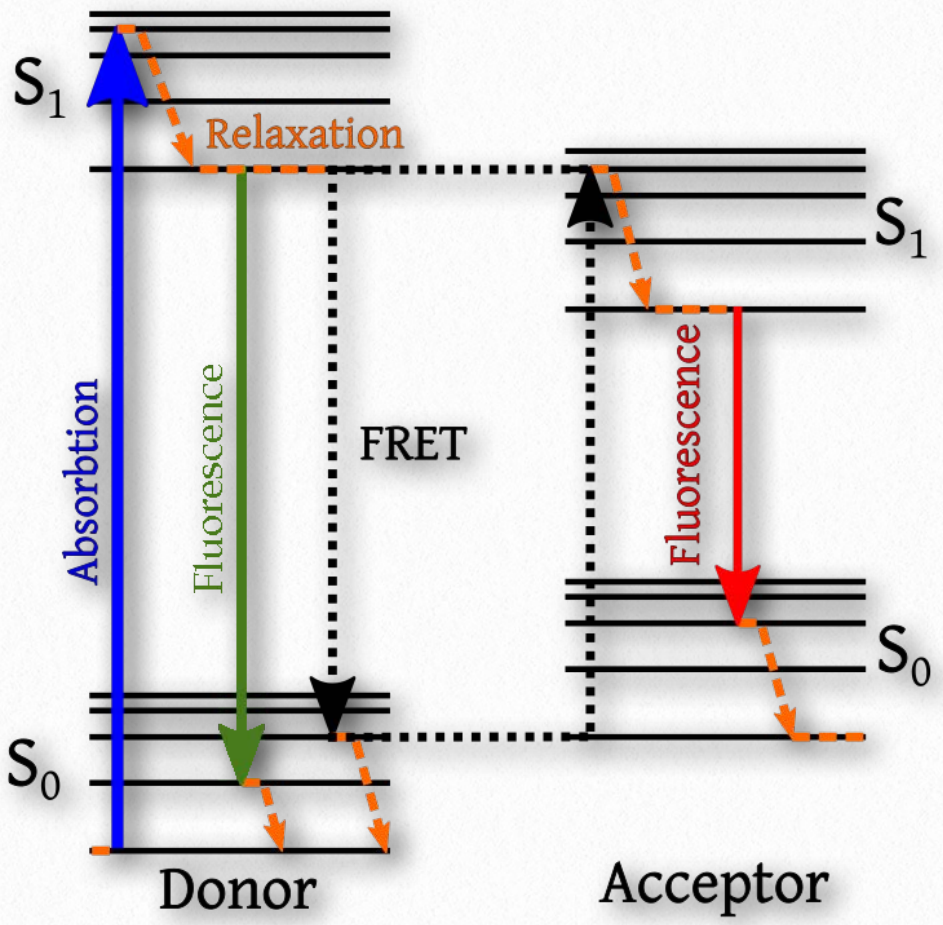


Figure 8.41 - Fluorescence resonance energy transfer between donor and acceptor chromophores

Image by Pehr Jacobson

nance energy transfer, resonance energy transfer (RET) or electronic energy transfer (EET). The technique is based on the observation that a molecule excited by the absorption of light can transfer energy to a nearby molecule if the emission spectrum of the first molecule overlaps with the excitation spectrum of the second (Figure 8.40) This transfer of energy can only take place if the two molecules are sufficiently close together (no more than a few nanometers apart).

Design

In the technique, a donor fluorophore or an acceptor fluorophore is covalently attached to two molecules of interest. The acceptor fluorophore is designed to accept energy from the donor molecule (orange dotted line in Figure 8.41) and fluoresce at a unique wavelength (red arrow) when it receives that energy from the donor.

Further, the wavelength of light that the donor absorbs is uniquely tailored for the donor fluorophore and has no effect on the acceptor fluorophore. The only way the acceptor can fluoresce is if it is close enough to receive energy transferred from the donor (red arrow). This fluorescence will have a unique wavelength, as well. If the donor and acceptor are not close enough together, the donor fluoresces and emits light corresponding to the green or black arrow. These are different wavelengths than that of the red arrow.

The experiment begins in the cell with one protein with a donor fluorophore and the other protein with an acceptor fluorophore. Light of a wavelength that excites the donor fluorophore is shined on the cell. If a protein with a donor interacts with the protein carrying an acceptor, then energy transfer occurs from the donor fluorophore to the acceptor and the unique fluorescence (red line) of the acceptor is detected. If the two proteins do not interact, then little or no fluorescence from the acceptor is detected.

Genome editing

The development of tools that would allow scientists to make specific, targeted changes in the genome has been the Holy Grail of molecular biology. An ingenious new tool that is both simple and effective in making precise changes is poised to revolutionize the field, much as PCR did in the 1980s. Known as the CRISPR/Cas9 system, and often abbreviated simply as CRISPR, it is based on a sort of bacterial immune system that allows bacteria to recognize and inactivate viral invaders.

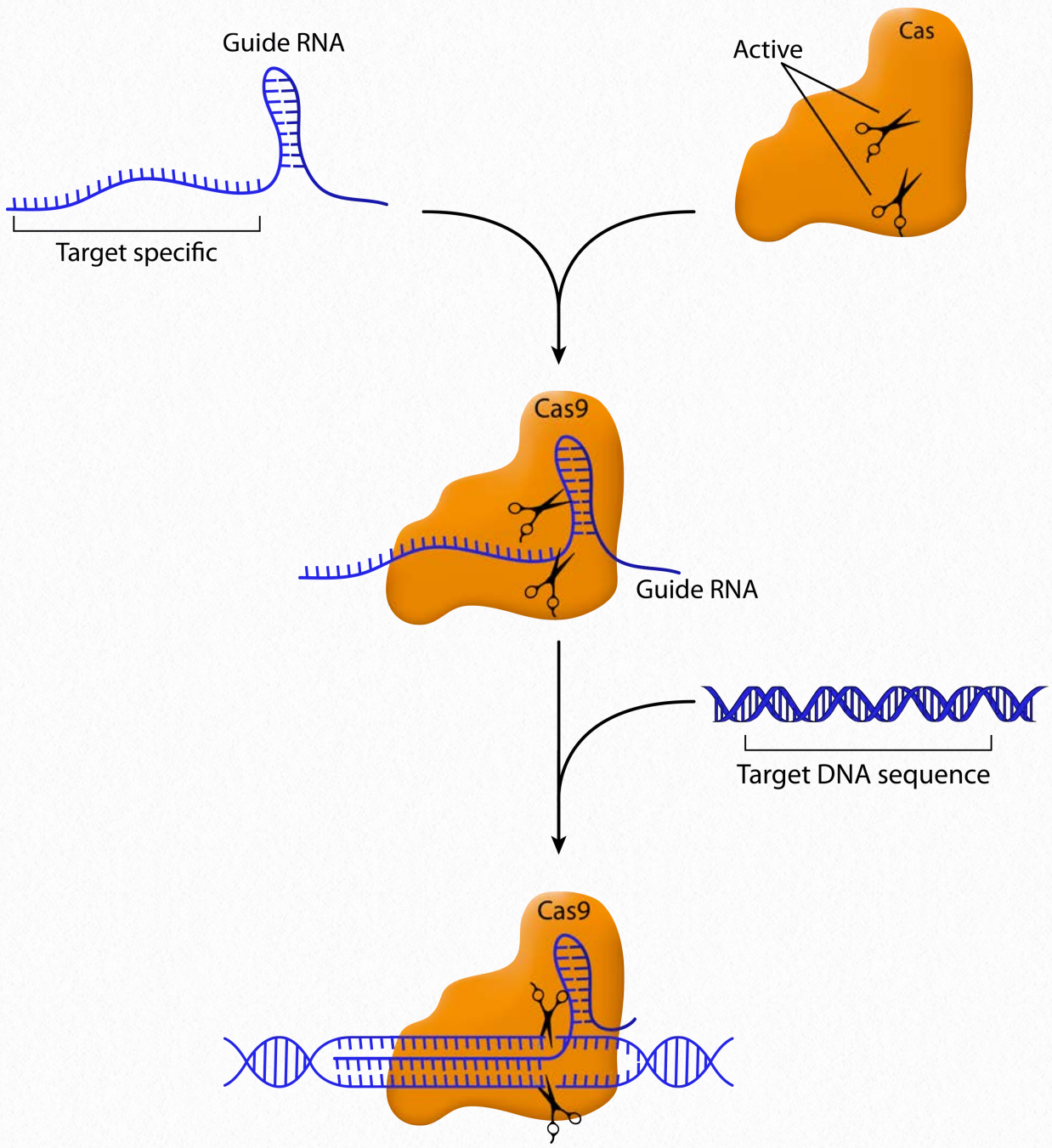


Figure 8.42 - A guide RNA directs the Cas9 nuclease to its target gene

Image by Pehr Jacobsen

CRISPR stands for Clustered Regularly Interspaced Short Palindromic Repeats, short repeated sequences found in prokaryotic DNA, separated by spacer sequences derived from past encounters with, for example, a bacteriophage. Like the glass slipper left behind by Cinderella that was later used to identify her, the pieces of the invader's sequences are a way for the bacteria to identify the virus if it attacks again. Inserted into the bacterial genome, these sequences can later be transcribed into a guide RNA that matches, and base-pairs with, sections of the viral genome if it was encountered again. A nuclease associated with the guide RNA then cleaves the sequence base-paired with the guide RNA. (The nucleases are named Cas for CRISPR-associated.)

The essential elements of this system are a guide RNA that homes in on the target sequence and a nuclease that can make a cut

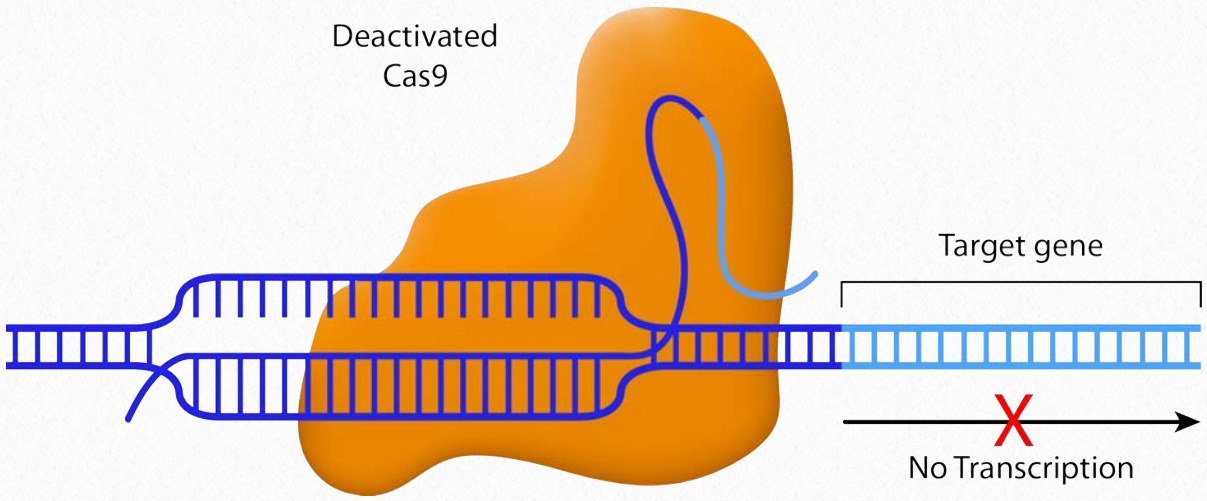


Figure 8.43 - Inactive Cas9 can block transcription of a target gene

Image by Pehr Jacobsen

in the sequence that is bound by the guide RNA. By engineering guide RNAs complementary to a target gene, it is possible to target the nuclease to cleave within that gene. In the CRISPR/Cas9 system, the Cas9 endonuclease cuts both strands of the gene sequence targeted by the guide RNA (Figure 8.42). This generates a double-strand break that the cell attempts to repair.

As you may remember, double-strand breaks in DNA can be repaired by simple, non-homologous end joining (NHEJ) or by homologous recombination. When a break is fixed by NHEJ, there is good chance that there will be deletions or insertions that will inactivate the gene they are in. Thus, targeted cleavage of a site by CRISPR/Cas9 can easily and specifically inactivate a gene, making it easy to characterize the gene's function.

Interactive Learning Module HERE

But, what if you wished to simply mutate the gene at a specific site to study the effect of the mutation? This, too, can be achieved. If a homologous sequence bearing the specific mutation is provided, homologous recombination can repair the break, and at the same time insert the ex-

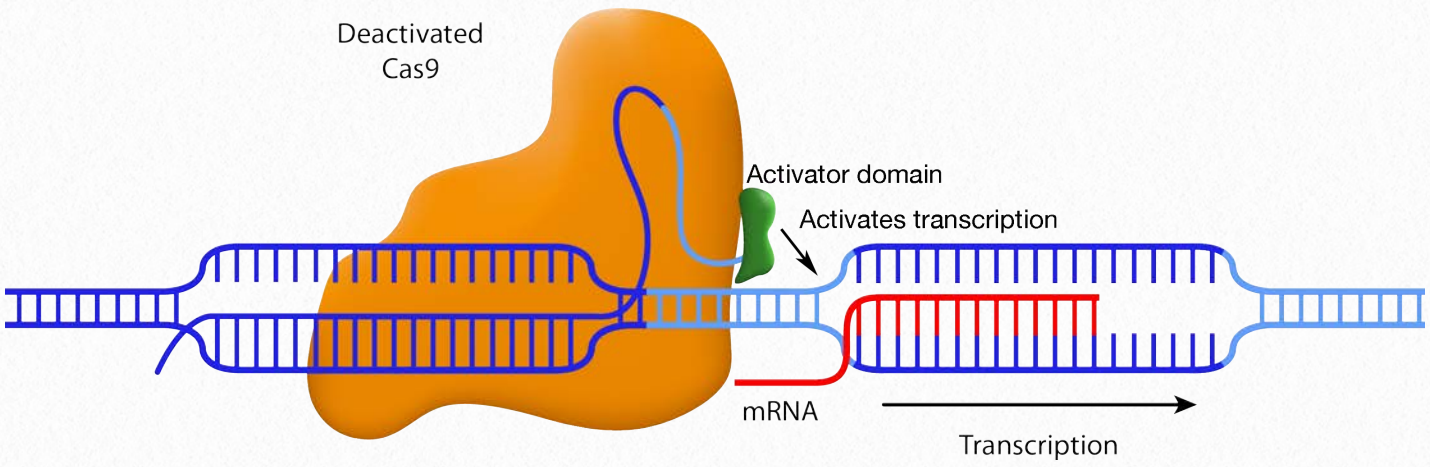


Figure 8.44 - A Cas9-activator domain fusion can activate transcription of a target gene

Image by Pehr Jacobsen

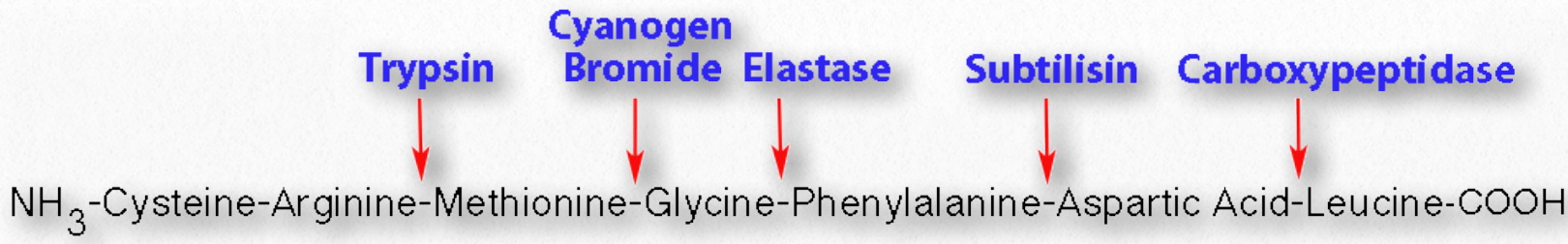


Figure 8.45 - Protease cleavage sites on a polypeptide

act mutation desired. It is obvious that if you can insert a mutation as just described, it should be possible to correct a mutation in the genome by cleaving at the appropriate spot and providing the correct sequence as a template for repair by homologous recombination. The simplicity of the system holds great promise for curing genetic diseases.

Scientists have also come up with some creative variations on the CRISPR/Cas9 system. For instance, one variant inactivates the nuclease activity of Cas9. The guide RNA in this system pairs with the target sequence, but the Cas9 does not cleave it. Instead, the Cas9 blocks the transcription of the downstream gene (Figure 8.43) This method allows specific genes to be turned off without actually altering the DNA sequence.

Another variation also uses a disabled Cas9, but this time, the Cas9 is fused to a transcriptional activation domain. In this situation, the guide RNA positions the Cas9-activator domain in a place where it can enhance transcription from a specific promoter (Figure 8.44). Other variations on this theme attach histone-modifying enzymes or DNA methylases to the inactive Cas9. Again, the guide RNA positions the Cas9 in the desired spot, and the enzyme attached to Cas9 can methylate the DNA or modify the histones in that region.

Protein Cleavage Agent Cut Sites

Subtilisin - C-terminal side of large uncharged side chains

Chymotrypsin - C terminal side of aromatics (Phe, Tyr, Trp)

Trypsin - C-terminal side of lysine and arginines (not next to proline)

Carboxypeptidase - N-terminal side of C-terminal amino acid

Elastase - Hydrolyzes C-side of small AAs (Gly, Ala)

Cyanogen Bromide (chemical) - Hydrolyzes C-side of Met

CRISPR has already been used to edit genomes in a wide variety of species (and in human cell cultures). It may not be long before the technique is approved for clinical use. In the meanwhile, CRISPR is transforming molecular biology.

Protein cleavage

Because of their large size, intact proteins can be difficult to study using analytical techniques, such as mass spectrometry. Consequently, it is often desirable to break a large polypeptide down into smaller pieces. Proteases are enzymes that typically break peptide bonds by binding to specific amino acid sequences in a protein and catalyzing their hydrolysis.

Chemical reagents, such as cyanogen bromide, which cleaves peptide bonds on the C-terminal side of a methionine residue can also be used to cut larger proteins into smaller peptides. Common proteins performing this activity are found in the digestive system and are shown in the yellow box below.

Determining mass and protein sequence

Mass spectrometry, as its name suggests, is a method that can be used to determine the masses of molecules. Once limited to analyzing small molecules, it has since been adapted and improved to allow the analysis of biologically important molecules like proteins and nucleic acids. Mass spectrometers use an electrical field to accelerate an ionized molecule toward a detector. The time taken by an ionized molecule to move from its point of ionization to the detector will depend on both its mass and its charge and is termed its time of flight (TOF).

Figure 8.46 - A desktop MALDI-TOF system

MALDI-TOF

MALDI-TOF (Matrix-assisted Laser Desorption Ionization - Time of Flight) is an analytical technique allowing one to determine the molecular masses of biologically relevant

molecules with great precision. It is commonly used in proteomics and determination of masses of large biomolecules, including nucleic acids. The development of MALDI, which permits the production of ionic forms of relatively large molecules, was crucial to the successful use of mass spectrometry of biomolecules. Figure 8.46 shows a compact MALDI-TOF system.

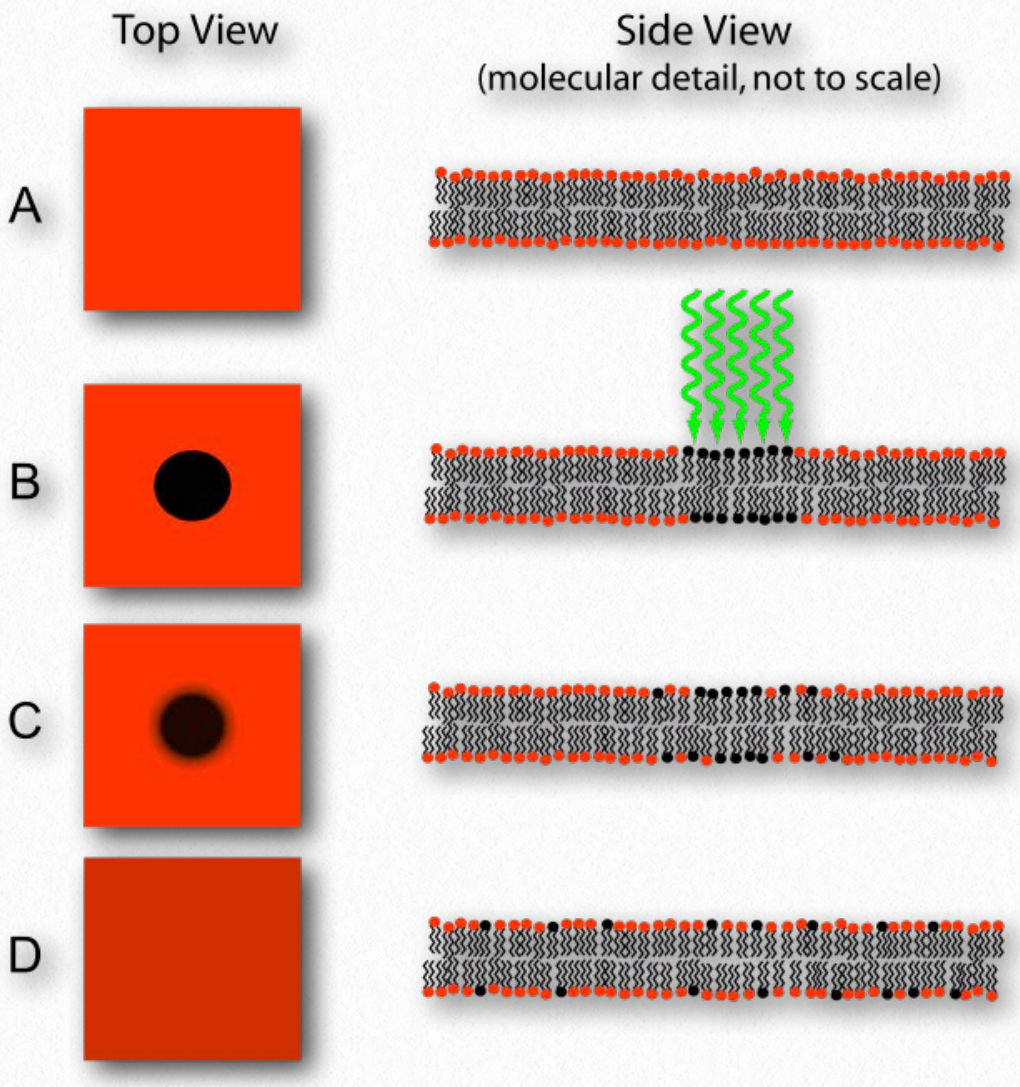
The MALDI-TOF process involves three basic steps. First, the material to be analyzed is embedded in solid support material (matrix) that can be volatilized in a vacuum chamber by a laser beam. In the second part of the process, a laser focused on the matrix volatilizes the sample, causing the molecules within it to vaporize and, in the process, to form ions by either gaining or losing protons. Third, the ions thus created in the sample are accelerated by an electric field towards a detector. Their rate of movement towards the detector is a function of the ratio of their mass to charge (m/z). An ion with a mass of 100 and a charge of +1 will move twice as fast as an ion with a mass of 200 and a charge of +1 and at the same rate as an ion with a mass of 200 and a charge of +2. Thus, by precisely determining the time it takes for an ion to go from ionization (time zero of the laser treatment) to being detected, the mass to charge ratio for all of the molecules in a sample can be readily determined.

Ionization may result in destabilization of larger molecules, which fragment into smaller ones in the MALDI-TOF detection chamber. The size of each of the sub-fragments of a larger molecule allows one to determine its identity if this is not previously known. This fragmentation can be intentionally enhanced by having the accelerated ions collide with an inert gas, like argon.

Fragmentation of a molecule may also be carried out prior to analysis, as for example, by cleaving a protein into smaller peptides by the use of enzymes or chemical agents. The amino acid sequence of a protein may be determined by using MALDI-TOF by analyzing the precise molecular masses of the many short peptide fragments obtained from a protein. When one amino acid, for example, fragments from a larger peptide, this can be detected as the difference in mass between the fragment with and without the amino acid, since each amino acid will have a characteristic molecular mass. By peptide mass fingerprinting and analysis of smaller fragments of individual peptides, the entire sequence of a polypeptide can, thus, be determined.

Membrane dynamics

Understanding the dynamics of movement in the membranes of cells is the province of the Fluorescence Recovery After Photobleaching (FRAP) technique (Figure 8.47). This optical technique is used to measure the two dimensional lateral diffusion of mole-



bleached using a laser (Step B). The spread of the bleached molecules is followed using a microscope (Step C). Information obtained in this manner provides data about the rate of lateral diffusion occurring in a lipid bilayer (Step D).

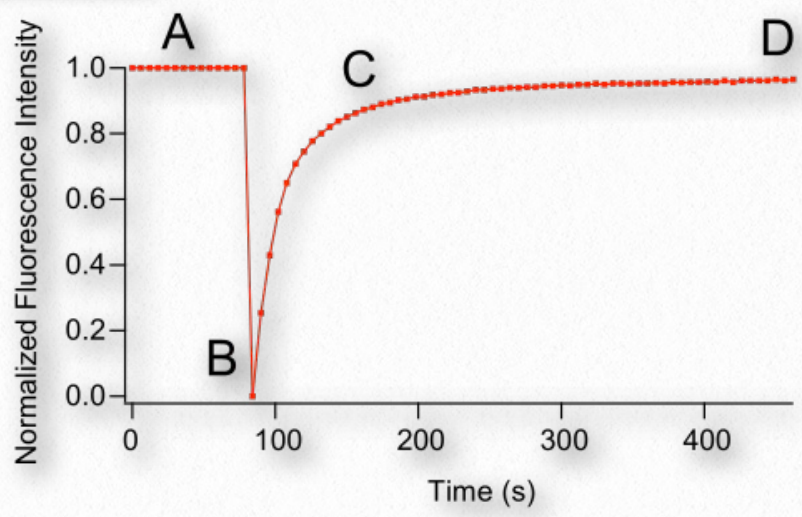


Figure 8.47 - FRAP analysis

Wikipedia

cules in thin films, like membranes, using fluorescently labeled probes. It also has applications in protein binding.

In the method, a lipid bilayer is uniformly labeled with a fluorescent tag (Figure 8.47, Step A) and then a subset of the tag is

Graphic images in this book were products of the work of several talented students. Links to their Web pages are below

Click HERE for Aleia Kim's Web Page

Click HERE for Pehr Jacobson's Web Page

Click HERE for Penelope Irving's Web Page

Click HERE for Martha Baker's Web Page

Problem set related to this section **HERE**

Point by Point summary of this section **HERE**

To get a certificate for mastering this section of the book, click **HERE**

Kevin Ahern's free iTunes U Courses - Basic / Med School / Advanced

Biochemistry Free & Easy (our other book) HERE / Facebook Page

Kevin and Indira's Guide to Getting into Medical School - iTunes U Course / Book

To see Kevin Ahern's OSU ecampus courses - BB 350 / BB 450 / BB 451

To register for Kevin Ahern's OSU ecampus courses - BB 350 / BB 450 / BB 451

Biochemistry Free For All Facebook Page (please like us)

Kevin Ahern's Web Page / Facebook Page / Taralyn Tan's Web Page

Kevin Ahern's free downloads HERE

OSU's Biochemistry/Biophysics program HERE

OSU's College of Science HERE

Oregon State University HERE

Email **Kevin Ahern** / **Indira Rajagopal** / **Taralyn Tan**

The Restriction Enzyme Song

To the tune of "*Chim Chim Cher-ee*"

Metabolic Melodies Website HERE

I'm obsessed with A-A-G-C-T-T
Cuz it is the binding site of Hin-d-III
Cutting up DNA most readily
The ends are not blunt when they're cut up you see

Five prime overhangs of A-G-C-T

Bacteria don't have an immune system so
They must fight off phages or they will not grow
Protection by chopping is their strategy
And one of the cutters we call Hin-d-III

On binding to A-A-G-C-T-T
The site recognition site's bent easily
Phosphodiester attacking meanwhile
Has water behaving as nucleophile

To stave off the phage for a little while

Why don't these enzymes cut cell DNAs?
The answer's provided by their methylase
Adding a methyl group on top of what
The sequence these enzymes would otherwise cut

So cells get protected in this simple way
From nuclease chewing of their DNA
The phage is not lucky in most every case
Unless methylases win the enzyme race

If that happens then, the cell gets erased

Lyrics by Kevin Ahern

No Recording Yet For This Song

I’ve Just Run a Gel

To the tune of “*I Just Saw a Face*”

Metabolic Melodies Website HERE

I’ve just run a gel. I do not think it went too well
I may have used a bit much SDS.
The stacker’s looking like a mess. It’s true
Oh now what will I do?

The protein sample’s my last one. To purify it was not fun
I spent three weekends working late.
The middle lanes aren’t looking great. I’m screwed
Good God what will I do?

Crawling.
I’m almost bawling
The boss is calling to follow through

I just loaded all I’ve got to make this final western blot
My fingers are both crossed for sure
I hope my protein product’s pure. I do
Then my thesis is through

Hating.
All of the waiting
I’m contemplating what I should do

Staining.
My eyes are straining
There’s no complaining. I say ‘wahoo’

‘Cuz it has the band I need
I’ll go and have it scanned to speed
The writing of my thesis and
Proceed onto the post-doct’ral plan
Oh that will be so grand

Pieces make up my thesis.
No more ‘phoresis. The promised land.

Writing so unexciting.
But no more biting. My nails again.

Writing is coinciding.
With reference citing. I’m at the end.

Lyrics by Kevin Ahern

No Recording Yet For This Song

The Proteins Marching One by One

To the tune of "*The Ants Go Marching One by One*"

Metabolic Melodies Website HERE

Oh there's a method you should know that's very huge
It's spinning round and round inside the centrifuge
The supernatant, pellet too
You choose the one that's right for you
And from there we pu-ri-fy
What's inside

To size exclude filtration is the way to go
The beads have pores small proteins can go in you know
The largest ones, they come out fast
The smallest ones eluting last
And the proteins purified
By their size

Electrons power gel e-lec-tro-pho-re-sis
The protein is denatured thanks to SDS
Proteins in a minus state
Get sorted by atomic weight
Smaller ones in speedy mode
To the anode

Ion exchange is special chromatography
To switch cations, you must have a minus bead
Upon this bead, the proteins bind
They're positive, not any kind
And the others wash right through
Out to you

Oh my this song has given you a mighty list
Perhaps we'll just skip over ol' dialysis
So study HPL and C
If you have questions, talk to me
You will get through protein hell
You'll do well.

Recording by David Simmons
Lyrics by Taralyn Tan