Lecture Notes for

Methods in Cell Biology

(TRMD 623)

Instructor:
Mark F. Wiser

TABLE OF CONTENTS

PREFACE—COURSE DESCRIPTION

This course provides students with a broad overview to the basic biochemical, molecular and immunological techniques that are commonly used in modern biomedical research. Lectures will describe the theories and principals behind each of the methods in addition to discussing the practical aspects and limitations in executing the various procedures. One of the course objectives is to assist students with their own research by providing them with sufficient background information so that they are able to design experiments and know which methods are best suited to address a particular research question or problem. A second course objective is to provide students a better access to the scientific literature in that a better understanding of the methods will allow the students to critically evaluate the results and conclusions of scientific papers. Students anticipating careers involving biological or medical research at any level will benefit from this course.

The course consists of three sections: 1) general biochemical and biophysical methods, 2) analysis and isolation of proteins and immunological procedures, and 3) analysis of nucleic acids and recombinant DNA. The first section will cover some basic biochemical procedures and equipment. Understanding these basic biochemical principals will assist in the subsequent discussions on proteins and nucleic acids. The section on characterization of proteins will describe some basic methods used to analyze proteins and provide an overview on protein purification. In addition, the generation of antibodies and their uses in various assays will also be covered. The final section on nucleic acids will describe the basic procedures used in molecular biology including gene cloning, PCR and sequence analysis.

These lecture notes approximately follow the course and are divided into four sections: 1) General Biochemical and Biophysical Methods (Chapters 1-6), 2) Analysis and Characterization of Proteins (Chapters 7-12), 3) Immunological Methods (Chapters 13-15), and 4) Nucleic Acids and Recombinant DNA (Chapters 16-23). Many of the chapters correspond to the lectures.

PART I

General Biochemical and Biophysical Methods

Topics covered:

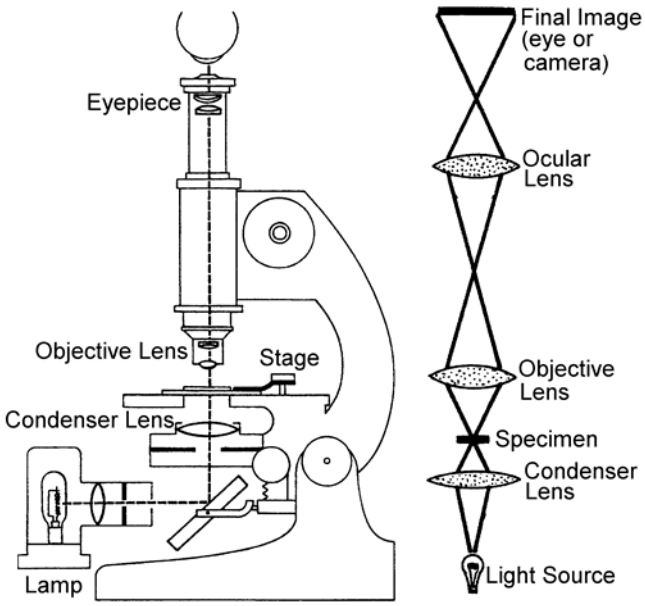
- Microscopy
- Spectrophotometry
- Fluorescence and Flow Cytometry
- Radioactivity
- pH and Buffers
- Centrifugation

CHAPTER 1--MICROSCOPY

Cells are small and in almost all situations a microscope is needed to observe them and their subcellular components. In fact the invention of the microscope led to the discovery and description of cells by Hooke in 1655. The microscope is still an extremely important tool in biological research. The light microscope has a limited capability in regards to the size of a particle that can be examined. The electron microscope provides additional resolution that allows for the examination of subcellular structures and even molecules.

LIGHT MICROSCOPY

The principal of light microscopy is to shine light through a specimen and examine it under magnification. The major optical parts of a microscope are the objective lens, the eyepiece, the condenser and the light source. The **objective lens** functions to magnify the object. The high degree of magnification of the objective lens results in a small focal length and the magnified image actually appears directly behind the objective. The **eyepiece** functions to deliver this image to the eye or camera. Eyepieces also magnify the image, but it is an empty magnification. In other words, the eyepiece enlarges the image but does not increase the ability to see fine details (i.e., the resolution). The **condenser** functions to focus the **light source** on the specimen. The condensor also eliminates stray light and provides an uniform illumination. An iris diaphragm which controls the amount of light reaching the specimen is also associated with the condenser lens. In addition, the light intensity can also be controlled by adjusting the voltage applied to the lamp on some microscopes.



Major components of a light microscope

1. Center light source and all components on optic axis.
2. Focus objective.
3. Focus condenser.
4. Adjust illumination.

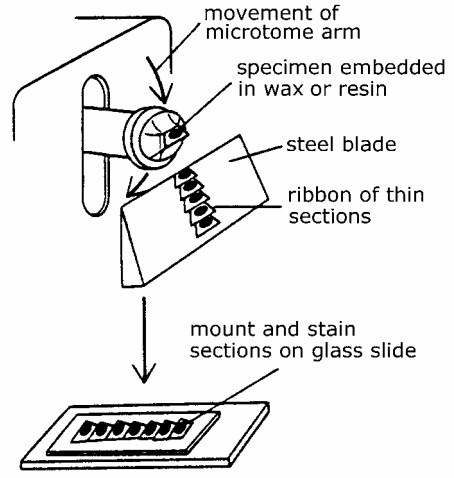
Before using a microscope (Box) it is also important to check that all of the optical components are centered on an optical axis so that the best image and resolution are obtained. Aligning the optical components is usually simple (see instructions manual for the particular microscope) and needs to be done periodically. The specimen is then placed on the stage and the objective lens is focused. The quality of the image produced is highly dependent on the illumination. The position of the condenser lens is adjusted so that the light is focused on the specimen and the intensity of the illumination is adjusted. On better microscopes the illumination can be controlled by both adjusting the diameter of the iris and by adjusting the voltage applied to the lamp. The amount of illumination is important for controlling resolution vs. the contrast and the depth of field (see Box for definitions). Resolution and contrast are antagonistic in that improving one results in a loss of the other. Resolution is increased by

increasing the amount of light. However, the brighter light leads to a loss in contrast. The user must decide upon the optimal mix of contrast and resolution by adjusting both the voltage (i.e., intensity or brightness) of the lamp and the iris diaphragm. The iris diaphragm also has some effect on the depth of field.

Resolution	The ability to discern fine details. Typically expressed as a linear dimension describing the smallest distance needed between 2 objects so that both are seen.
Contrast	Contrast refers to the number of shades in a specimen. More shades decreases the contrast, but increases the amount of information (also called dynamic range).
Depth of Field	Refers to the thickness of the specimen that will be in acceptable focus.

Sample Preparation

Specimens can be examined by simply placing them on a glass microscope slide under a glass cover slip. However, it is usually necessary to prepare and stain the samples before examination by microscopy. **Fixation** is a process by which cells are preserved and stabilized. Common fixatives include: acids, organic solvents, formaldehyde and glutaraldehyde (see Appendix for more discussion about aldehyde fixatives). These treatments affix macromolecules in position. For example, glutaraldehyde chemically cross-links the primary amines of neighboring proteins and organic solvents precipitate proteins and other macromolecules.



Thick samples, such as tissues, will need to cut into thin sections. Following fixation the sample or cells are embedded into a supporting medium. Paraffin is a common embedding medium for light microscopy as well as various plastic resins. **Sectioning** is carried out with a microtome (Figure). The microtome cuts the specimen into thin slices, or sections, of a specified thickness. It is also possible to collect the successive slices, called serial sections, and therefore ascertain the three dimensional aspects of the tissue or specimen being examined.

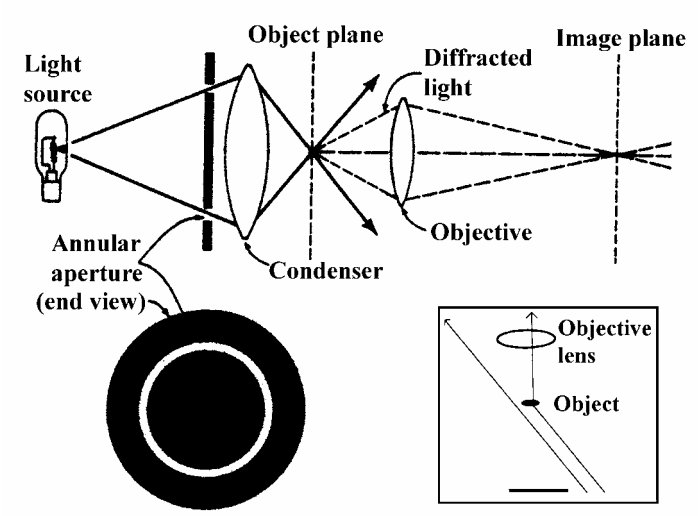
The image generated by microscopy depends upon different components in the sample interacting with and impeding the light waves differentially. Biological samples are fairly homogeneous (i.e., carbon-based polymers) and do not greatly impede light. Therefore, it is often necessary to **stain** cells with dyes to provide more contrast. Different dyes have different affinities for different subcellular components. For example, many dyes specifically interact with nucleic acids (i.e., DNA and RNA) and will differentially stain the cytoplasm and nucleus. These stained subcellular components will differentially absorb the light waves and result in less light reaching the eyes or camera, and thus appears darker. Furthermore, since the dyes only absorb certain wavelengths of light, the

various structures within the specimen will exhibit different colors. (See chapter on Spectrophotometry for a more extensive discussion of chromaphores and light absorption.)

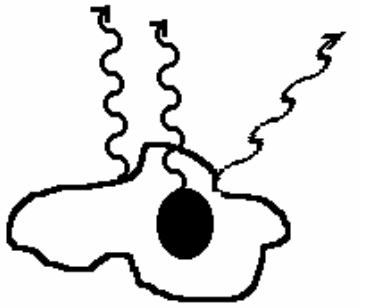
- Dark Field
- Phase Contrast
- Differential Interference Contrast (or Normarski)
- Confocal Scanning
- Fluorescence
- Image Enhancement

Variations to bright field (transmission) microscopy

Many modifications of light microscopy that have specialized applications have been developed (Box). In **dark-field** microscopy the specimen is illuminated from the side and only scattered light enters the objective lens which results in bright objects against dark background. This is accomplished through the use of an annular aperture that will produce a hollow cone of light that does not enter the objective lens (see Figure). Some of the light hitting objects within the specimen will be diffracted into the objective lens (see Figure Inset). The images produced by dark-field microscopy are low resolution and details cannot be seen. Dark-field microscopy is especially useful for visualization of small particles such as bacteria.



Dark Field Microscopy Optics



Phase Shift vs Diffraction

Both **phase contrast** microscopy and **differential-interference-contrast** allow objects that differ slightly in refractive index or thickness to be distinguished within unstained or living cells. Differences in the thickness or refractive index within the specimen result in a differential retardation of light which shifts the phase or deviates the direction of the light (Figure). During phase contrast microscopy the phase differences are converted to intensity differences by special objectives and condensers. Normarski optics use special condensers and objectives to recombine incident and diffracted light waves from a single source at the plane of the image. In both methods the interference effects between the incident and diffracted light enhance small differences in the refractive index or thickness of the specimen and leads to an increased resolution without staining.

In **fluorescence microscopy** a fluorochrome is excited with ultraviolet light and the resulting visible fluorescence is viewed. This produces a bright image in a dark background.

Confocal microscopy uses the objective lens as both the objective and the condenser. This allows the illuminating light to be focused on a relatively thin plane. In addition, a 'pin-hole' is used to further minimize the light coming from other planes. Minimizing the interference from other planes increases apparent resolution. (Fluorescence and confocal microscopy are discussed in greater detail in the chapter on Fluorescence.)

Video cameras and image processing have had a major impact on microscopy. Images are digitized and can be manipulated electronically. This can correct imperfections in optical systems and can overcome limitations of human eye. In particular, the human eye is not very effective in dim light and cannot distinguish small differences in intensity against a bright background. Image enhancement can remedy both of these limitations. However, image enhancement cannot increase the resolution. This is due in part to the limit of resolution which is determined by the wavelength (λ) of visible light (see Box).

The theoretical **limit of resolution** is defined as 0.61λ/N.A., where N.A. (numerical aperture) is a property of the objective lens determined by its magnification, diameter and refractive index. Typical ranges for the N.A. are 0.25-1.32. Visible light has an average wavelength of approximately 0.5 μm making the maximum limit of resolution approximately 0.2 μm. Mitochondria are about the smallest subcellular structures that theoretically can be seen. No amount of refinement of the optical systems can overcome this physical barrier, even though the image can be enlarged indefinitely. In addition, the practical limit of resolution will be less than the theoretical limit of resolution due to optical aberrations in the lenses (see Table).

Typical Limits of Resolution for Common Objective Lenses

Objective Magnification	N.A.	Theoretical (μm)	Practical (μm)
4X	0.10	3.05	3.4
10X	0.25	1.22	1.3
40X	0.65	0.47	0.52
100X	1.30	0.24	0.26

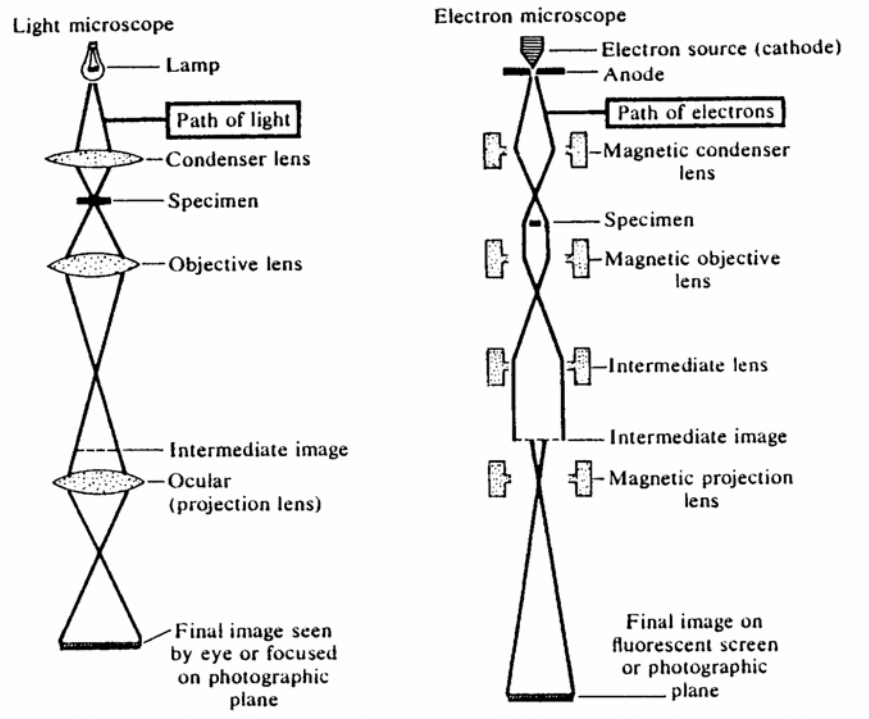
ELECTRON MICROSCOPY

The relationship between the limit of resolution and the wavelength of the illumination holds true for any form of radiation. Thus resolution can be increased in theory by using radiation of lower wavelengths. However, the human eye is only capable of detecting radiation with wavelengths in the range of 0.4-0.8 μm, or the visible spectrum. These problems associated with the limit of resolution have been overcome by using electrons to generate an image of the specimen. Particles, such as electrons, travelling near the speed of light behave as a wave (i.e., radiation) and their effective wavelength is inversely proportional to electron's velocity. Therefore increased resolution can be achieved by examining a specimen with high velocity electrons.

The general principal of electron microscopy is analogous to light microscopy (Figure) except that electrons are used to analyze the specimen instead of visible light. The illumination source is a white-hot tungsten filament, which emits high velocity electrons. The electron beam is focused by a condenser lens onto the specimen. The condenser lens, however, is an electromagnet instead of a glass. These electrons are differentially impeded by the various structures within the specimen. In other words, some of the electrons are scattered or absorbed by the atoms of the specimen. The electrons which pass through the specimen are focused with a series

of magnetic objective lens onto either a photographic plate or a fluorescent screen. The electrons interact with the photographic plate or fluorescent screen as if they were photons (i.e., light) and generate an image. The differential loss of electrons due to the substractive action of the sample will generate an image in much the same way as the absorption of light creates an image in light microscopy.

Comparison of Microscope Optics



Sample preparation

1. Fixation
2. Dehydration
3. Embedding
4. Sectioning
5. Staining

It is not possible to view living material with an electron microscope. Biological samples are usually fixed with glutaraldehyde, which cross-links proteins (see Appendix), and treated with osmium tetroxide, which stabilizes lipid bilayers and proteins. Osmium tetroxide is reduced by many organic compounds, especially lipids, which results in cross-linking. Since electrons have very little penetrating power, samples must be embedded in special plastic resins and cut into thin sections of 0.05-0.1 μm. Removing all water from the specimen is necessary for the proper polymerization of the plastic resin. Following fixation the samples are dehydrated by exposing them to series of increasing alcohol concentrations until reaching 100%. The dehydrated sample is then put into a solution containing monomers of the embedding resin and polymerization is induced. This 'block' containing the sample is sectioned with the ultramicrotome and the ultrathin sections are place onto copper or nickel grids coated with a thin carbon or plastic film for support.

Contrast in electron microscopy is dependent upon atomic number of the atoms in

the sample. Biological materials, primarily made of carbon, exhibit low atomic number and exhibit a similar electron scattering as the carbon films on the support grid. To obtain more contrast, samples are stained with salts of heavy metals, such as osmium, uranium and lead. Staining can be carried out before the dehydration and embedding or after sectioning. Different cellular compartments and structures stain differently with the heavy metals.

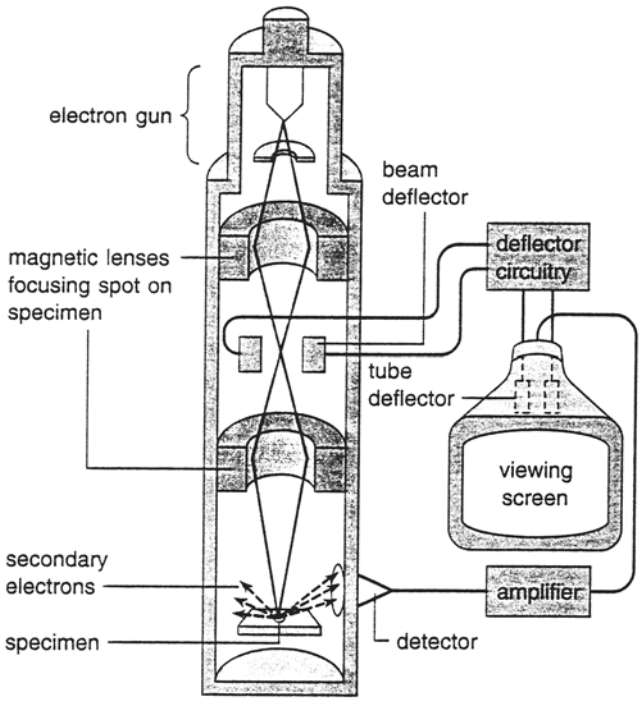
Electron microscopes are expensive instruments and require substantial training to operate. Electron microscopy will usually require collaboration with someone having expertise in electron microscopy. Many universities have shared instrument facilities in which users pay a fee that includes use of the instrument and technical assistance. Typically fixed samples are provided to the electron microscopy service for further processing and analysis.

Variations in Electron Microscopy

- Transmission (TEM)
- Scanning (SEM)
- Shadow-casting
- Freeze-fracture
- Freeze-etching
- Negative staining
- cryoEM

The standard form of electron microscopy involves shooting an electron beam through the sample. This is called **transmission electron microscopy**, often abbreviated TEM. **Scanning electron microscopy** (SEM) detects the electrons that are scattered by the specimen to form a 3-dimensional image. The sample is fixed and coated with a thin layer of a heavy metal such as platinum to form a replica of the specimen. This replica is then scanned with a thin beam of electrons and the quanitity of electrons scattered along each successive point of the specimen is measured by detectors which surround the sample (see Figure). Since the amount of electron scattering depends on the angle and depth of the surface relative to the beam, the image has highlights and shadows that give it a three dimensional appearance. The resolution of SEM is not very high (approximately 10 nm with an effective magnification of up to 20,000 times) and only suface features can be examined. Therefore, the technique is generally used to study whole cells or tissues.

A 3-dimensional appearance with higher resolutions than SEM can be obtained by TEM by **shadowing**. In this case the metal coating is applied at an angle resulting in a replica reflects the height and depth of the specimen. Shadowing is often used in conjunction with other techniques. For example, in **freeze-fracture** and **freeze-etching** cells are frozen in cryoprotectant and cut with a knife. Freeze-fracture will often split the lipid bilayer membranes which are then shadowed with platinum. Alternatively, in freeze-etching, the water is sublimated and replicas formed.



Scanning Electron Microscope

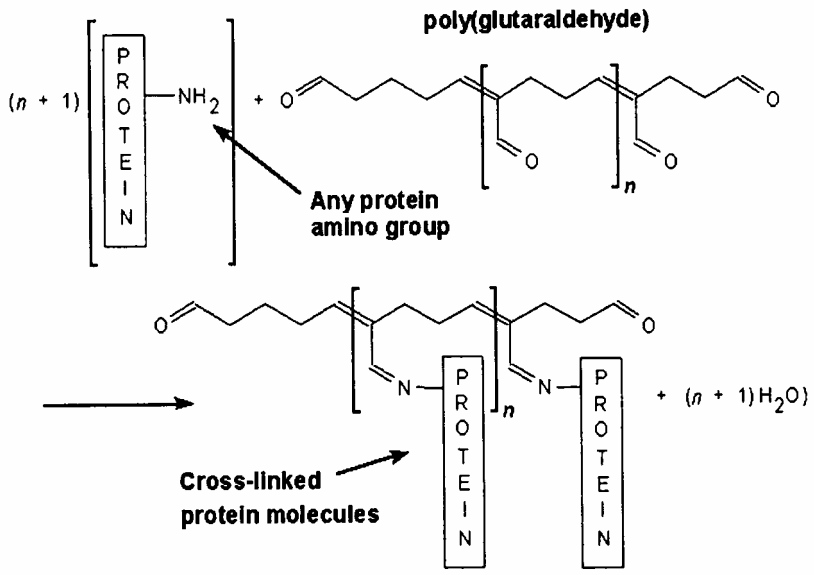
Negative staining can be used to visualize macromolecules and supramolecular structures such as virus particles or cytoskeletal filaments. The samples are placed on the electron transparent carbon grids and stained with heavy metals. Areas with biological structures appear more electron transparent.

The fixation and manipulation of the specimen will often distort cells. Cryo-electron microscopy is used to overcome this problem. Special holders, which keep hydrated specimen at -160oC, allows viewing without fixation, staining or dehydration.

MICROSCOPY APPENDIX. ALDEHYDE FIXATIVES

Formaldehyde and glutaraldehyde are a commonly used fixatives for both light and electron microscopy. Formaldehyde is a small molecule (HCHO). The formaldehyde monomers form polymers in aqueous solutions. The liquid known as formalin is 37-40% formaldehyde by weight and most of the polymers are 2-8 units long. Higher polymers (n up to 100) are insoluble in water and sold as a white powder called paraformaldehyde. To be useful as a fixative, the solution must contain monomeric formaldehyde. Dilution of formalin with a buffer at physiological pH results in an almost instantaneous formation of monomers. Conversion of paraformaldehyde to monomers requires heat (typically 60^{o}C) and the addition of hydroxide ions. Commercial formalin contains about 10% methanol and small amounts of formate ions, whereas a formaldehyde solution prepared from paraform-aldehye initially does contain any methanol or formate.

Formaldehyde's mechanism of action is based on the reaction of the aldehyde group with primary amines in proteins. A cross-link between neighboring proteins can also be formed if the primary amines are close enough together. The initial reaction of formaldehyde with protein is complete within 24 hours, but the formation of cross-links, called methylene bridges, proceeds much more slowly (several weeks). Lipids, nucleic acids and carbohydrates are trapped in a matrix of insoluble and cross-linked proteins. In practical terms, formaldehyde penetrates tissues rapidly (because of its small size), but it slowly cross-links the proteins.



Glutaraldehyde contains two aldehyde groups separated by three methylene bridges (HCO-[CH_2]$_3$-CHO). These two aldehyde groups and the flexible methylene bridge greatly increases the cross-linking potential of glutaraldehyde over formaldehyde. In solution glutaraldehyde exists as polymers of various sizes which exhibit an enormous potential for cross-linking proteins (Figure). In contrast with formaldehyde, the chemical reaction of glutaraldehyde with protein is fast, but the penetration of tissue is slower, especially for the

larger oligomers. Therefore, an 'EM grade' glutaraldehyde, which contains low polymers, should be used for fixation. In addition, fixation with glutaraldehyde results many leftover free aldehyde groups which cannot be washed out of the tissue. For many applications these free aldehyde groups needs need to be removed or blocked. A common blocking method is to treat with glycine or another small primary amine.

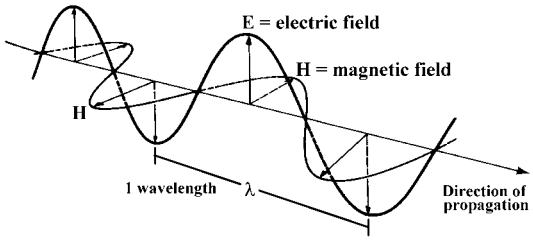
The combination of formaldehyde and glutaraldehyde is also used as a fixative for electron microscopy. This takes advantage of the rapid penetration of formaldehyde molecules, which quickly stabilizes the structure of the tissue, followed by a more thorough cross-linking of proteins mediated by the more slowly penetrating glutaraldehyde.

CHAPTER 2--SPECTROPHOTOMETRY

Spectrophotometry is a versatile analytical tool. The underlying principle of spectrophotometry is to shine light on a sample and to analyze how the sample affects the light. Advantages of spectrophotometry are: 1) it is often non-destructive (i.e., can measure and recover sample), 2) it is selective (often a particular compound in a mixture can be measured without separation techniques), 3) it has a short time interval of measurement (10^{-14} seconds).

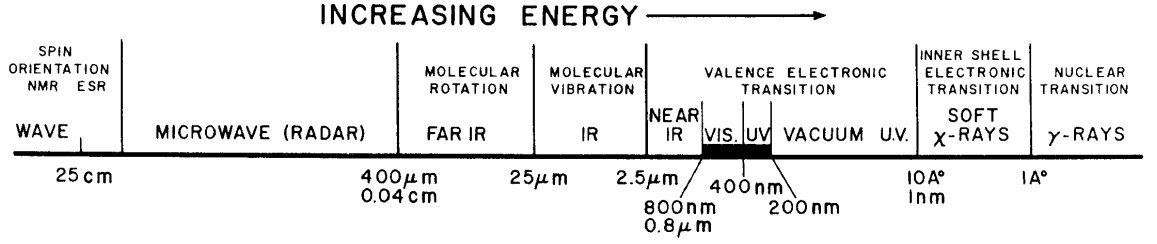
SPECTROPHOTOMETRY THEORY

Light can be described as a wave. This wave has an electric component and a magnetic component which are perpendicular to each other (Figure). Electromagnetic radiation exhibits a direction of propagation and wave-like properties (i.e., oscillations). The energy of electromagnetic radiation is defined as:

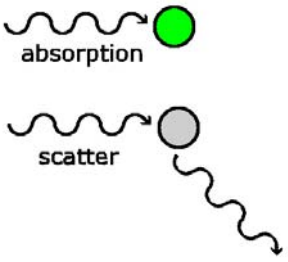


$$E = hc/\lambda = h\upsilon$$

where E = energy, h = Planck's constant, c = the speed of light, λ = the wave length, and υ = frequency. Light behaves both as a wave and as a particle. The conceptual particle of light is called a **photon** and is represented by $h\upsilon$. Electromagnetic radiation exhibits a wide spectrum and specific ranges of wavelengths have names (Figure). The energy of electromagnetic radiation is inversely proportional to its wavelength.



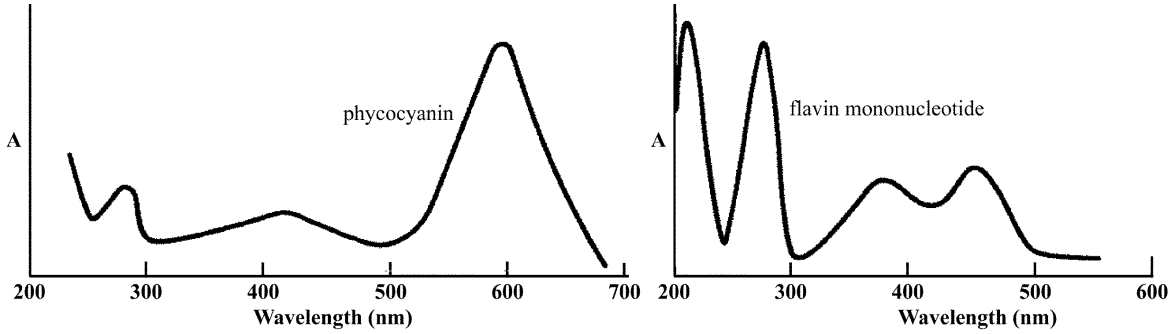
When a light wave encounters a particle, or molecule, it can be scattered (i.e., direction changed), absorbed (energy transferred), or unaffected. Molecules only absorb discreet packets of energy, or quanta, and absorption occurs when the energy of the photon corresponds to differences between energy levels in that particular molecule. These discrete energy levels, called electronic energy levels, are a property of the particular molecule and are determined by the spatial distribution of the electrons.



Absorption of the energy from the photon elevates the molecule to an excited electronic state (see Figure in Chapter 3) by causing an electron to move from one orbit to another. These electronic energy levels are further subdivided into vibrational levels. The vibrational levels correspond to stretching and bending of various covalent bonds. The transitions to the excited

state can occur between different vibrational levels giving a range of energy that can be absorbed by the molecule.

A molecule or substance that absorbs light is called a **chromophore**. Chromophores exhibit unique absorption spectra (Figure) and can be defined by a wavelength of maximum absorption, or λ_{max}, of a broad absorbtion band due to the vibrational levels. The absorption spectra can consists of several absorption maxima of various amplitudes. A large number of biological molecules absorb light in the visible and ultraviolet (UV) range.



The net affect of absorption is that the intensity of the light decreases as it passes through a solution containing a chromophore. The amount of light absorbed depends on the nature of the chromophore, the concentration of the chromophore, the thickness of the sample, and the conditions (eg., pH, solvent, etc.) under which absorption is measured. Absorption is governed by the Beer-Lambert Law:

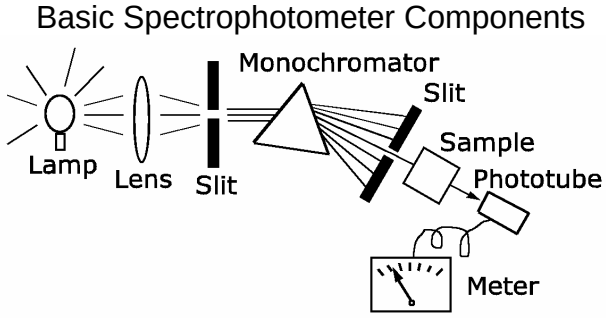
$$I = I_o 10^{-\varepsilon dc} \quad \text{or} \quad \log(I/I_o) = -\varepsilon dc$$

where I = final light intensity, I_o = initial light intensity, ε = molar extinction coefficient, d = thickness, and c = molar concentration. Absorption (A) will be defined by:

$$A = -\log(I/I_o) = \varepsilon dc$$

The molar extinction coefficient (ε) is defined as the A of 1 M of pure compound under standard conditions and reflects something about the nature of the chromaphore. The units of ε are liter/cm·mole. However, the extinction coefficient can be expressed in other units. For example, it can be expressed in terms of mM concentration. The thickness of the sample (d) is almost always 1 cm and therefore can be ignored in calculations. Sometimes, though, the extinction coefficient units is expressed in cm^2/mole (by converting liters to cubic centimeters) and care should be taken in making calculations. In cases where the molecular weight of the substance is not known, or varies, $E^{1\%}$ is used as the extinction coefficient. $E^{1\%}$ is defined as the A of a 1% (w/v) solution. It is important to precisely record the units of ε when looking it up or determining it experimentally since these units will determine the concentration. It is also important to record the conditions (eg., pH, solvent, temperature, etc.) for an extinction coefficient (see below).

INSTRUMENTATION



Spectrophotometers produce monochromatic light and then accurately measure the light intensity. The major components of a spectrophotometer are the light source, a monochromator, sample holder, a light detector (phototube), and a meter. In most instruments a tungsten lamp is used for the visible range and either high pressure H_2 or D_2 lamps are used for UV range. Monochromatic light is generated by either 1) a movable prism, 2) a diffraction gradient, or 3) filters. Monochromatic light is projected through the sample and then measured by a photomultiplier tube. A photomultiplier tube converts the energy of the light photons into electrons (see appendix). The voltage resulting from these electrons is measured by a meter and converted to an absorbance value. The I_o (initial intensity) is determined by calibrating the instrument with a 'buffer blank'. The relative difference in the light intensity between the blank and the sample is then expressed as the absorbance (A). Spectrophotometers often include accessories such as chart recorders or microprocessors for data analysis.

APPLICATIONS OF SPECTROPHOTOMETRY

Determining the concentration of substances in solution is the most common use of the spectrophotometer. Exact concentrations can be determined in cases where ε are known and the measurement is carried out under the prescribed conditions. The substance being measured does not necessarily need to absorb radiation if it can scatter radiation. For example, measuring the A_{600} is a quick and easy way to monitor bacterial growth and determine the number of bacteria in cultures. In addition, since compounds exhibit unique absorption profiles, spectrophotometry can also be used to identify unknown compounds.

Spectrophotometry is also a convenient method to measure enzyme activity in cases where the substrate and the product exhibit different λ_{max}. Either the disappearance of substrate or the appearance of product over time is measured. The change in the A_λ per unit time (generally per minute) is calculated. The change in A_λ of a blank (= identical sample without enzyme source) is subtracted from this value. The enzyme activity in terms of amount of product formed per unit time per mg protein can be calculated by factoring in the amount of enzyme, dilution factors and the extinction coefficient (see example in appendix). A typical example of a formula for the calculation of enzyme activity is:

$$\mathbf{activity = (\Delta A_{sample} - \Delta A_{blank})\cdot volume\cdot 10^6/\varepsilon\cdot \Delta time\cdot mg\ protein}$$

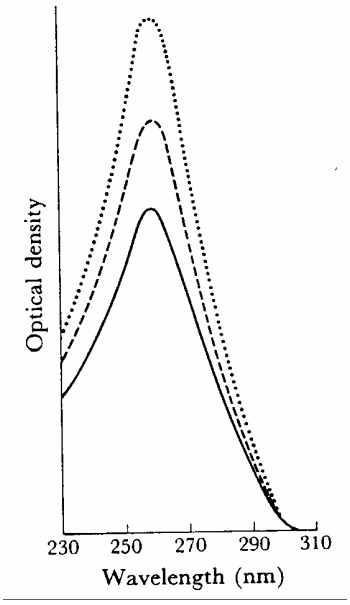
where activity is expressed as μmoles product formed/min/mg protein; ΔA_{sample} is the change in absorbance of the sample containing enzyme; ΔA_{blank} is the change in absorbance of a sample containing everything except the enzyme; the volume in the cuvette expressed in the same units as ε; 10^6 μmoles per mole (assuming ε is expressed in moles); ε is the molar extinction coefficient; Δtime is the time in minutes the reaction was measured; and mg protein in the

cuvette. In deriving such formulas it is important to match the units. The units of ε may also include the 1 cm thickness of the cuvette which is ignored in the calculations.

FACTORS AFFECTING ABSORPTION

Although the absorption spectrum is primarily determined by the chemical structure of the chromophore, environmental factors can affect λ_{max} and ε. The pH determines the ionization of chromophore which in many cases will affect the absorption properties of a chromophore. Indicator dyes and pH paper are examples of this phenomenon.

The polarity of the solvent or neighboring molecules can also affect absorption. Because of this effect, spectrophotometry can also be used to determine structural features of macromolecules. For example, whether particular amino acid residues are buried within the internal portion of proteins or exposed on the aqueous solvent at the protein surface can be determined by spectrophotometry.



Spectra of T7 DNA

The relative orientation of neighboring chromophores also affects absorption. Hypochromism of nucleic acids (Figure) is an example of orientation effects. The absorbance (i.e., ε) decreases as free nucleotides (dotted line) are polymerized into single-stranded DNA (dashed line) and decreases further in double-stranded DNA (solid line).

VARIATIONS IN SPECTROPHOTOMETRY

It is also possible to use other spectra besides UV/visible range. For example, vibrations between the atoms of molecules can be analyzed using infrared (IR) and Raman spectrophotometries. Many molecules will exhibit characteristic 'signatures' and therefore can be identified.

Other Forms of Spectrophotometry

Spectrophotometry	**Comment**
Infrared (IR)	vibrational levels
Raman	
Optical Rotary Dispersion (ORD)	polarize light
Circular Dichroism (CD)	
Nuclear Magnetic Resonance (NMR)	magnetic moments
Electron Spin Resonance (ESR)	

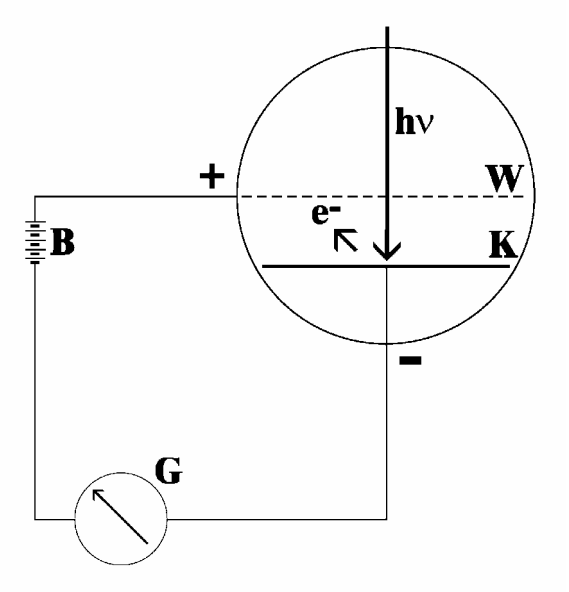
Light can be polarized so that all of the waves are in the same orientation. The study of the absorption of polarized light can yield more information about the structure of molecules if the chromophores have optically active centers. Circular dichroism (CD) measures the ability of chromophores to differentially absorb left and right circularly polarized light. Optical rotary dispersion (ORD) measures the ability of a optically active chromophore to rotate plane polarized light. Both CD and ORD are useful in structural studies or proteins and nucleic acids. For example, it is possible to approximate the amount of α-helix, β-sheet and random coil in

proteins.

The effects of molecules on the magnetic component of radiation can also be analyzed. Nuclear magnetic resonance (NMR) is a spectroscopic method capable of yielding information on the structure of molecules, interactions between molecules and molecular motion. This method is based upon the principle that a spinning charge (i.e., the nucleus) generates a magnetic field. Similarly, an electron also possesses a spin magnetic moment which can be analyzed by electron spin resonance (ESR). A common use for ESR in biological sciences is to monitor the fluidity of membranes.

APPENDIX I. PHOTOMULTIPLIER TUBE

Photomultiplier tubes, also called photoelectric cells, convert light intensity into an electrical current. The basic principle of a photomultiplier tube is the ejection of electrons from a metal surface and the measurement of the resulting current (Figure). A receiver (K) is coated with potassium or other metal in a highly evacuated tube. The tube also contains a wire screen (W) which is connected to the receiver through a battery (B) and a sensitive galvanometer (G). Light hitting the receiver ejects ejects electrons which are attracted to the positive charge of the screen. The resulting current is directly proportional to the number of electrons ejected per unit time, which is proportional to the light intensity. In other words, the energy of the photon (hν) is transferred to a single electron within the metal. If this energy is sufficeintly large then the electron will be ejected from the surface of the metal and retain some kinetic energy which depends upon the energy of the photon that ejected it. The number of electrons ejected will depend on the number of photons (i.e., light intensity). This is known as the photoelectric effect.

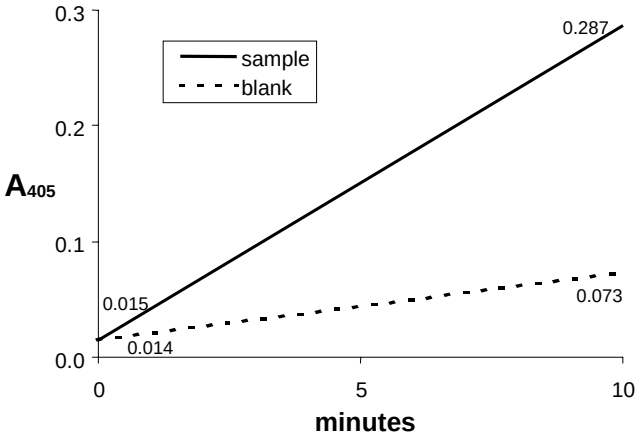


APPENDIX II. CALCULATION OF ENZYME ACTIVITY

Acetylcholine esterase (ACHE) removes the acetate group from choline. This activity can be measured with the non-natural substrate acetylthiocholine. ACHE cleaves the sulfide bond producing a free thiol (thiocholine) and acetate. This free thiol will spontaneously react with dithionitrobenzoate (DTNB) to produce 5-thionitrobenzoic acid (5TNB), which has a red color. Therefore ACHE activity can be measured using the following coupled reaction:

acetylthiocholine $\rightarrow$ thiocholine + acetate (catalyzed by ACHE)

thiocholine + DTNB $\rightarrow$ 5TNB + oxidized thiocholine (occurs spontaneously)



You have a sample with a protein concentration of 0.34 mg/ml and would like to calculate the activity of ACHE (units = μmoles product formed/ min·mg protein). Thirty μl of the sample are added to a cuvette in final volume of 1 ml containing all the necessary substrates. The A_{405} is plotted over time as depicted by the figure on the right. The A_{405} increases from 0.015 to 0.287 in 10 minutes. The A_{405} of a blank (containing all the same reagents except the protein sample) goes from 0.014 to 0.073 in 10 minutes. The ε_{405} for 5TNB is 13.3×10^3 liter/cm·mole

Calculations and Explanations

The following formula (see derivation below) can be used to calculate the enzyme activity:

$(\Delta A_{sample} - \Delta A_{blank}) \cdot$cuvette vol.$\cdot 10^6/\varepsilon \cdot \Delta$time$\cdot$sample vol.$\cdot$protein conc.

ΔA_{sample} This represents the change in the absorbance from the beginning to the end of the measurement period. Obtained by subtracting 0.015 from 0.287.

ΔA_{blank} Some DTNB will spontaneously convert to 5TNB and this needs to be substracted from sample values. Obtained by subtracting 0.014 from 0.073.

cuvette vol Absorbance is proportional to concentration and must be multiplied by the volume to calculate the absolute number of moles product formed. The units must match ε; liters in this example.

10^6 This converts the moles of ε to μmoles. The conversion factor will change according to the desired final units and units of ε.

ε — The molar extinction coefficient converts absorbance values to concentrations. ε represents an inherent property of the chromophore and is usually looked up. It is important that the conditions of assay are similar to the conditions used to determine ε. The thickness unit (cm) can be ignored since the standard cuvette size is 1 cm.

time — Enzyme activity is expressed per unit time (usually minutes) and this represents the time interval the absorbance was measured. If the times for the samples and blank do not match then it will be necessary to calculate ($\Delta A_{sample}/\Delta time - \Delta A_{blank}/\Delta time$).

sample vol — All of the parameters above represent the activity in the cuvette. Therefore it is necessary to divide by the sample volume so that the activity is expressed per unit volume of the sample.

protein — Dividing by the protein concentration will provide information about the amount of enzyme activity per unit protein (i.e., specific activity).

Plugging the values:

(0.272 - 0.059)(0.001 L)(10^6 μmole/mole)/(13.3x10^3 L/~~cm~~·mole)(10 min)(0.03 ml)(0.34 mg/ml) =

0.16 μmoles thiocholine formed per min per mg protein

Example of formula derivation.

There are several strategies and formulas that can be used to calculate enzyme activities. All of them start with the Beer-Lambert Law:

$$\mathbf{A = \varepsilon dc}$$

rearranging the equation to solve for concentration and adding the thickness results in:

$$\mathbf{c = A/\varepsilon(1\ cm)}$$

The 1 cm will not affect the calculation and can be ignored as long as the units of ε are corrected to account for this. Measuring enzyme activity involves determining the change in concentration over time. In addition, the spontaneous formation of product (blank without enzyme) needs to be subtracted from the sample resulting in:

$$\mathbf{\Delta c/\Delta time = (\Delta A_{sample}/\Delta time - \Delta A_{blank}/\Delta time)/\varepsilon}$$

If the sample and blank are measured for the same time periods the equation can be convert

to:

$$\Delta c/\Delta t = (\Delta A_s - \Delta A_b)/\varepsilon \cdot \Delta t$$

Multiplying both sides of the equation by the cuvette volume (same units as ε) and 10^6 μmoles/mole (assuming ε is expressed in moles) will result in:

$$\mu\text{mole product formed/min} = (\Delta A_s - \Delta A_b)(\text{vol.})10^6/\varepsilon \cdot \Delta t$$

This formula will represent the total enzyme activity in the cuvette. Dividing by the volume of sample used in the assay will give the amount of enzyme activity per unit volume of the original sample:

$$\mu\text{mole/min} \cdot \text{ml} = (\Delta A_s - \Delta A_b)(\text{vol.})10^6/\varepsilon \cdot \Delta t \cdot (\text{sample vol.})$$

Alternatively enzyme activity is often expressed per unit of protein (or specific activity) and therefore needs to be divided by the amount of total protein added to the cuvette instead of the sample volume, or:

$$\mu\text{mole/min} \cdot \text{mg protein} = (\Delta A_s - \Delta A_b)(\text{vol.})10^6/\varepsilon \cdot \Delta t \cdot (\text{mg prot.})$$

The mg protein can be calculated by multiplying the protein concentration (in mg/ml) by the sample volume (in ml).

Additional Practice Problem

Dehydrogenase activities are easily measured in spectrophotometric assays because of strong absorbance of reduced flavin nucleotides (NADH and NADPH) at 340 nm as compared to the oxidized nucleotides (NAD^+, $NADP^+$). The ε_{340} of either NADH or NADPH is 6.22 $x10^3$ liter/cm·mole.

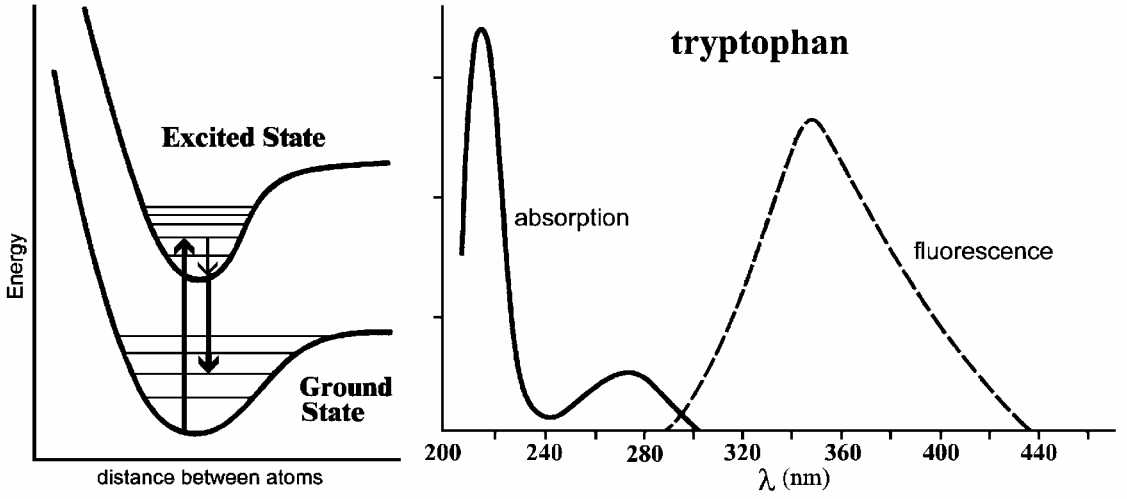
Calculate the amount of glutamate dehydrogenase (GDH) activity expressed in μmoles glutamate formed per min per mg protein using the following reaction parameters: the protein concentration in the sample is 0.23 mg/ml; 0.01 ml of sample was measured in a final volume of 1 ml; the A_{340} decreased from 0.60 to 0.52 in 10 minutes; and the A_{340} of a blank (containing no enzyme) decrease from 0.60 to 0.59 in 10 minutes. GDH catalyzes the following reaction:

$$\alpha\text{-ketoglutarate} + NH_4^+ + \text{NADPH} \leftrightarrow \text{glutamate} + NADP^+$$

[Answer is 0.49 μmoles glutamate/min·mg protein]

CHAPTER 3--FLUORESCENCE

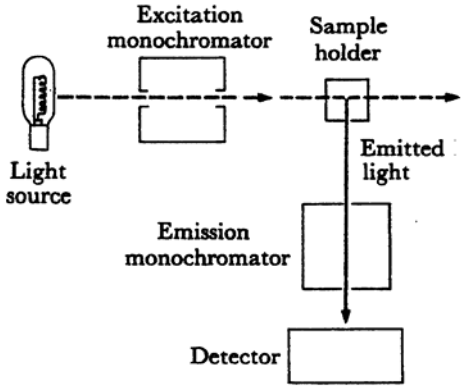
Molecules absorbing the energy of electromagnetic radiation (i.e., photons) will be elevated to a higher energy level, or excited state. These excited molecules will return to the ground state and some molecules emit radiation on their return to the ground state. This phenomenon is known as fluorescence and fluorescent molecules are known as **fluorochromes**. Fluorochromes have distinct absorption spectra as well as emission spectra. The wavelengths of the emitted radiation are longer than the absorbed wavelengths (i.e., lower energy).



Fluorescence is widely used in the biological research. In particular, many molecules absorb in the ultraviolet range (UV) and emit radiation in the visible range. These fluorochromes can be used as sensitive probes to investigate biological phenomenon and are utilized in many different applications (Box).

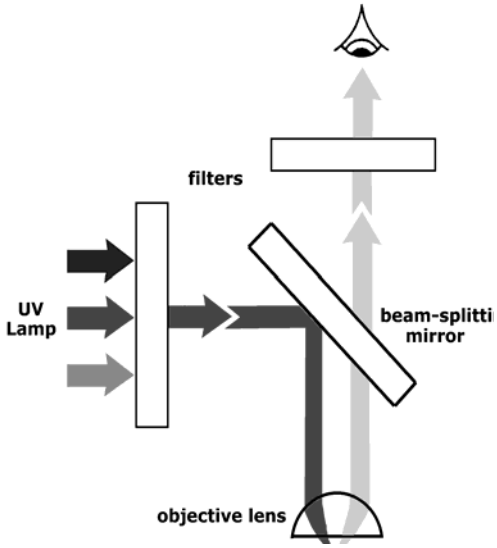
- enzyme assays
- detection of nucleic acids (gels)
- microscopy
- flow cytometry

ENZYME ASSAYS AND FLUOROMETERS



Enzyme substrates are available in which either the product or the substrate is fluorescent. As in the case of spectrophotometry, this provides a convenient method to either follow the appearance of product or the disappearance of the substrate. In general fluorescent substrates provide greater sensitivity than conventional spectrophotometric assays. The amount of fluorescence is measured using a fluorometer. The fluorometer is similar to a spectrophotometer, except that the photomultiplier tube to detect the emitted light is located at a right angle from the path of the excitation light (Figure). The sample is exposed to UV light of the desired wavelength and emitted light of the appropriate wavelength range is detected by a photomultiplier tube. The intensity of the emitted light is quantified and usually expressed as relative fluorescence units.

FLUORESCENCE MICROSCOPY



Since many fluorescent molecules emit light in the visible range it is possible to view fluorescence in conjunction with microscopy. An UV light source is used to illuminate the sample through the objective lens using a beam-splitting mirror. The fluorescence emitted from the sample (**epifluorescence)** passes through this same mirror, but the UV light does not. Filters before the beam-splitting mirror will control the excitation wavelength and filters before the eyepiece (or camera) will control the wavelength of the emitted light.

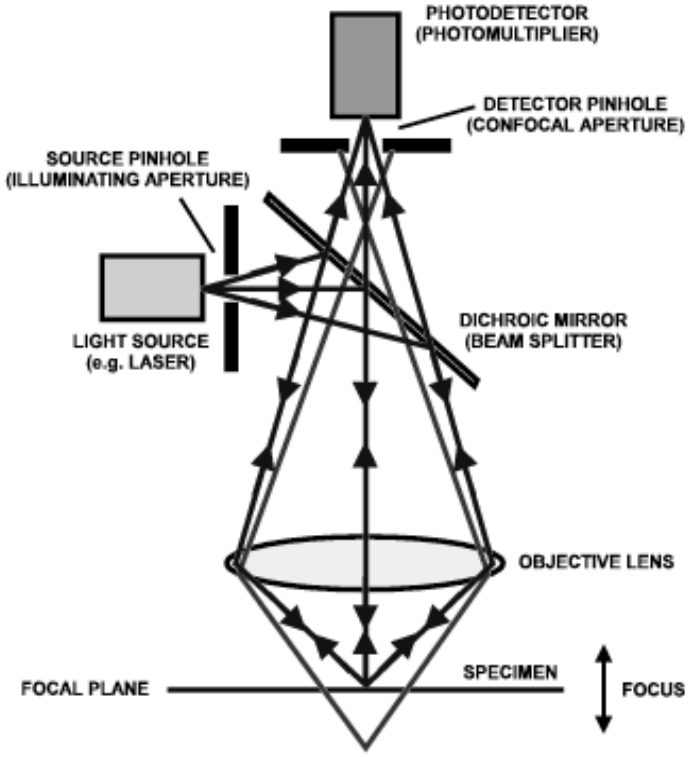
- ethidium bromide (DNA/RNA)
- DAPI (only DNA)
- rhodamine 123 (mitochondria)
- C_5-CMB-ceramide (Golgi)
- free Ca^{2+} indicators
- pH indicators
- membrane potential indicators
- antibodies against specific proteins

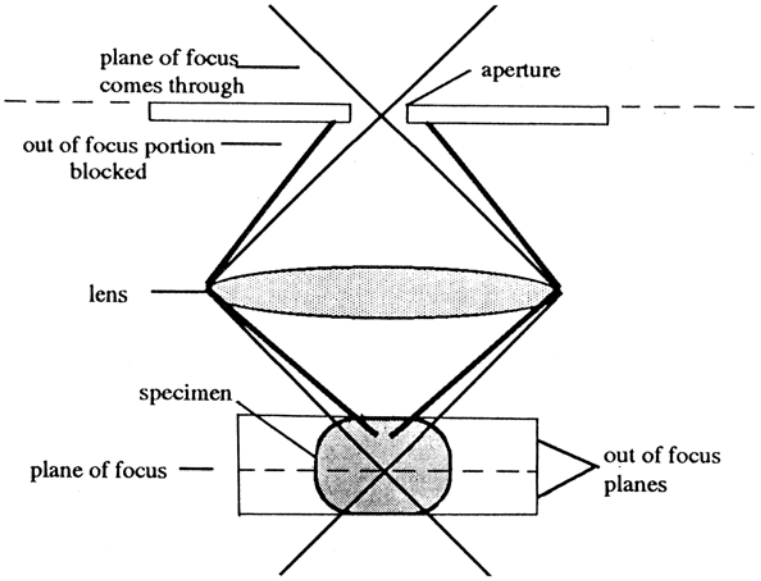
Fluorescent probes can be used to identify individual cells or subcellular components (Box). The location of fluorochromes associated with cells will appear as light objects against a dark background. Immunofluorescence (discussed in greater detail in immunoassays) uses antibodies with a conjugated fluorescent molecule (eg., fluorescein or rhodamine) to determine the location of a protein within a cell. Numerous fluorescent dyes which bind to specific subcellular components are also available. For example, acridine orange fluoresces when it binds to either RNA and DNA. Acridine orange bound to RNA fluoresces a yellow to green color and acridine orange bound to DNA fluoresces orange to red. Other fluorescent probes which bind to specific subcellular compartments are also available. In addition, there are many fluorescent probes available that are sensitive to pH, divalent cations such as Ca^{2+}, and membrane potentials. Such probes permit physiological measurements in individual living cells.

Dual-labeling experiments in which fluorochromes with different emission spectra are co-incubated with the sample can also be carried out. The different fluorochromes will appear as different colors due to the differences in their emission spectra. If the emission spectra are far enough apart then it is possible to use filters so that the two fluorochromes are examined separately. Dual-labeling experiments are especially useful for determining whether two proteins (or other substances) are localized to the same subcellular compartment.

Confocal scanning laser microscopy (CSLM). The confocal microscope increases the resolution, as compared to conventional fluorescent microscopy, by minimizing the light from out of focus planes. In CSLM the illuminating light, a laser, is focused onto the specimen by the objective lens. Light reflected back from the illuminated area is collected by the objective and directed through a pinhole placed in front of the detector. This diameter of the pinhole can be controlled and determines the thickness of the focal plane. This confocal pinhole rejects light that did not originate from the focal plane (see Figures). A thinner optical plane (i.e., smaller pin

hole diameter) provides a higher degree of resolution since interference from other planes is minimized.





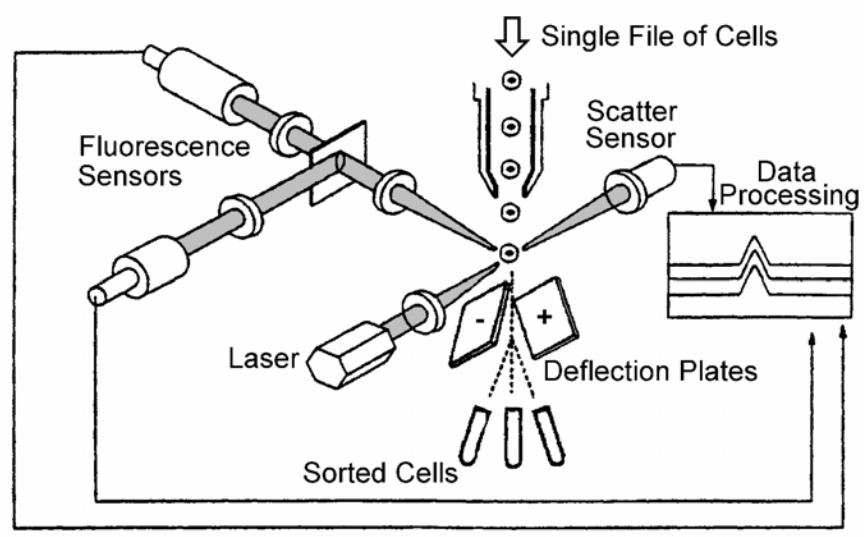
The image is produced by scanning the optical field pixel by pixel (x, y dimensions) with the laser and collecting the fluorescence intensity with a photomultiplier tube. A computer generates the image from this fluorescence data. In the case of dual- or multiple fluorchromes the data on the different wavelengths is collected separately and assigned different colors. Images of the separate fluorochromes or a merge image can then be displayed or exported into a file.

It is also possible to section a specimen into several optical planes by moving the stage vertically (z-dimension). This requires a computer-controlled motor to precisely move the stage according to the specified distances. The process of making these 'optical sections' involves setting the focal plane at the bottom of the specimen, recording the image in digital form, and moving the stage up a prescribed distance. Another digital picture is taken and the process is repeat throughout the thickness of the specimen resulting in images of consecutive focal planes separated by a specific interval. These digitized images are stored and can be viewed in sequence. In addition, three-dimensional images can be reconstructed from the optical sections through the use of computers.

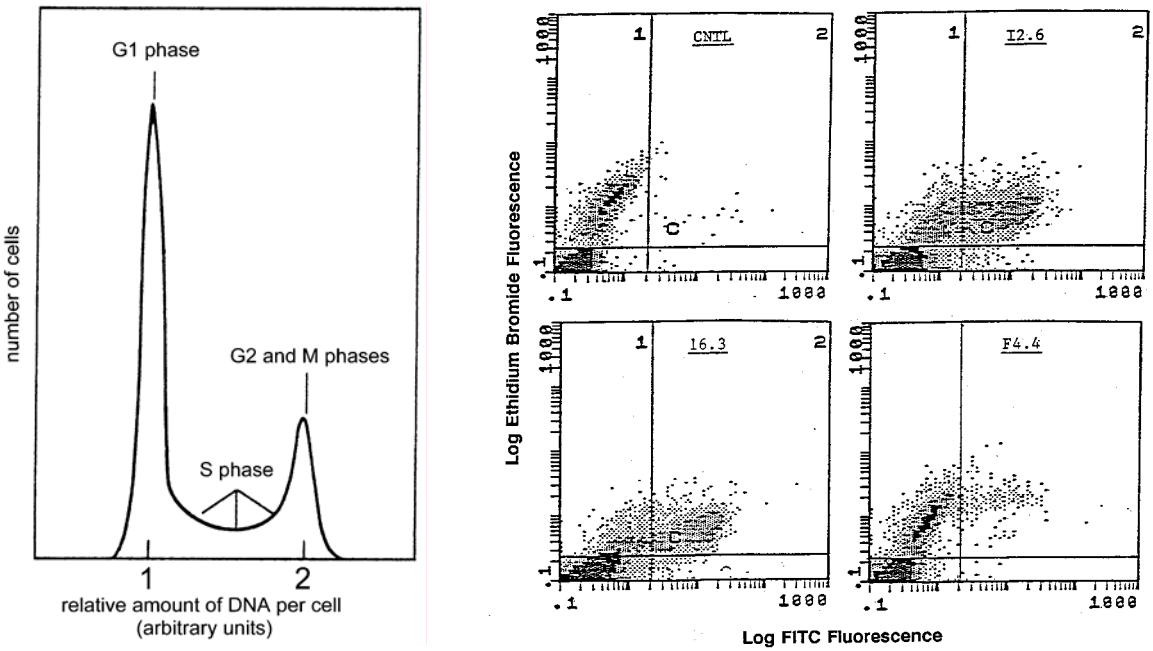
FLOW CYTOMETRY

Fluorescent microscopy provides qualitative information about individual cells. However, obtaining quantitative data is difficult and labor intensive. Fluorometry provides quantitative data, but represents the average of the population being analyzed. In other words, fluorometry does not reveal information about heterogeneity within that population. Flow cytometry can provide quantitative information on individual cells.

The principal of flow cytometry is to pass cells (or particles), in single file through a laser beam and analyze the interaction of the laser with individual cells (Figure). For example, the scatter (i.e., direction change) of the laser provides information about the size and shape of the cell. In addition, cells can be labeled with fluorochromes which are excited by the laser. The amount of fluorescence associated with the cells provide quantitative data on specific target molecules or their subcellular constituents.



Flow cytometry can quantify virtually any cell-associated factor or organelle for which there is a fluorescent probe (or natural fluorescence). For example, the amount of DNA per cell can be measured and provide information about the proportions of cells in various phases of the cell cycle (Figure). Through the use of multiple detectors and lasers it is also possible to analyze two or more fluorochromes with different emission spectra simultaneously. The data for the various cellular parameters being measured are accumulated for each cell and sub-populations can be identified and characterized. For example, the data can be plotted as dots defined by the fluorescence values of the two fluorochromes (Figure). The number of cells with particular characteristics are tabulated and their distributions in the cell population determined.



In addition, many flow cytometers are equipped with a cell sorter, or a fluorescence activated cell sorter (FACS). Cell sorting is accomplished by passing the cells between a charged plate and applying a pulse of current to cause the cell to change directions. In this way cells with particular characteristics (i.e., predetermined levels of fluorescence, scatter, etc.) can be separated from a mixture of cells. The separated cells can then be subjected to other analyses.

CHAPTER 4--RADIOCHEMISTRY

Radioactivity is widely used in biological research. Radioisotopes are used to label molecules of interest and these molecules can be analyzed. One advantage of radioactivity, as compared to other analytical techniques, is its sensitivity. For example, detecting 10^{-12} moles (pmole) is feasible in many applications. The other major advantage relates to the ability to label molecules with radioisotopes and not greatly affect their chemical properties. The use of chromophores or fluorochromes for analytical work usually involves adding another chemical group to the molecule of interest which will have an impact on the properties of the molecule. These two features, sensitivity and chemical similarity, make radioisotopes powerful experimental tools. For example, its is possible to follow the conversion of one molecule into another through the use of 'tracer' techniques. Radiolabled compounds can also be used as ligands in binding studies or as substrates (eg., enzyme assays, macromolecular synthesis or permeability studies).

ATOMIC STRUCTURE AND RADIOACTIVE DECAY

Atoms are composed of a nucleus, containing protons and neutrons, surrounded by electrons. Protons (p^+) have a mass of 1 atomic mass unit (amu) and a charge of +1, neutrons (n) have a mass of 1 amu and are neutral (i.e., no charge), and electrons (e^-) have a charge of -1 and a negligible mass (0.0005 amu). The mass of an atom is equal to the sum of the number of protons and neutrons. An atomic mass unit is defined as the mass of a proton or neutron and approximately equals 1.66×10^{-24} grams. The number of electrons orbiting the nucleus is equal to the number of protons.

Major Types of Radiation

α-particles	2p2n (He nucleus)
β-particles	electron (e^-)
γ-rays	photon ($h\nu$)

Only β- and γ-radiation are widely used in biological research.

The atomic number of an atom is equal to the number of protons and defines the chemical element. **Isotopes** are species of the same chemical element with different atomic masses (i.e., a different number of neutrons). Isotopes can be either stable or unstable. Unstable isotopes will undergo nuclear rearrangements, or decompose, and radiation is emitted as part of this decomposition. This emitted radiation can be the ejection of subatomic particles (i.e., neutrons, protons, electrons) from the nucleus or the emission of electromagnetic radiation (Table). Alpha-radiation is not widely used in biomedical research since the relatively large α-particles are difficult to detect.

Radioisotopes are found in nature and any particular chemical element is composed of a mixture of isotopes both radioactive and non-radiactive. Radioisotopes can also be synthesized. For example, bombarding nitrogen with neutrons (n) will displace a proton (p) and result in the formation of radioactive carbon (^{14}C) according to:

$$^{14}N_7 + n \rightarrow {}^{14}C_6 + p^+$$

The superscripts are the atomic masses (neutrons + protons) and the subscripts are the atomic numbers (i.e., protons). The resulting ^{14}C is unstable and will decay by converting a neutron into a proton. (A neutron is composed of a proton plus an electron and this nuclear electron is

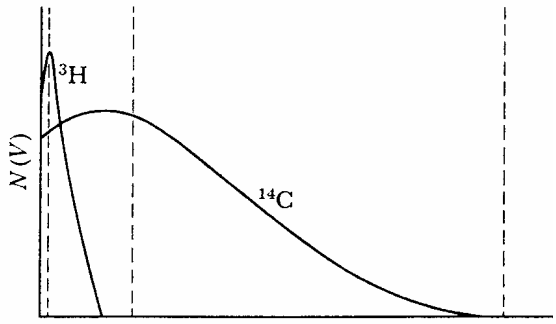
ejected as a β-particle). Conversion of a neutron into a proton results in the formation of nitrogen according to:

$$^{14}C_6 \rightarrow {}^{14}N_7 + \text{ß (or } e^-) + \text{neutrino}$$

Also emitted from the decaying nucleus is a neutrino, an entity with little mass and no charge.

PROPERTIES OF RADIOISOTOPES

Each radioactive decay, or emission, has a specific energy. These emissions have distinct energy spectrums, or ranges for each particular isotope (Figure). In the case of b-radiation this energy corresponds to the velocity at which the electron is ejected from the nucleus. Each isotope is characterized by a maximal energy, or E_{max}, that can be emitted and an average energy of the emissions. The average energy of the emissions is approximately one-third of the E_{max}.



Energy spectra of ^{3}H and ^{14}C. On the y-axis is the frequency, or number of emissions with a particular energy (x-axis).

Radioactive decay is a random process in that the probability that a particular atom will decay is the same for each of the atoms of that particular isotope (i.e., one decay does not affect another decay). However, each isotope has a distinctive rate of decay which is defined by a **half-life**. The half-life of an isotope is the time it takes for half of the radioactivity to decay and is independent of the amount of radionuclide. Radioisotopes can be defined by the energy of their emissions and their half-life (Table).

Properties of Commonly Used β-emitters in Biological Sciences

Isotope	*E_{avg}	*E_{max}	Half-Life
^{3}H	0.0055	0.015	12.3 years
^{14}C	0.05	0.15	5500 years
^{32}P	0.7	1.71	14.3 days
^{35}S	0.0492	0.167	87.1 days

*Energy expressed in MeV (mega electron volts)

Units of Radioactivity

Curie (Ci)	3.7×10^{10} dps 2.2×10^{12} dpm
becquerel (Bq)	1 dps

The ICRU recommends the use of the Bq as the official unit of measure.

Units of radioactivity are expressed as disintegrations (or decays) per unit time. For example, the amount of radioactivity in a sample can be expressed as disintegrations per second (**dps**) or disintegrations per minute (**dpm**). The recommended unit of radioactivity is the Becquerel (Bq) which is defined as one dps (Table). However, the curie (Ci) is still commonly used. The definition of one curie was originally the dps in one gram of ^{226}Ra. This definition has been standardized to 3.7×10^{10} dps. Radioactivity is often expressed as counts (= detected disintegrations) per unit time (i.e., **cps** or **cpm**). All of these units express how much radioactivity is present. Because each disintegration represents a loss of radioactivity, this value changes with time according to the half-life of that isotope. In other words, there is only half of the amount of radioactivity remaining after one half-life.

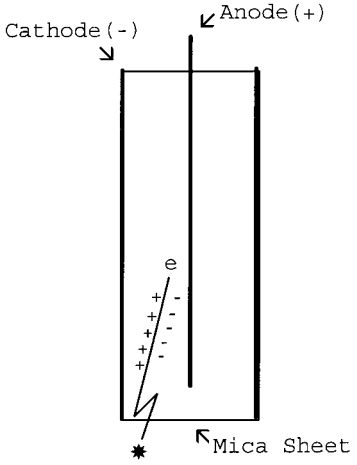
Two other terms used to describe radioactive compounds are **specific activity** and **radiochemical purity**. The specific activity refers to the disintegration rate per unit mass (eg., GBq/mmole) and reflects the mixture of non-radioactive and radioactive isotopes. The maximum possible specific activity is determined by the half-life. In other words, it is an inherent property of the isotope. The specific activity provides a means to convert units of radioactivity into mass units (eg., mg, moles, etc).

Radiochemical purity refers to the variety of compounds that are radiolabeled in the preparation. This is expressed as a percent and indicates the degree of purity. This can change over time due to decomposition of the compound of interest into other compounds.

MEASUREMENT OF ß-RADIATION

- Geiger-Mueller Counting
- Autoradiography (film exposure)
- Scintillation Spectrophotometry

Three methods are commonly used to measure β-radiation (Box). None of these methods measure β-particles directly, but measure the effect of collisions between β-particles and some component of the assay system.



Geiger-Mueller Counter

The G-M counter is not widely used as a quantitative instrument, but is used most often for monitoring possible radioactive contamination and estimating the amount of radioactivity. A typical G-M counter consists of a cylindrical cathode with a central anode. The tube is filled with an inert gas and sealed on one end by a mica sheet. β-particles passing through the mica sheet will dislodge electrons from the inert gas resulting in ion pairs. The ion pairs are then detected as an electrical potential between the anode and cathode. The G-M counter is unable to detect low energy isotopes since radiation needs to penetrate the mica sheet and dislodge enough electrons to create a detectable electrical potential.

Autoradiography

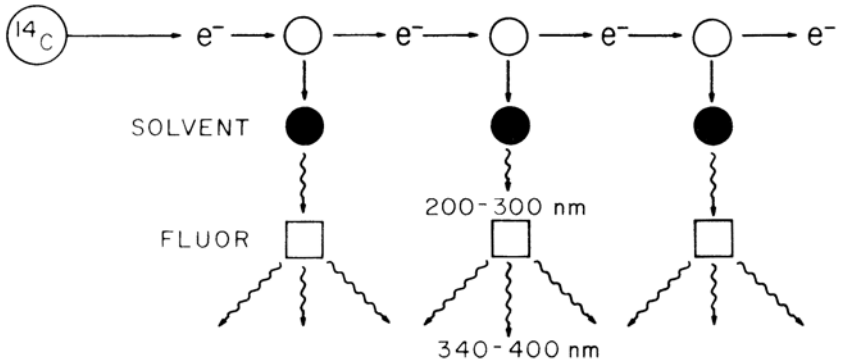
Autoradiography is the exposure of x-ray film (or a photographic emulsion) to a radioactive sample. The radioactivity behaves as light and results in the exposure of the photographic emulsion on the film. Radiation activates the silver halide in the photographic emulsion. During development the activated silver halide is converted to metallic silver. The metallic silver will appear as 'grains' on the film in positions where radioactivity is located. In other words the exposed film will be darken in proportion to the amount of radioactivity it was exposed to. If the film is juxtaposed to the sample it is possible to precisely localize the position of the radioactivity.

The two most common uses of autoradiography is in conjunction with gel electrophoresis or in conjunction with microscopy. Autoradiography of gels allows the identification and quantitation of specific proteins or nucleic acids (discussed later). It is also possible to localize radioactive markers to particular cells in tissue sections or subcellular structures within cells by auto-radiography. The typical procedure is to cover cells with a photographic emulsion (Box) and develop the emulsion after sufficient exposure. Dark grains (or spots) will appear over the cells or structures which contain the radioactive marker. It is usually desirable to use lower energy isotopes in autoradiography. High energy isotopes will travel greater distances and activate more silver halide atoms thus producing long 'tracks'.

1. Fix tissue or cells to slide.
2. Cover with a photographic emulsion.
3. Expose and develop.
4. Examine under microscope.

Scintillation Spectrophotometry

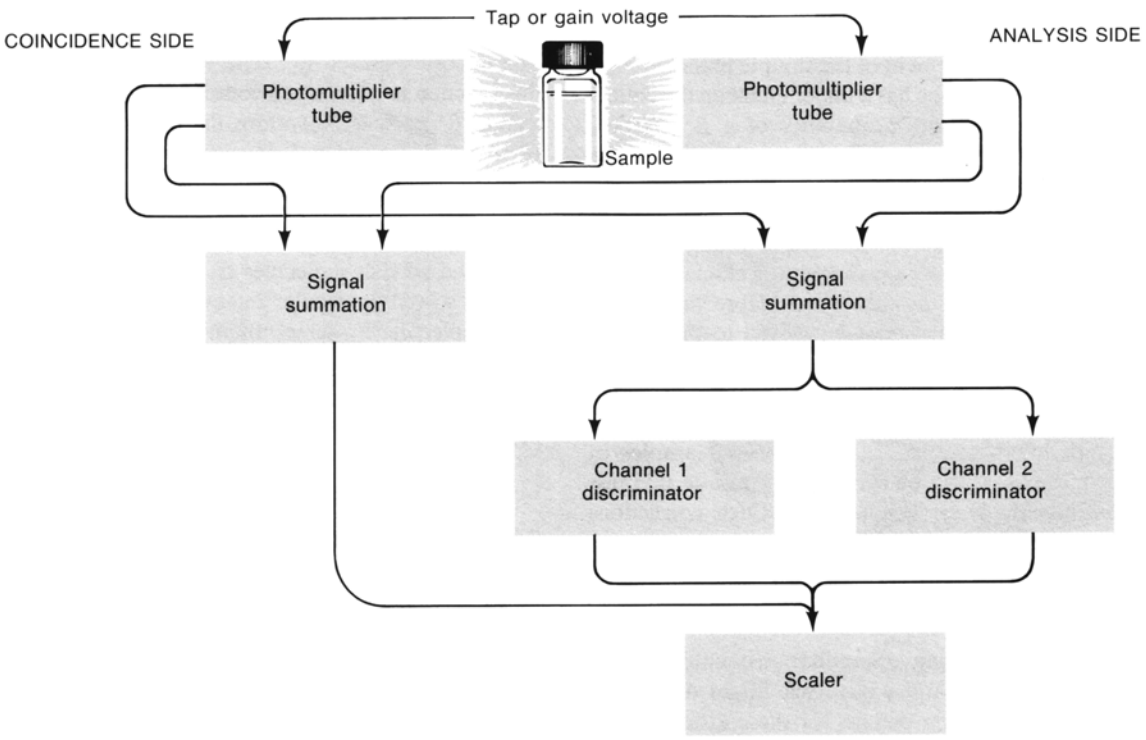
The scintillation spectrophotometer efficiently detects and measures low energy isotopes and low levels of radioactivity. In liquid scintillation counting (LSC) the radioactive sample is mixed with a **scintillation cocktail** containing chemicals known as fluors. The ß-particles collide with solvent molecules and transfer energy to these solvent molecules. The excited solvent molecules will then transfer this excess energy to the fluors (= fluorochromes). The excited fluors will emit a flash of light (i.e., fluoresce) when they return to its ground (unexcited) state. This flash of light is then detected and evaluated with photomultiplier tubes. The scintillation spectrophotometer records the number of flashes of light, which are converted to counts per minute (cpm).



Instrumentation. The instrument typically consists of two or more photomultiplier tubes (Figure). As in the case with spectrophotometers the photomultiplier tubes will convert light into an electrical signal. In addition, the scintillation counter has coincident circuits which only count events that are simultaneously recorded on both photomultiplier tubes (signal summation), thus minimizing background counts from the photomultiplier tubes.

ß-particles with higher energy levels will excite more solvent molecules which will activate a greater number of fluor molecules. This will result in more photons being emitted which generates an electrical pulse with a higher voltage. Therefore, not only can radioactive disintegrations be counted, but the energy of each disintegration can also be determined.

Through the use of channel discriminators this allows the number of counts within particular energy ranges to be determined.



Background. Background radiation sources (Box) can interfere with the accurate measurement of radioactivity in experiments. The earth is constantly being bombarded by cosmic radiation and radioactive isotopes are found in nature. Cosmic and other natural radiation are generally not a problem since they are relatively constant. However, it does vary geographically. Lead shielding around the counting chamber eliminates much of the cosmic and natural radiation. Another potential source of background is ^{40}K found in glass. Low potassium glass, or plastic, should be used in the manufacture of scintillation vials. The instrument is also a source of background radiation due to spontaneous electrical pulses. Instrument background is controlled for by the coincidence circuits.

- cosmic
- natural
- machine
- sample

The most problematic source of background is often the sample itself. The nature of sample background depends upon the experiment. Sample background is determined from 'blanks' which contain radioactivity but no sample. For example, if radioactivity is being used to determine enzyme activity, then a blank containing all components except the enzyme needs to be counted and subtracted from the sample counts.

Statistics of Counting. Nuclear disintegrations are random and infrequent. The accuracy of counting depends

68% confidence interval
- counts ± $(counts)^{1/2}$

95% confidence interval
- counts ± $1.96(counts)^{1/2}$

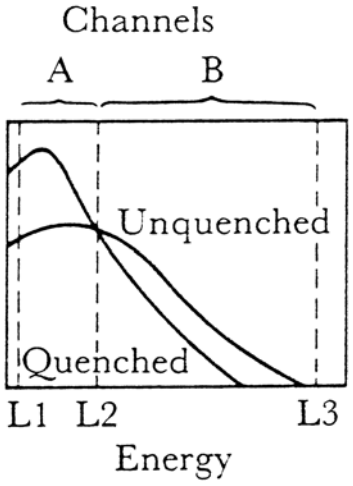
upon total number of disintegrations observed (Box). The proportional error at a 95% confidence level is defined by $1.96(\text{counts})^{1/2}/(\text{counts})$. In other words you need ~10,000 total counts to be 95% certain that the number determined is within ±2% of the true value. In practical terms this means that samples with low levels of radioactivity need to be counted longer. For example, if you count a sample for 1 minute and it contains 1000 cpm (i.e., 1000 total counts) the error term is ±6%. If you count the same sample for 10 minutes and it contains 1000 cpm (i.e., 10,000 total counts) the error term is now ±2%.

Efficiency of counting. Not all disintegrations are detected and recorded as counts during scintillation spectrophotometry. The efficiency of counting refers to the ratio of the cpm to the dpm (i.e., cpm = dpm x efficiency). The inability to detect a disintegration can be due to the light pulse being below the threshold of detection. In other words, the light intensity is too low to result in an electrical signal. The efficiency of detecting radioactivity is generally directly proportional to the energy level of the isotope.

Quenching will also affect the efficiency of counting. Quenching is the reduction in the transfer of energy from the β-particles to the photomultiplier tube. **Chemical quenching** is the prevention of fluorescence of the fluors by substances in the sample (eg., strong acids or bases). **Color quenching** is the absorption of the energy of the β-particle or photon emitted by fluor or solvent by substances in the sample (eg., chromophores). The efficiency of counting is not a problem if the efficiency is the same for all the samples. However, if the efficiency (i.e., quenching) is different between the samples then it is necessary to determine the dpm in the each of samples. Three methods by which efficiency can be determined are: 1) the internal standard method, 2) the channels ratio method, and 3) the external channels ratio method.

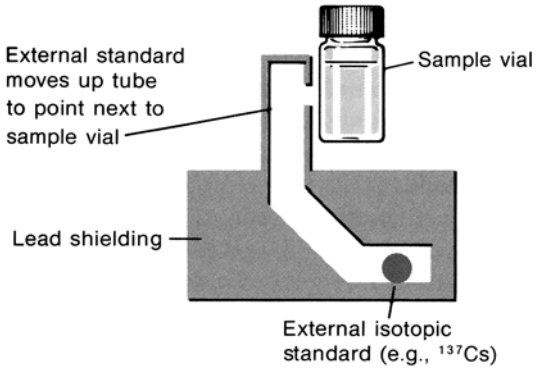
1) internal standard

After the samples have been counted a known amount of radioactivity (usually ^{3}H-toluene or ^{14}C-toluene) is added to each sample and the samples are recounted. From this data the efficiency can be determined for each sample (see Appendix 1 for calculations). The limitations of this method are that it is tedious, expensive, and the sample cannot be recovered.



2) channels ratio

Quenching reduces the efficiency by shifting the energy spectrum to a lower level. Different amounts of the quenching agent are added to a known amount of radioactivity and the cpm in two 'channels' are determined. A standard curve is generated by plotting efficiency vs. the channel ratios (see Appendix 2). The efficiency of each sample is then determined by comparing the channel ratios of each sample to the standard curve. Many scintillation counters can be programmed to carry out this analysis and



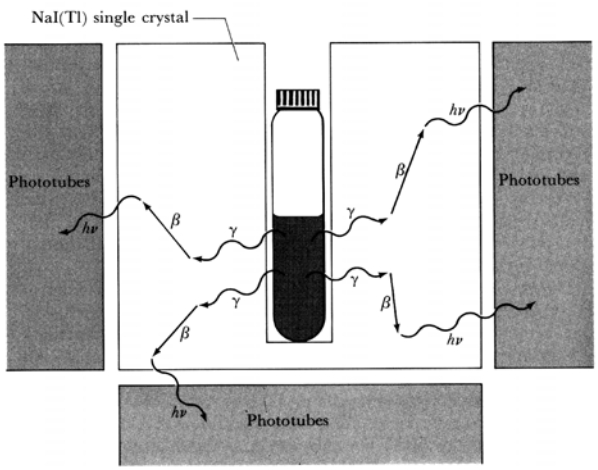
to calculate the results. The channels ratio method is not very accurate in cases of low radioactivity.

3) external channels ratio

An external radiation source (usually ^{137}Cs) is placed next to the sample and the radioactivity determined. Electrons are ejected from solvent molecules by the γ-rays and behave like β-particles. This method incorporates principles of both internal standard and channels ratio methods. The calculation is complex and carried out by the instrument.

Counting multiple isotopes. Scintillation spectrophotometry can also determine the amount of different isotopes in a mixture. In the case of isotopes that have little overlap in their energy spectrums, such as ^{3}H and ^{32}P, it is relatively easy. Channels are set so that one channel contains low energy isotope and the other channel contains the high energy isotope. In cases where the energy spectrums overlap it will be necessary to use channel ratios to determine amount of each isotope. Radioactive standards are use to determine how much of each isotope is found in each channel. If possible, the lower energy channel should be set so that it contains all of the lower energy isotope. In other words, virtually no counts from the lower energy isotope should be detected in the higher energy channel. (See Appendix 3 for sample calculation.)

GAMMA-RAY DETECTION



Gamma-rays can be detected and measured by the same methods as β-particles (eg., G-M counter, autoradiography, and LSC). Like β-particles, γ-rays are ionizing radiation and will be detectable with a G-M counter. Similarly, γ-rays are the same as X-rays (i.e., photons) and will expose photographic emulsions. In addition, electrons (i.e., β-particles) are ejected when γ-rays interact with other substances. Although it is possible to measure γ-rays with a liquid scintillation counter, it is more common to use an instrument specifically designed to count γ-radiation. The counting chamber of a γ-counter is surrounded by a thallium-activated NaI crystal. When γ-rays interact with the NaI(T) secondary electrons (β-particles) are ejected. The β-particles then result in radiation in the visible range (i.e., light) which is detected and quantitated by photomultiplier tubes as in scintillation spectrophotometry. The advantage of the γ-counter is that no scintillation cocktail is necessary.

SAFETY

Radioisotopes are a health hazard since ionizing radiation can lead to cellular damage and DNA mutations. Radioisotopes need to be handled safely at all times and precautions against exposure must always be taken. The time of exposure to radioactivity should be kept at a minimum. Whenever possible plexiglass shielding should be placed between the user and the radioactive sample when moderate and high energy isotopes are being used. Lab coats and disposable gloves should be worn at all times. Avoid ingestion of isotopes by observing the no eating, drinking, smoking, mouth-pipetting, etc. rules in the laboratory. Some users tend to be more relaxed with the use of low energy isotopes such as ^{3}H than with higher energy isotopes such as ^{32}P. However, ^{3}H incorporated into the cell can cause more damage than ^{32}P since ^{3}H only travels a short distance and will have more of a chance to damage cellular structures. Because of its high energy, ^{32}P will be more likely to exit the cell without doing any damage.

APPENDIX 1. INTERNAL STANDARDS

Calculation of efficiency using the internal standard method.

1. Measure the radioactivity in all of the samples (=cpm_1).

2. Add a known amount of radioactivity (dpm_{std}) to each of the samples and recount (=cpm_2).

3. Calculate the efficiency (E) for each sample using the following equation:

$$E = (cpm_2 - cpm_1)/dpm_{std}*$$

4. Divide the cpm for each of the samples (cpm_1) by the efficiency (E) for that particular sample resulting in the dpm_{sample}.

*Derivation of equation:

1) $cpm_2 = cpm_1 + cpm_{std}$ (The cpm after the addition of the radioactivity will be equal to the original cpm + the cpm of the standard.)

2) $E = cpm_{std}/dpm_{std}$ (This is the definition of Efficiency.)

3) $cpm_{std} = (E)dpm_{std}$ (Rearrange eq. 2 and solve for cpm_{std}.)

4) $cpm_2 = cpm_1 + (E)dpm_{std}$ (Substitute the value of cpm_{std} from eq. 3 into eq. 1.)

5) $E = (cpm_2 - cpm_1)/dpm_{std}$ (Solve eq. 4 for Efficiency.)

APPENDIX 2. CHANNELS RATIO

Calculation of efficiency using the channels ratio method.

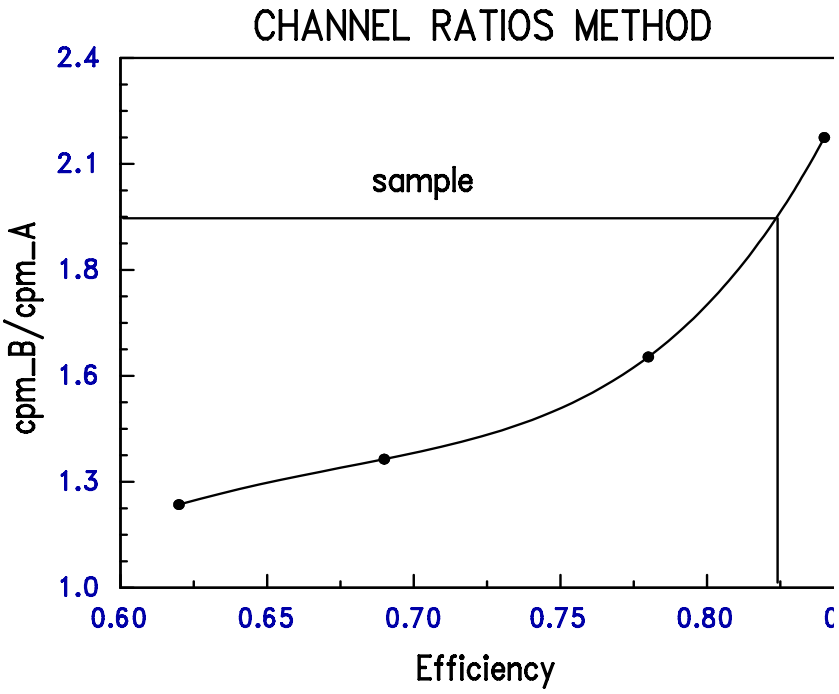
1. Take a known amount of radioactivity and add increasing amounts of the suspected quenching agent and determine the cpm in two channels (A and B).

2. Calculate the efficiency for each of the standards containing the quenching agent using eq. 2 from above.

3. Plot the channel ratios (cpm_B/cpm_A) on the y-axis vs. the efficiency on the x-axis to make a standard curve like the figure below. (Can also use cpm_A/cpm_B as the channels ratio.)

4. From the channels ratio of each sample determine the efficiency of each sample using the standard curve.

5. Divide the total cpm ($cpm_A + cpm_B$) in each sample by the efficiency for that sample to determine the dpm for each sample.



APPENDIX 3. MULTIPLE ISOTOPES

Example for calculation of multiple isotopes in mixed samples.

1. Set channels so that virtually all of the counts of the lower energy isotope (eg., ^{3}H) are in channel A.

2. Determine the cpm in both channels A and B for all of the samples as well as a ^{3}H standard and a ^{14}C standard. (Standards should be counted under same conditions as samples. If samples exhibit different levels of quenching then the dpm for the samples and standards will need to be determined.)

3. Use the channel ratios of the standards to calculate the amount of each isotope in each channel. Sample data:

	Radioactivity (cpm)		Channel Ratios	
Sample	Channel A	Channel B	B/A	A/B
Experimental	500	800		
^{3}H standard	16,840	25	0.0015	-
^{14}C standard	4250	10,200	-	0.416

- ^{3}H in channel B: (0.0015)(500) < 0.75 cpm (since some of the 500 cpm are due to ^{14}C, therefore
- ^{14}C in channel B ≈ 800 cpm*
- ^{14}C in channel A = (0.416)(800) = 333 cpm
- ^{3}H in channel A = 500 - 333 = 167
- Total ^{3}H = **167 cpm**
- Total ^{14}C = 800 + 333 = **1133 cpm**

*In other words, at these channel settings there is no significant amount of ^{3}H in channel B and all of the radioactivity can be assumed to be ^{14}C.

CHAPTER 5--pH AND BUFFERS

Maintaining the proper pH during experimental procedures is often one of the most critical factors in determining whether an experiment will work. Buffers are the components in solutions that maintain constant pH at a defined value. The pH is maintained due to an equilibrium established between weak acids and bases. The same principles can also be applied to other equilibrium phenomenon.

DEFINITIONS

An **acid** donates a proton (H^+) by dissociating into its **conjugate base**:

$$\mathbf{acid \leftrightarrow conjugate\ base^- + H^+}$$

A **base** accepts a proton (H^+) forming a **conjugate acid**:

$$\mathbf{base + H^+ \leftrightarrow conjugate\ acid^+}$$

A **strong acid** is one that completely ionizes in solution while **weak acids** only partially ionize in solution. The weak acid and its conjugate base are in equilibrium. The equilibrium between the acid and its conjugate base will determine the pH and weak acids are used as pH buffers. **pH is defined as the -log [H^+]** (where [] brackets refer to molar concentration).

HENDERSON-HASSELBACH

The ionization of weak acids is defined in part by the chemical properties of the individual weak acid. Therefore, it is possible to calculate the pH of a solution containing a weak acid and its conjugate base. The ionization of a weak acid can be defined by the following equation:

$$\mathbf{HA \leftrightarrow A^- + H^+}$$

The dissociation constant, or $K_a = [H^+][A^-]/[HA]$. Solving for $[H^+]$ results in:

$$\mathbf{[H^+] = [HA]K_a/[A^-]}$$

Since $pH = -\log[H^+]$, take the -log of both sides of the equation:

$$\mathbf{-\log[H^+] = (-\log K_a) + (-\log[HA]/[A^-])}$$

Substitutions and rearrangements yield the Henderson-Hasselbach equation:

$$\mathbf{pH = pK_a + \log[A^-]/[HA]}$$

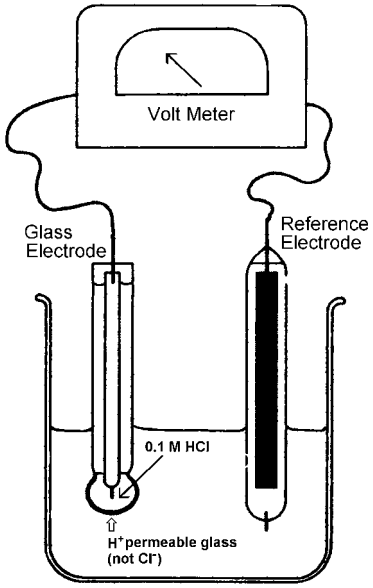
The pK_a is equal to the pH when the concentrations of the protonated acid and the ionized form

are equal. The pK_a value is dependent on the particular weak acid or ionizable group and needs to be either looked up or determined. The equation is also written as:

$$\mathbf{pH = pK_a + log[salt]/[acid]}$$

The Henderson-Hasselbach equation can be used to calculate the ratio of acid and salt (= conjugate base) that is needed to produce a buffer solution of desired pH. (See appendix for examples.)

INSTRUMENTATION



The pH meter is composed of 1) a reference electrode, 2) a glass electrode, and 3) a voltmeter. (The two electrodes are usually combined into a single electrode.) The glass electrode is made of a special glass that is permeable to H^+, but not other cations or anions. The electrode is filled with 0.1 N HCl. The H^+ diffuses from higher concentration to lower concentration leaving behind Cl^-, thus creating an electrical potential. This electrical potential is proportional to the pH of the solution being measured as expressed by:

$$\mathbf{Voltage = E^o + (2.303RT/F)\cdot pH}$$

where E^o = summation of electrical potentials (i.e., voltage) that are not pH related, R = gas constant, T = absolute temperature, F = Faraday constant. A voltmeter will measure this electrical potential of the glass electrode (i.e., E) as compared to the reference electrode (i.e., E^o) and convert it to a pH value. It is necessary to calibrate the meter using a solution of known pH before use due to changes in electrode HCl concentration and other factors that affect E^o.

COMPLICATIONS

- ionic concentration
- Na^+ error
- temperature

Several factors can affect pH and/or the measurement of pH (Box). Generally, these factors will have a minimal effect. In addition, some electrodes are incompatible with certain buffers; the most notable being Tris buffers.

Ionic Concentration. The concentration of ions and in particular the concentration of the buffer itself can affect pH. Although Henderson-Hasselbach is independent of concentration the equation is most accurate at low buffer concentration. This is especially important when stock solutions that will be diluted to the final buffer concentration are used. Therefore it is generally best to make the final pH adjustment at the time the buffer is diluted.

Na^+ Error. Sometimes pH electrodes will read the $[Na^+]$ in addition to the $[H^+]$. This occurs most often at high pH; therefore it is better to use KOH to adjust the pH of alkaline

buffers.

Temperature. The temperature affects the pK_a of a buffer. The degree that this occurs is dependent on the buffer. For example, Tris has a $\Delta pK_a/C^o = -0.03$. Therefore, it is important to adjust the pH at the temperature at which the buffer will be used.

BUFFERING CAPACITY

Buffering capacity refers to the amount of acid or base you can add to the system without greatly affecting the pH. Buffering capacity (ß) is defined in operational terms:

$$ß = \Delta B/\Delta pH$$

where B = the mole equivalent of strong base or strong acid added and ΔpH is the resulting change in pH. Buffering capacity will be the greatest at the pK_a of the buffer.

BUFFER SELECTION

There are no well defined criteria for choosing a buffer except that the pK_a should be close to the desired pH range (± 1 pH unit) since the buffering capacity is greatest at the pK_a. Buffers should not participate in reactions that are being analyzed or otherwise affect the outcome of the experiment. For example, the buffer should not bind to other components in the mixture. The most common criteria for choosing a buffer is the scientific literature. In other words, use the same buffer for a particular procedure that has worked before. The particular experiment also needs to be taken into account. For example, if a reaction produces H^+, then the buffer should have a pK_a lower than the desired pH. This will provide more buffering capacity since the pH will move towards the pK_a as the reaction proceeds.

Good's Buffer Criteria

- pK_a between 6 and 8
- high solubility in aqueous solutions
- exclusion by biological membranes
- minimal salt effects
- minimal effects on dissociation due to concentration, temperature and ionic composition
- well-defined or non-existent interactions with cations
- chemical stability
- insignificant light absorption between 240 and 700 nm
- readily available in pure form

During the 1960's Dr. Norman Good examined a series of zwitterionic compounds as biological buffers based upon some ideal properties (Box). A series of such buffers--often called Good buffers--with pK_a values ranging from 6-10 was developed. These buffers have long names and are more commonly known by their abbreviations (eg., HEPES).

BUFFER APPENDIX

Example 1. How many grams of acetic acid (CH_3CO_2H) and sodium acetate are needed to make up one liter of a 50 mM acetate buffer with a pH of 5.0?

5.0 = 4.76 + log[Naacetate]/[acetic a.]	Substitute values into Henderson-Hasselbach
0.24 = log[Naacetate]/[acetic a.]	Subtract 4.76 from both sides.
1.74 = [Naacetate]/[acetic a.]	Take anti-log (10^X) of both sides
[Naacetate] = 1.74[acetic a.]	Multiply both sides by [acetic a.] (Eq. 1)
[Naacetate] + [acetic a.] = 0.05	The sodium acetate and acetic acid concentrations must add up to the desire final molar concentration of the buffer. (Eq. 2)
1.74[acetic a.] + [acetic a.] = 0.05	Substitute eq. 1 into eq. 2.
2.74[acetic a.] = 0.05	
[acetic a.] = 0.018 moles/liter	Eq. 3
[Naacetate] + 0.018 = 0.05	Substitute eq. 3 into eq. 2
[Naacetate] = 0.032 moles/liter	Eq. 4

0.018 moles/liter x 60 g/mole x 1 liter = **1.09 grams of acetic acid**

0.032 moles/liter x 82 g/mole x 1 liter = **2.62 grams of sodium acetate**

Example 2. After making the calculations for example 1, you realize that there is no available sodium acetate. However, you do have a 1 M NaOH solution. How is the buffer made now?

The total acetate concentration must now come from the acetic acid, so:

0.05 moles/liter x 60 g/mole x 1 liter = **3.0 grams of acetic acid**

The calculations from example 1 are still valid. The only change is to substitute [NaOH] for [Naacetate] in eq. 4. Concentrations are easily calculated using:

$C_1V_1 = C_2V_2$

Therefore (0.032M)(1 liter) = (1M)V_2, so V_2 = **32 ml of 1 N NaOH**

Example 3. How many grams of NaH_2PO_4 and Na_2HPO_4 are needed to make 50 ml of a 0.1 M phosphate buffer with a pH of 7.2? Phosphate has three ionizable groups with different pK_a values. Choose the pK_a closest to the desired pH. Since pH = pK_a use equal molar amounts of the monobasic and dibasic phosphates.

NaH_2PO_4 = 0.05 moles/liter x 138 g/mole x 0.05 liter = **0.345 g**

Na_2HPO_4 = 0.05 moles/liter x 142 g/mole x 0.05 liter = **0.355 g**

Example 4. How many grams of Tris and TrisHCl are needed to make up 500 ml of 1 M Tris buffer with a pH of 8.0? Although Tris is a base the calculations are the same. Make sure that the base form and the conjugate acid form are correctly identified.

8.0 = 8.3 + log[Tris]/[TrisHCl]

[Tris] = 0.5[TrisHCl] and [Tris] + [TrisHCl] = 1

[TrisHCl] = 0.67 M x 157.6 g/mole x 0.5 l = **52.8 g**

[Tris] = 0.33 M x 121 g/mole x 0.5 l = **20 g**

ADDITIONAL PRACTICE PROBLEMS

How many grams of acetic acid and how many ml of 1 M NaOH are needed to make 1 liter of 50 mM acetate buffer, pH 5.0? The pK_a of acetate is 4.76 and molecular weight of acetic acid is 60.

How many grams of HEPES and how many ml of 1 M NaOH are needed to prepare 1 liter of a 0.1 M HEPES buffer at pH 7.4? The pK_a of HEPES is 7.5 and the molecular weight is 238.

How many grams of each Tris and TrisHCl are need to make up 1 liter of 0.5 M Tris buffer at pH 8.8? The pK_a of Tris is 8.3 and the molecular weights of Tris and TrisHCl are 121 and 157 respectively.

Buffer	Mol. Wt.	pK_a
acetate	60 (free acid) 82 (Na salt)	4.76
phosphate	138 (monobasic) 142 (dibasic)	2.15 7.2 12.3
MES	195 (free acid) 212 (Na salt)	6.15
HEPES	238 (free acid)	7.5
TRIS	121 (free base) 157 (TrisHCl)	8.3
Tricine	179 (free base)	8.15

How many grams of MES (mol. wt. = 195) and the Na^+ salt of MES (mol. wt. = 217) are needed to make 100 ml of 0.5 M MES buffer at pH 6.0. pK_a of MES = 6.15?

How many grams of Tricine (mol. wt. = 179) and how many ml of 1 N HCl are needed to make of 200 ml of 0.1 M Tricine buffer at pH 8.5. The pK_a of Tricine = 8.15?

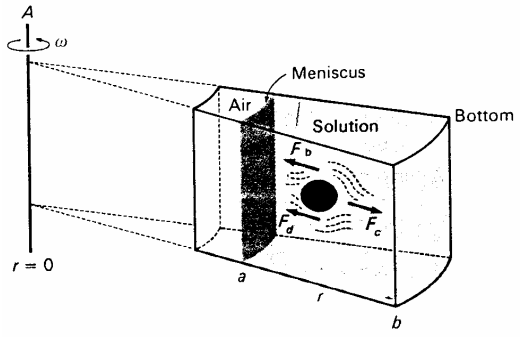
CHAPTER 6--CENTRIFUGATION

- Ultracentrifuge
  - ♦ Analytical
  - ♦ Preparative
- High Speed
- Table Top
- Clinical
- Microfuges

The centrifuge is an essential instrument in cell and molecular biology research. It is primarily used to separated biological components based upon differential sedimentation properties. Many types of centrifuges are available for various applications (Box). All centrifuges basically consist of a motor which spins a rotor containing the experimental sample. The differences between centrifuges are in the speeds at which the samples are centrifuged and the volumes of samples.

SEDIMENTATION THEORY

Spinning around an axis creates a centrifugal field (Figure). A particle in a centrifugal field will experience a centrifugal force defined by:



Centrifugal Force

$$F_c = m\omega^2 r$$

where F_c = the centrifugal force, m = mass of the particle, ω = angular velocity and r = distance from the axis. This force will be opposed by a buoyant force (F_b) and a frictional force (F_f).The buoyant force represents the force it takes to displace solvent as the particle moves through the centrifugal field. The frictional force represents the drag on the particle as it passes solvent molecules. These two forces are respectively defined as:

$$F_b = -m_o\omega^2 r \text{ and } F_f = -fv$$

where m_o = the mass of the displaced solution, f = frictional coefficient and v = velocity of the particle. The particle will move at a velocity such that the total force equals 0, thus:

$$F_c + F_b + F_f = 0, \text{ or } m\omega^2 r - m_o\omega^2 r - fv = 0$$

substituting $m\upsilon\rho_s = m_o$, where υ = partial specific volume of the particle and ρ_s = density of the solvent, and solving for v results in:

$$v = \omega^2 rm(1 - \upsilon\rho_s)/f = \omega^2 rm\upsilon(\rho_p - \rho_s)/f$$

This equation (expressed in the two different forms) tells us several things about sedimentation:

1. The more massive a particle, the faster it moves in a centrifugal field

2. The denser a particle (i.e., the smaller its υ) the faster in moves in a centrifugal field.

3. The denser the solution, the slower the particle will move in a centrifugal field.

4. The greater the frictional coefficient (factors such as viscosity, particle shape, etc. influence this parameter), the slower the particle will move.

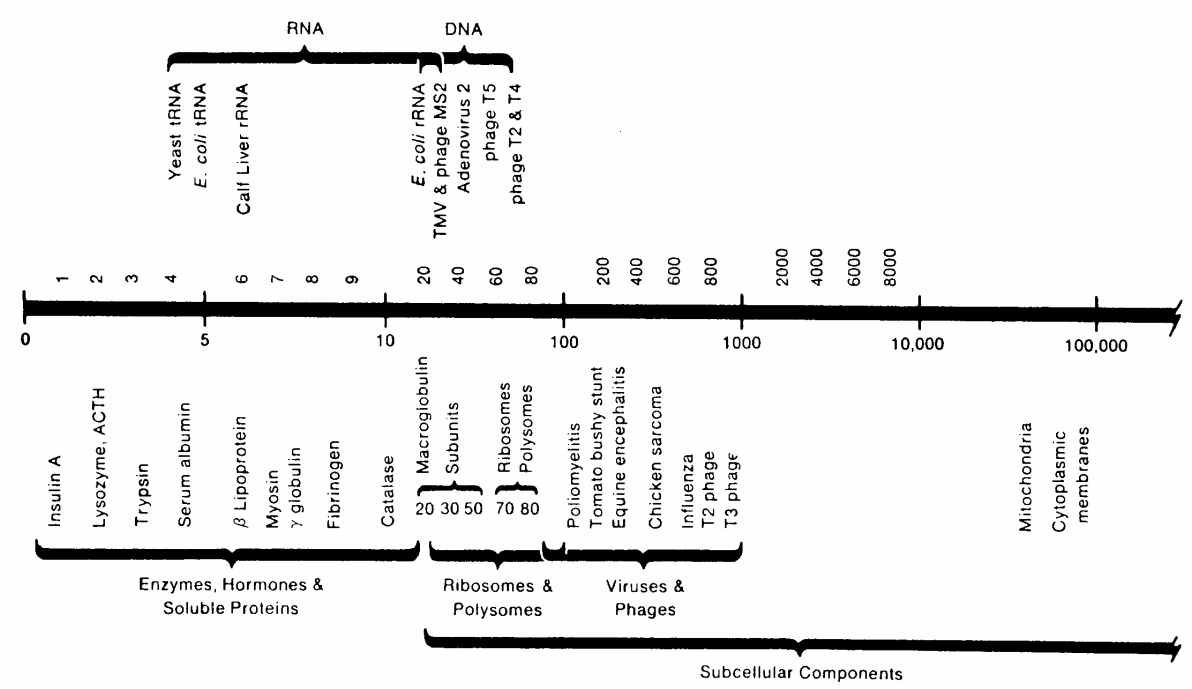
5. The particle velocity is 0 when the solution density is greater than the particle density.

6. The greater the centrifugal force ($\omega^2 r$) the faster the particle sediments.

The velocity per unit force will be defined as the sedimentation coefficient (s), or:

$$s = v/\omega^2 r = m(1 - \upsilon\rho_s)/f$$

When mass is expressed in g and f in g/sec s ranges from 10^{-13} to 10^{-11} sec. This is normally expressed in **Svedberg** (S) units where 1 S = 10^{-13} sec. The higher the S value, the faster that particle will sediment. For example, subcellular compartments and macromolecules have sedimentation coefficients which reflect their size, shape and density (Figure).

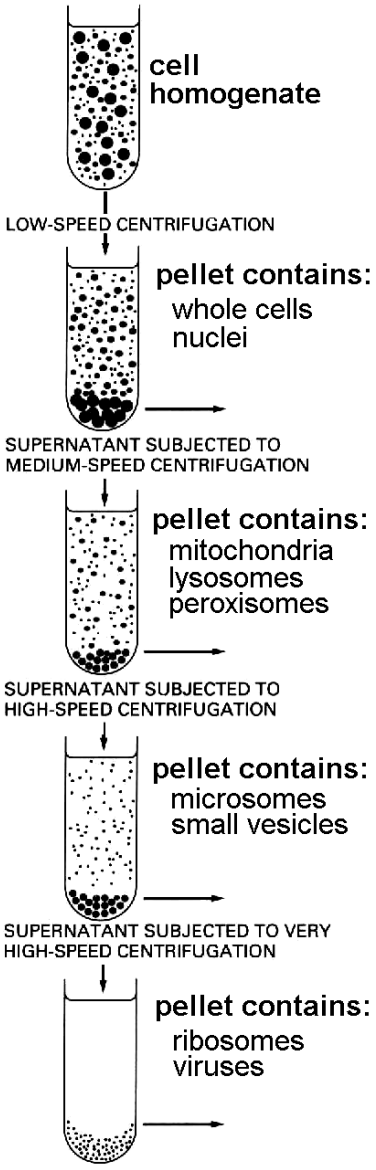


The sedimentation coefficient is determined by measure the velocity of a particle in a known centrifugal field. S units need to be normalized according to the composition of the medium. Sedimentation coefficients are usually determined with an analytical ultracentrifugation, also referred to as a model E ultracentrifuge. The rotor in the model E centrifuge allows the sample to be monitored spectrophotometrically as the sample is centrifuging. As the sample migrates in the centrifugal field the absorbance across the chamber will change. The sedimentation rate can be calculated from this change in absorbance. In addition, physical characteristics (such as size, density and shape) of a particle, or molecule, can be

determined from the sedimentation rate of a particle in a medium of known composition.

PREPARATIVE CENTRIFUGATION

The centrifuge is used most often in the biological sciences to collect material or to separate particles based upon their sedimentation properties. Preparative centrifugation takes advantage of the fact that more massive particles will sediment faster than less massive particles. For example, organelles and other subcellular components can be isolated by differential centrifugation (Figure). Differential centrifugation is carried out by centrifuging a sample at low speed and separating the supernatant and pellet. The supernatant is then recentrifuged at higher speed and the supernatant and pellet separated again.



Relative centrifugal force (RCF) is defined as the ratio of the centrifugal force to the force of gravity:

$$RCF = F_c/F_g = \omega^2 r/980$$

ω (expressed in radians/sec) is converted to revolutions per minute (rpm) by substituting $\omega = \pi(rpm)/30$, resulting in:

$$RCF = 1.119 \times 10^{-5} (rpm)^2 r$$

where r is expressed in cm. RCF units are expressed as "x g". (See appendix for nonograph.)

Sedimentation of a particular particle is dependent upon the RCF (see box) and how long the centrifugal force is applied. Samples must be centrifuged long enough for the particles at the top of the tube to reach the bottom. The size and shape of the centrifuge tube will also affect the centrifugation time. In addition, there is a difference in the RCF between the top and bottom of the tube. These factors need to be taken into consideration when scaling up from small pilot experiments carried out in small centrifuge tubes into large scale experiments. The term g·min refers to the product of the minutes a sample is centrifuged and the g-force. In general, assuming that the sizes and dimensions of the centrifuge tubes are similar, the product of the g-force and the time of centrifugation can be used to determine the conditions needed to to completely sediment a particle. For example, if it takes 60 min to completely sediment a particle at 1000 x g, then it should only take 10 min to sediment the same particle at 6000 x g.

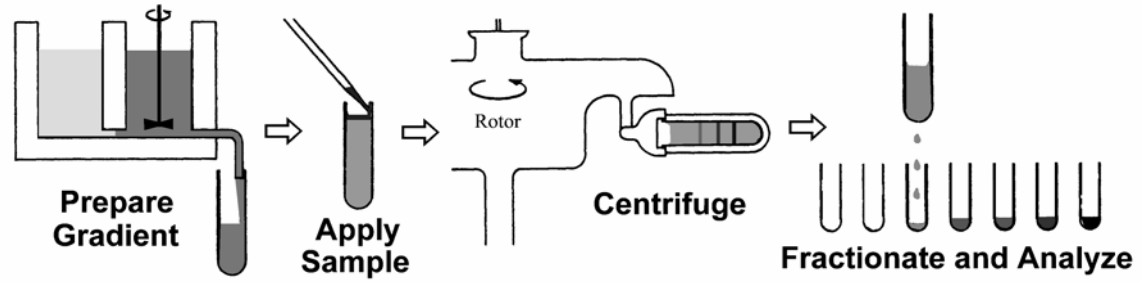
Centrifugation Through Density Gradients

- Sucrose
- CsCl
- Ficoll
- Hypaque
- Percoll

Fast sedimenting particles will be contaminated with slow sedimenting particles. The reason for this is that by the time large particles near the top of the tube are pelleted, some of the small particles near the bottom of the tube have also pelleted. In addition, mechanical vibrations, thermal gradients and convection currents can also affect the sedimentation properties. Centrifugation through a dense medium, or density gradient centrifugation, can partially alleviate these problems. In addition, density gradient centrifugation will allow for better separation of particles with similar properties. Several different media are commonly used in density gradient centrifugation depending upon the exact application (Box).

The two types of density gradient centrifugation are **rate zonal** and **isopycnic** (or equilibrium). In rate zonal centrifugation the density of the particles being separated are greater than the density of the solvent. Separation is based primarily upon size (i.e., larger particles will sediment faster). It is important to determine the optimal length of centrifugation for separating the particle of interest. If the centrifuge is not turned off soon enough all of particles will pellet. In isopycnic centrifugation the solvent density encompasses density of particles. The separation is based upon particle density. Centrifugation is carried out until equilibrium is reached (i.e., all particles have banded at densities corresponding to their own).

Density gradients can be preformed or formed during centrifugation. In the case of isopycnic density gradients it is possible to mix the sample with the desity gradient medium and carry out the centrifugation until the density gradient forms. The various components in the sample will then be found at a position which corresponds to their density. This is especially useful in the case of Percoll which rapidly forms density gradients when subjected to centrifugation. Another common example in which self-forming gradients are commonly used is the separation of nucleic acids on CsCl gradients (see Nucleic Acid Struction and Isolation).



Preforming the gradients before centrifugation decreases the amont of time and centrifugal force needed to reach equilibrium. In the case of rate zonal density gradients it is necessary to use preformed gradients. Gradient makers are used to produce continuous gradients, or step-wise gradients can be prepared by layering successive solutions of lesser density. Samples are layered onto the gradients and subjected to centrifugation. In the case of isopycnic gradients the sample can also be underlaid at the bottom of the tube and the various particles will 'float' to their correct densities during centrifugation. Following centrifugation the gradient is divided into fractions corresponding to different densities and analyzed.

The fractions are analyzed for the component(s) of interest and if needed the densities of the various frations can be determined. One method of determining the density is to measure the refractive index of the fractions and to then calculate the densities based on a standard curve prepared from solutions of known density made with the same medium of the desity gradients (eg., sucrose, etc.). Another method of determining the density is to prepare an identical gradient with 'marker beads' of known density. These marker beads are different colors and the position in the gradient is determined by measuring the distance from the top of the tube or determining which fraction they are associated with. This can then be use to estimate the range of densities in the other tubes containing sample. In subcellular fractionation experiments the various organelles can be evaluated by measuring components (eg., enzymes) known to be exclusively associated with the various subcellular compartments. This will identify the fraction(s) containing the organelle of interest as well as provide information about contamination with other organelles.

Subcellular Fractionation and Marker Enzymes

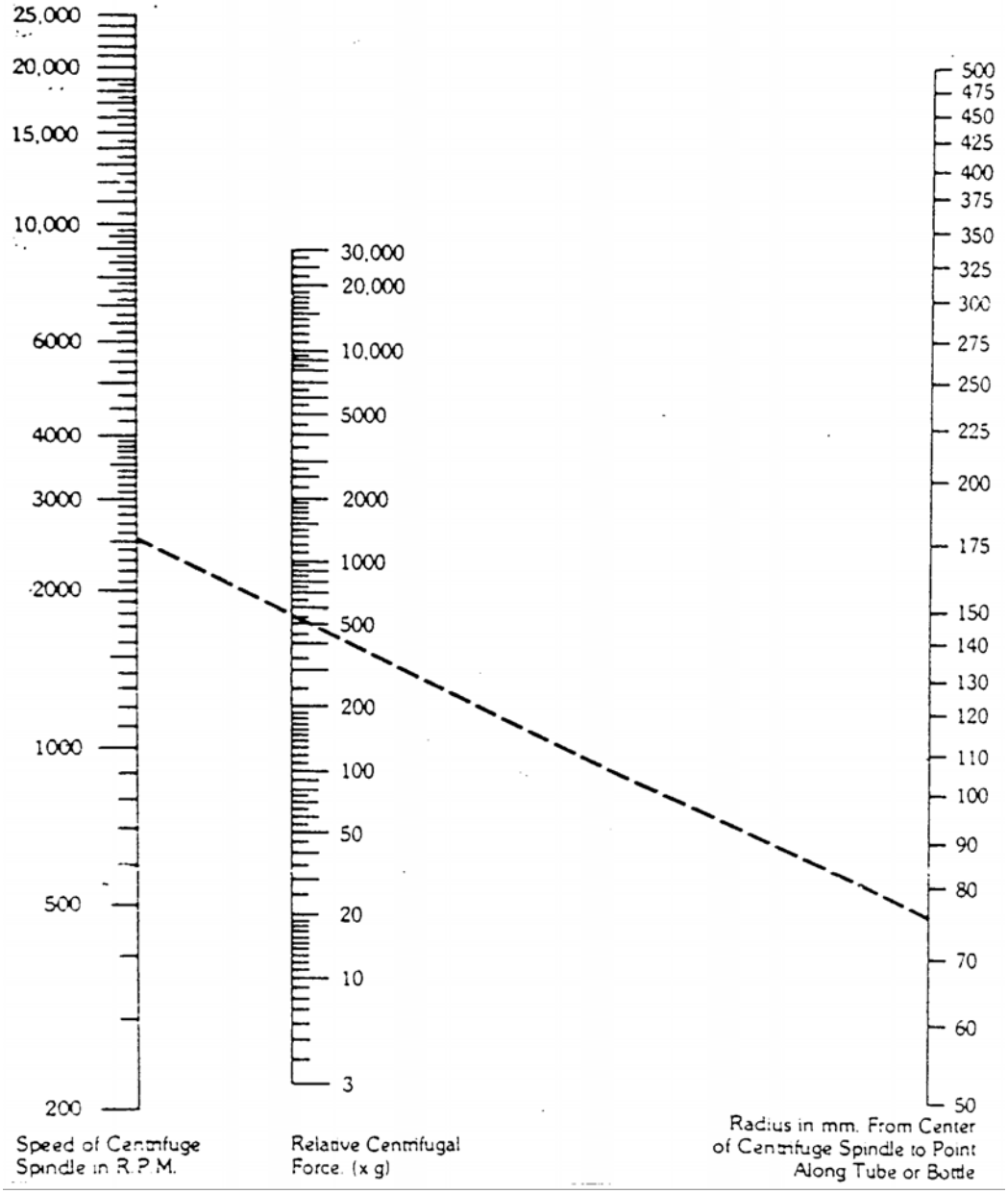
Organelle	**Marker**
nuclei	DNA
mitochondria	cytochrome oxidase
lysosome	hydrolases
peroxisome	catalase
Golgi	α-mannosidase
plasma membrane	adenylate cyclase
cytosol	lactate dehydrogenase

CENTRIFUGATION APPENDIX 1. RCF CALCULATION

The relative centrifugal force (**RCF**) can be calculated from the following equation:

$$\mathbf{RCF = (1.119 \times 10^{-5})(rpm)^2(r)}$$

where rpm is the speed of rotation expressed in revolutions per minute and r (radius) is the distance from the axis expressed in cm. The RCF units are "**x g**" where g represents the force of gravity. RCF can also be determined from the nomograph below. Place a straight edge to intersect the radius and the desired RCF to calculate the needed rpm. Alternatively place the straight edge on the radius and the rpm to calculate the g-force. For example, spinning a sample at 2500 rpm in a rotor with a 7.7 cm radius results in a RCF of 550 x g.



PART II

Analysis and Characterization of Proteins

Topics covered:

- Protein Structure and Assays
- Differential Solubility
- Chromatography
- Membranes and Detergents
- Electrophoresis
- Overview on Protein Purification

CHAPTER 7--INTRODUCTION TO PROTEINS

Proteins typically make up more than half the dry weight of cells. They contribute to the structure of a cell and are responsible for cellular functions such as catalysis and molecular recognition.

L-α-AMINO ACIDS

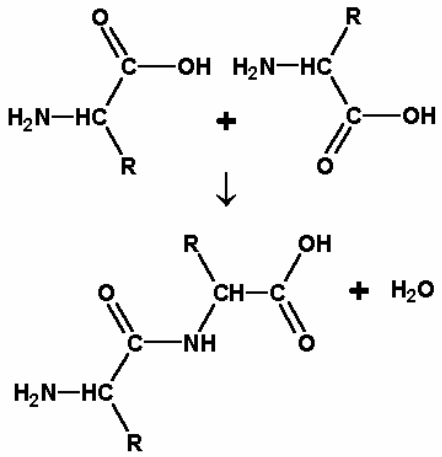
Nonpolar	Polar
Alanine	Arginine
Glycine	Asparagine
Isoleucine	Aspartic a.
Leucine	Cysteine
Methionine	Glutamic a.
Phenylalanine	Glutamine
Proline	Histidine
Tryptophan	Lysine
Valine	Serine
	Threonine
	Tyrosine

PROTEIN STRUCTURE

Proteins are polymers of L-α-amino acids. The α refers to a carbon with a primary amine, a carboxylic acid, a hydrogen and a variable side-chain group, usually designated as 'R'. Carbon atoms with four different groups are asymmetric and can exhibit two different arrangements in space due to the tetrahedral nature of the bonds. The L refers to one of these two possible configurations the four different groups on the α-carbon can exhibit. Amino acids of the D-configuration are not found in proteins and do not participate in biological reactions.

Twenty different amino acids, distinguished by their side-chain groups, are found in proteins (Box). The side-chain groups vary in terms of their chemical properties such as polarity, charge and size. These various side-chain groups will influence the chemical properties of proteins as well as determine the overall structure of the protein (see Appendix). For example, the polar amino acids tend to be on the outside of the protein where they interact with water and the nonpolar amino acids are on the inside forming a hydrophobic core.

The covalent linkage between two amino acids is known as a peptide bond. A peptide bond is formed when the amino group of one amino acid condenses with the carboxyl group of another amino acid to form an amide (Figure). This arrangement gives the polypeptide chain a polarity in that one end will have a free amino group, called the N-terminus, and the other end will have a free carboxyl group, called the C-terminus.



Peptide bonds tend to be planar which gives the polypeptide backbone some rigidity. However, rotation can occur around both of the α-carbon bonds resulting in a polypeptide backbone with different potential conformations in regards to the relative positions of the R-groups. (Conceptually this can be viewed as the R-groups projection into or out from the page in the figure.) Although many conformations are theoretically possible, interactions between the R-groups will limit the number of potential conformations and proteins tend to only fold into a single functional conformation. In other words, the conformation or shape of the protein is due to the interactions of the side-chain groups with one another and with the polypeptide backbone. The interactions can be between amino acids that

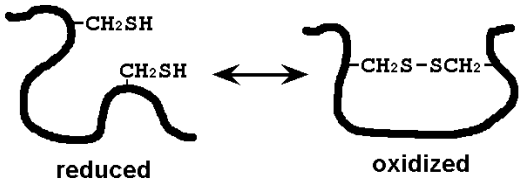
are close together in a polypeptide or between amino acids that are far apart or even on different polypeptides. These different types of interactions are often discussed in terms of primary, secondary, tertiary and quaternary protein structure (Table).

Levels of Protein Structure

Primary	Refers to the amino acid sequence and the location of disulfide bonds between cysteine residues (i.e., covalent bonds).
Secondary	Refers to interactions between amino acids that are close together (eg., α-helix, β-sheet, β-turn, random coil).
Tertiary	Refers to interactions between amino acids that are far apart (eg., motifs, domains).
Quaternary	Refers to interactions between two or more polypeptide chains (i.e., protein subunits).

The primary amino acid sequence and positions of disulfide bonds strongly influence the overall structure of protein. In regarads to the primary amino acid sequence, certain side-chains will permit, or promote, hydrogen-bonding between neighboring amino acids of the polypeptide backbone resulting in secondary structures such as β-sheets or α-helices. Alternatively, certain R-groups may interfere with each other and prevent certain conformations.

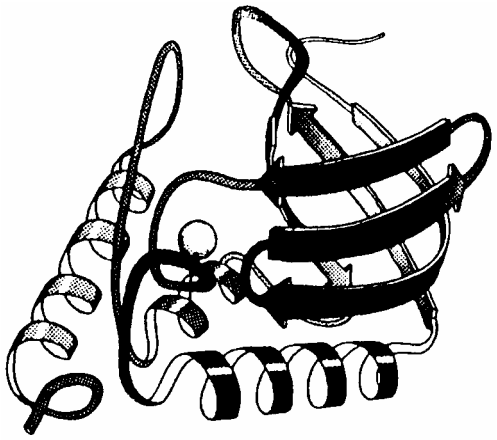
In the α-helix conformation the peptide backbone takes on a 'sprial staircase' shape which is stabilized by H-bonds between carbonyl and amide groups of every fourth amino acid residue. This restricts the rotation of the bonds in the peptide backbone resulting in a rigid structure. β-sheets are also rigid structures in which the polypeptide chain is nearly fully extended with the R-groups alternating between pointing up and down. β-sheets interact either in parallel (both with same orientation in regards to N- and C-termini) or anti-parallel fashion. Certain amino acids promote the formation of either α-helices or β-sheets due to the nature of the side-chain groups. Some side-chain groups may prevent the formation of secondary structures and result in a more flexible polypeptide backbone, which is often called random coil conformation.



The other aspect of primary protein structure is the position of disulfide bonds. The amino acid cysteine has a free thiol group that can be oxidized to form a covalent bond with another cysteine (Figure). These disulfide bonds can form between cysteine residues that are relatively close or far apart within a single polypeptide chain, or even between separate polypeptide subunits with a protein. In this regard, disulfide bonds can contribute to secondary, tertiary and quaternary aspects of protein structure. Proteins containing disulfide bonds will be sensitive to reducing agents (such as β-mercaptoethanol) which can break the disulfide bond.

The various secondary structures can interact with other secondary structures within the same polypeptide to form motifs or domains (i.e., tertiary structure). A motif is a common combination of secondary structures and a domain is a portion of a protein that folds independently. The tertiary structure will represent the overall three dimensional shape of a

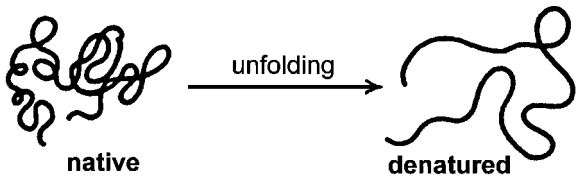
polypeptide. A typical protein structure is a compact entity composed of the various secondary structural elements with protruding loops of flexible (ie, random coil) sequence. This is often depicted in a ribbon diagram (Figure) in which β-sheets are drawn as flat arrows with the arrowhead representing the N-terminal side and α-helices are drawn as flat spirals. The flexible loops are represented by the strings connecting the secondary structural elements.



Many proteins are composed of multiple subunits, or distinct polypeptide chain that interact with one another. This is referred to as quaternary structure.

PROTEIN STABILITY

Proteins are often fragile molecules that need to be protected during purification and characterization. Protein denaturation refers the loss of protein structure due to unfolding. Maintaining biological activity is often important and protein denaturation should be avoided in those situations. Elevated temperatures, extremes in pH, and changes in chemical or physical environment can all lead to protein denaturation (Table). In general, things that destabilize H-bonding and other forces that contribute to secondary and tertiary protein structure will promote protein denaturation. Different proteins exhibit different degrees of sensitivity to denaturing agents and some proteins can be re-folded to their correct conformations following denaturation.



Factors Affecting Protein Stability

Factor	Possible Remedies
temperature	Avoid high temperatures. Keep solutions on ice.
freeze-thaw	Determine effects of freezing. Include glycerol in buffers. Store in aliquots.
physical denaturation	Do not shake, vortex or stir vigorously. (Protein solutions should not foam.)
solution effects	Mimic cellular environment: neutral pH, ionic composition, etc.
dilution effects	Maintain protein concentrations > 1 mg/ml as much as possible.
oxidation	Include 0.1-1 mM DTT (or β-ME) in buffers.
heavy metals	Include 1-10 mM EDTA in buffers.
microbial growth	Use sterile solutions, include anti-microbials, and/or freeze.
proteases	Include protease inhibitors. Keep on ice.

The optimal conditions for maintaining the stability of each individual protein need to

be determined empirically. In general, though, protein solutions should be kept cold (< 4^o C) except during assays and other procedures requiring specific temperatures. Many proteins are especially labile and need to be stored at -20^o or -80^o. However, repeated freezing and thawing of protein solutions is often deleterious. Adding 50% glycerol to storage buffers will lower the freezing point and allow storage at -20^o. Solutions for working with proteins will often contain heavy-metal chelators and/or antioxidants as protectants.. In addition, proteases may be released during cell disruption and it may therefore be necessary to include protease inhibitors.

PROTEIN ASSAYS

Numerous spectrophotometric methods have been developed to estimate the amount of protein in a sample. Proteins are chromophores with absorption maximum in the UV range. Some proteins, such as cytochromes and hemoglobin, will have distinct spectral characteristics due to prosthetic groups. In addition, several indirect ways to measure protein concentrations spectrophotometrically have been developed.

UV Absorption

A simple method to measure protein concentration is to determine the absorption at 280 nm. Tyrosine and tryptophan residues which have A_{max} at 275 and 280, respectively, are responsible for this absorption. The distribution of tyrosine and tryptophan is fairly constant among proteins so it is not absolutely necessary to determine an extinction coefficient for each individual protein. Typically, a 1 mg/ml protein solution will result in an A_{280} of approximately one (1). In the case of purified proteins the exact extinction coefficient will depend on the exact amount of tyrosine and tryptophan in that particular protein. For example, a 1 mg/ml IgG solution has an A_{280} of approximately 1.5. The simplicity and ability to completely recover the sample are the major advantages of this method. The lower limit of sensitivity for UV absorption is 5-10 μg/ml.

A potential problem with using A_{280} values to calculate protein concentration is the absorption due to contaminating substances, and in particular, nucleic acids which have an A_{max} at 260 nm. It is possible to use correction factors that permit the determination of protein concentrations. A particularly convenient formula is:

$$(A_{235} - A_{280})/2.51 = \text{mg/ml protein}.$$

Indirect spectrophometric assays (eg., Lowry, Bradford) for the determination of protein concentrations overcome some of the problems associated with interfering substances in protein samples. However, the measured protein cannot be recovered in such assays and they take longer to perform.

Folin-Ciocalteu or Lowry

Historically, one of the most widely used protein assays was the Lowry assay. This assay is a modification of a previous assay known as the Biuret. In the Lowry assay proteins

react with alkaline Cu^{2+} reducing it to Cu^{+}. The reduced Cu^{+} and the side-chain (R) groups of tryptophan, tyrosine and cysteine react with the Folin-Ciocalteu reagent (complex of inorganic salts) to form a blue color that is proportional to the amount of protein. The $A_{600\text{-}750}$ is determined and protein concentration is calculated from a standard curve. The assay is linear 1-300 μg and the lower limit of sensitivity is 1-5 μg/ml. Substances in the sample such as detergents can interfere with the results and therefore appropriate controls and blanks need to be carried out.

Bradford or Coomassie Blue G-250

The Bradford assay has replaced the Lowry as the standard protein assay. The major advantage is that it is carried out in a single step and that there are very few interfering substances. The principle of the assay is based on a shift of the A_{max} of the Coomassie-blue (G-250) dye from 465 nm to 595 nm in the presence of protein due to a stabilization of the anionic form of the dye. The dye reacts primarily with arginine residues and to lesser extent with his, lysine, tyrosine, tryptophan and phenylalanine. Protein concentrations are determined by developing a standard curve with known amounts of proteins and extrapolating the absorbance values of the samples. The standard curves are not linear over a wide range of protein concentrations. This assay can also carried out in 96-well plates and read on automated ELISA readers. Programs are available that will automatically calculate the protein concentrations based upon a standard curve.

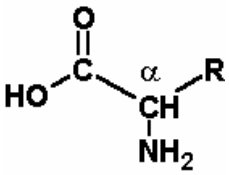
- Specificity
- Sensitivity
- Accuracy (Quantification)
- Rapid
- Easy to Perform

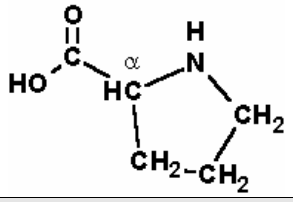
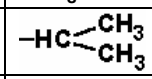
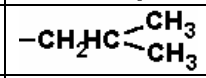
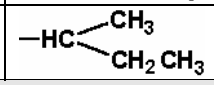
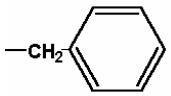
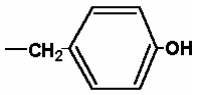
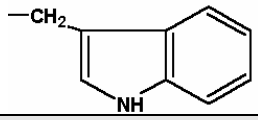
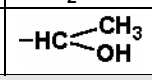
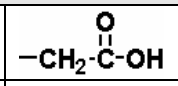
Assay Of Specific Proteins

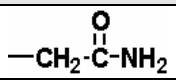
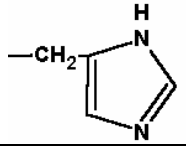
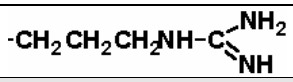
In addition to measuring the total amount of protein, it is often necessary to estimate the amount of a specific protein in a mixture of proteins. Measuring a specific protein will depend upon the availability of an assay that is specific for the protein of interest. Protein assays should be practical in addition to being specific and accurate (Box). Typically protein assays are based upon the biological activity of the protein of interest. For example, enzyme assays will detect the conversion of a substrate to a product. Enzymes assays can be based upon colorimetric, fluorescent or radioactive substrates (or products). Many proteins bind to ligands or other substances and this binding activity is measured. Bioassays measure a change in some biological property (eg., stimulation of cell division). In cases where the protein of interest has no measurable activity or the activity is unknown it may be possible to generate antibodies against the protein and develop an immunoassay (to be discussed later). If antibodies against such a protein are not available, the assay may simply be the amount of a protein band on a Commassie blue-stained gel following electrophoresis (to be discussed later).

APPENDIX. AMINO ACIDS: PROPERTIES AND STRUCTURE

Proteins are composed of 20 different amino acids which are distinguished by their side-chain groups, designated as R (see figure at right). The amino acids are grouped below according to the chemical nature of their side-chain groups.



Name/Abbreviations			R-group	Comments
Glycine	Gly	G	-H	The lack of a side chain provides for the greatest possible conformational flexibility.
Proline	Pro	P		The aliphatic side-chain forms a cyclic imino compound (entire amino acid is shown). This imposes rigid constraints on the rotation around the backbone and has significant affects on protein conformation.
Amino Acids with Aliphatic Side Chains				
Alanine	Ala	A	$-CH_3$	These alkyl side-chain groups are important for hydrophobic interactions and provide for a variety of surfaces and shapes.
Valine	Val	V		
Leucine	Leu	L		
Isoleucine	Ile	I		
Amino Acids with Aromatic Side Chains				
Phenylalanine	Phe	F		The aromatic groups are important for hydrophobic interactions and may be especially important for interacting with other flat molecules.
Tyrosine	Tyr	Y		
Tryptophan	Trp	W		
Amino Acid Alcohols				
Serine	Ser	S	$-CH_2OH$	The hydroxyl groups are weakly ionizable (pK_a~13) and participate as active groups in some enyzmes.
Threonine	Thr	T		
Acidic Amino Acids				
Aspartatic acid	Asp	D		These side-chain groups are generally negatively charged at neutral pH (pK_a = 4.3-4.7).
Glutamatic acid	Glu	E	$-CH_2CH_2CO_2H$	

Amides of the Acidic Amino Acids				
Asparagine	Asn	N		These side-chain groups do not ionize, but are relatively polar.
Glutamine	Gln	Q	$-CH_2CH_2CONH_2$	
Basic Amino Acids				
Histidine	His	H		The imidazole group participates in the active site of many enzymes, as well as binding metal ions.
Lysine	Lys	K	$-CH_2CH_2CH_2CH_2NH_2$	These side-chain groups are generally positively charged at neutal pH (pK_a >10).
Arginine	Arg	R		
Sulfur Containing Amino Acids				
Cysteine	Cys	C	$-CH_2SH \leftrightarrow -CH_2S-SCH_2-$	Cysteines participate in redox reactions and can form disulfide links between two residues (i.e., oxidized state).
Methionine	Met	M	$-CH_2CH_2SCH_3$	Met is rather hydrophobic, but the thioether group is a potent nucleophile.
Other Abbreviations				
Asp or Asn	Asx	B	The amides are converted to the acids by the procedures used to identify and quantify amino acid residues. Therefore they are designated as either in such situations..	
Glu or Gln	Glx	Z		
Any amino acid	Xaa	X	Used when the amino acid is unknown or does not matter.	

CHAPTER 8--DIFFERENTIAL PRECIPITATION OF PROTEINS

Proteins are responsible for many aspects of cell function and structure. To learn more about their precise roles in cellular biology it is often necessary to analyze purified proteins. A protein of interest can be separated from other proteins based upon its unique physical and chemical properties. The first step in protein isolation is to extract the protein from the tissue or cell in a soluble form. In some cases it is desirable to first disrupt the cell and to isolate the appropriate subcellular compartment before solubilizing the protein. Generally, cytosolic proteins are found in the supernatant following cell disruption and centrifugation at 100,000 xg for 30-60 min. Membrane proteins or proteins associated with organelles will require additional solubilization techniques such as detergents or chaotropic agents (see Chapter 10).

Proteins are soluble in aqueous solutions because the charged and polar side-chain groups of the amino acids interact with water. In other words, proteins become hydrated. If the protein-solvent interaction is prevented, proteins will interact with one another and form aggregates that precipitate out of solution. Different proteins will behave differently in terms of interacting with either the solvent or other protein molecules. Therefore it is possible to separate different proteins based upon different solubility properties. For example, as the salt concentration of a solution is increased the amount of water that is available to interact with proteins is decreased. This will result in the more hydrophobic domains of proteins interacting with hydrophobic domains of other proteins. The interacting proteins will form large masses, or a precipitate, that can be collected by centrifugation. Proteins with a larger proportion of hydrophobic amino acid residues exposed on their surface will be more sensitive to effects of salts than highly charged proteins. Therefore, it is possible to selectively precipitate some proteins under conditions where other proteins remain soluble. One of the most common salts used for this differential precipitation of proteins is $(NH_4)_2SO_4$.

PROCEDURE FOR $(NH_4)_2SO_4$ PRECIPITATION.

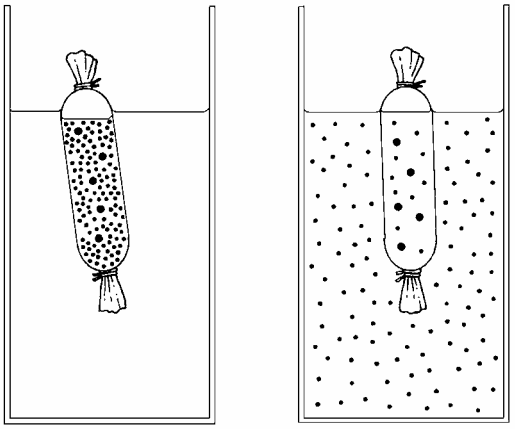
A common method for differentially precipitating proteins is by 'salting-out' with $(NH_4)_2SO_4$. As the $(NH_4)_2SO_4$ is added, H_2O molecules interact with the salt ions, thus decreasing the amount of water available to bind to protein. Proteins are differentially precipitated by $(NH_4)_2SO_4$ due to their differences in hydrophobicity. A more hydrophobic protein will precipitate at a lower salt concentration than a more hydrophilic protein.

1. The $(NH_4)_2SO_4$ is slowly added to a protein solution while gently stirring. The amount to add to precipitate the protein of interest will have to be initially determined empirically. The concentration of $(NH_4)_2SO_4$ needed to precipitate a particular protein in usually expressed as a percent of saturated $(NH_4)_2SO_4$. Either solid or a saturated solution can be added. For large volumes, adding solid is more convenient.

2. Once dissolved, the $(NH_4)_2SO_4$/protein suspension is allowed to slowly stir (usually while kept cold) until equilibrium is reached. This is typically for one hour, however, some proteins require overnight incubations. The precipitated protein are collected by centrifugation. Typically 10,000 xg for 10 minutes is sufficient to pellet precipitated proteins.

3. The protein pellet is dissolved in the buffer of interest and the remaining $(NH_4)_2SO_4$ is removed by either dialysis or desalting column. Dialysis tubing has defined pore size which allows small molecules to pass through freely while retaining larger molecules. Multiple buffer changes accelerates the process.

$(NH_4)_2SO_4$ precipitation can also be carried out as a two-step procedure. The maximum amount of $(NH_4)_2SO_4$ that does not precipitate the protein of interest is added and this pellet is discarded. $(NH_4)_2SO_4$ is then added to the supernatant to bring the concentration to the minimal % saturation that completely precipitates the protein of interest. Nomograms for calculating the amounts of $(NH_4)_2SO_4$ needed to achieve the desired percent saturation are available (see appendix). The pellet is retained and the supernatant discarded at this step.



Another variation of $(NH_4)_2SO_4$ precipitation is 'back extraction'. In this case the total protein is precipitated with 80-90% saturated $(NH_4)_2SO_4$ and the pellet is extracted with the appropriate $(NH_4)_2SO_4$ concentration to solubilize the protein of interest.

- salt (eg, $(NH_4)_2SO_4$
- solvents
- acidic pH
- high temperature

OTHER PROTEIN PRECIPITATION METHODS

Organic solvents can also be used to differentially precipitate proteins. Water miscible solvents (eg., acetone, ethanol, methanol) are mixed with an aqueous solution. As the water molecules interact with the organic solvent molecules less water is available for hydration and proteins will differentially precipitate as in the case of added salts.

Proteins also tend to precipitate out at high temperatures or at extreme pH values. Proteins will differentially denature under conditions of high temperature or extreme pH. The unfolded, or denatured, proteins will have hydrophobic residues exposed and will aggregate. Most of the time precipitation with high temperatures or acidic pH results in precipitated protein which can not be resolubilized. However, it is possible to remove unwanted proteins in cases where the protein of interest is acid or heat stable.

APPENDIX. AMMONIUM SULFATE NOMOGRAM

Initial Concentration of Ammonium Sulfate (% saturation)	Final Concentration of Ammonium Sulfate (% saturation)													
	20	25	30	35	40	45	50	55	60	65	70	75	80	90
	Grams of solid $(NH_4)_2SO_4$ to be added to 1 liter of solution.													
0	114	144	176	209	243	277	313	351	390	430	472	516	561	662
20		29	59	91	123	155	189	225	262	300	340	382	424	520
25			30	61	93	125	158	193	230	267	307	348	390	485
30				30	62	94	127	162	198	235	273	314	356	449
35					31	63	94	129	164	200	238	278	319	411
40						31	63	97	132	168	205	245	285	375
45							32	65	99	134	171	210	250	339
50								33	66	101	137	176	214	302
55									33	67	103	141	179	264
60										34	69	105	143	227
65											34	70	107	190
70												35	72	153
75													36	115
80														77

Nomogram for determining the amount of ammonium sulfate which will yield the desired % saturation of $(NH_4)_2SO_4$. A saturated $(NH_4)_2SO_4$ solution is 4.1 M and 3.9 M at 25^o and 0^o, respectively. Find the appropriate initial and desired final concentrations of $(NH_4)_2SO_4$ on the vertical and horizontal scales, respectively. The point of intersection is the number of grams needed per liter of solution. Multiply this number by the number of liters and add that amount of $(NH_4)_2SO_4$ to the solution. (Modified from Methods of Enzymology 1:76).

Example: you have 100 ml of protein solution and would like to remove the proteins precipitate at 35% saturation $(NH_4)_2SO_4$ and collect the remaining proteins precipitated at 60%. Add 20.9 grams of $(NH_4)_2SO_4$ to the 100 ml. Centrifuge at 10,000 x g for 10-20 minutes, discard the pellet and determine the volume of the supernatant. Multiply 164 by the volume in liters and add that number of grams of $(NH_4)_2SO_4$ to the solution. Centrifuge as above, discard the supernatant and dissolve the pellet in appropriate volume of buffer.

CHAPTER 9--CHROMATOGRAPHY

Grossly dissimilar molecules are relatively easy to separated. For example, lipids, proteins and DNA can usually be separated from one another based on differences in solubility in various solvents. Separation of substances with similar chemical and physical properties is more complex and subtle. Although individual proteins are unique in terms of their structures, the overall chemical and physical properties are somewhat similar in that they are all polymers of amino acids. Therefore differential solubility has a limited ability to separate proteins. Chromatography provides a means to refine the separation of substances.

BASIC PRINCIPALS

The basis of chromatography is to place substances to be separated into a system with two phases: a **mobile phase** and a **stationary phase**. Substances are then separated based upon their differential interaction with these two phases as the mobile phase moves across the stationary phase. In the case of liquid chromatography the mobile phase is a solvent. Molecules of interest (called the **solute**) are dissolved in the solvent and the solvent then flows across a solid matrix (i.e., the stationary phase). Solutes interact with the stationary phase by reversibly binding to the stationary phase. The strength of the binding between the solute and the stationary phase will determine how fast the solute is carried by the mobile phase. For example, substances which do not bind or interact with the solid phase will be carried unimpeded by the solvent. Whereas substances that interact with the solid phase will be temporarily retained. Therefore two substances that interact with the solid phase to different degrees can be separated from one another.

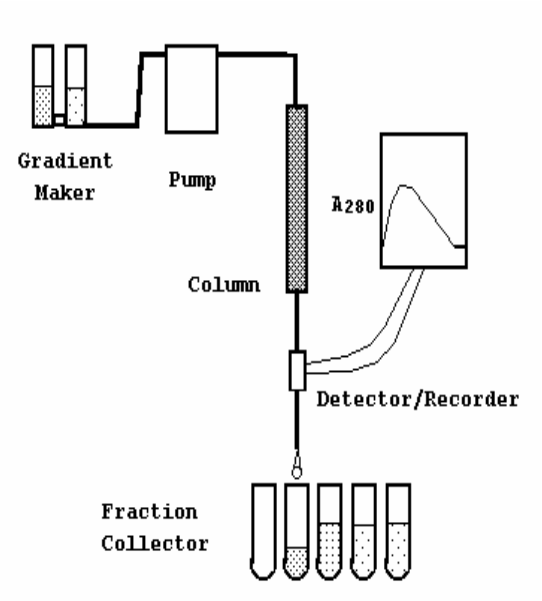
In most applications the mobile phase is liquid. The major exception is gas-liquid chromatography in which the mobile phase is a gas and the stationary phase is a liquid absorbed to a solid support. In liquid chromatography, the stationary phase can be in a column configuration or in a thin layer. The column is probably the most common way to hold the solid support and is especially convenient for preparative work such as the isolation of the solute. In particular, proteins are generally isolated by column chromatography. The column is a cylinder or tube which holds the solid phase matrix and the liquid phase is passed through this column.

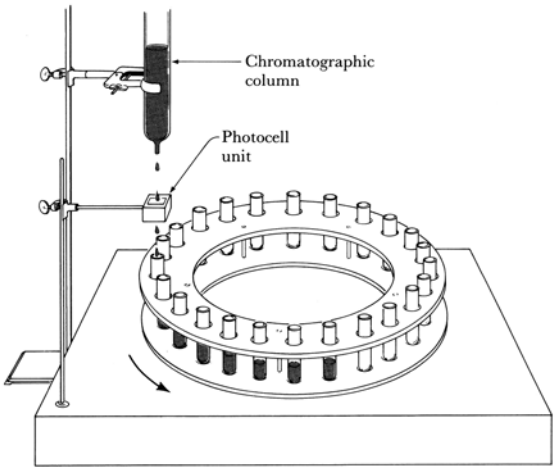
There are several distinct types of solid phases used in the isolation and analysis of proteins (see Table). These various types of solid supports will separated proteins based upon different chemical and physical properties (discussed below).

CHROMATOGRAPHY	DISCRIMINATION
Ion Exchange	Charge
Gel Filtration	Size and Shape
Hydrophobic	Surface Hydrophobicity
Reverse Phase	Total Hydrophobicity
Affinity	Specific Amino Acids
Adsorption	Amino Groups?

EQUIPMENT

The equipment needed for column chromatography can range from expensive workstations to pasteur pipettes. The central component is the **column**. A **pump** may be needed to control the flow rate of buffers through the column. However, gravity can also be used. The solvent used in the mobile phase will often need to gradually change. This gradual change is accomplished by a **gradient maker**. The elution of substances can be monitored during chromatography with an in line spectrophotometer which measures the absorbance of material coming off the column. In the case of protein chromatography the **detector/recorder** monitors the A_{280}. Detectors to record radioactivity or fluorescence during chromatography are also available.





If chromatography is used as part of a protein purification scheme a **fraction collector** is needed. A fraction collector is a device that will automatically collect the liquid flowing from the column in separate tubes. Most fraction collectors will allow fractions to be collected per unit time or per unit volume (i.e., number of drops). In either case, the end result is a series of tubes containing approximately equal volumes. The tubes can then be evaluated for the substances of interest. The amounts of the substances being measured are often plotted against either the fraction number, volume or time. These three variables can be easily inconverted from the flow rate and volume of the individual fractions. Some fraction collectors can also be interfaced with the detector/recorder and programmed to collect only the peaks.

High Performance (Pressure) Liquid Chromatography. HPLC is not a distinct chromatographic technique, but an advancement of technology. The flow rates in conventional chromatography is limited because of the compression of the support matrices used in the columns. These low flow rates result in diffusion and loss of resolution. New resins that can withstand packing and high flow rates allow for higher resolution have been developed. This allows for separations to be carried out under higher pressures (i.e., high flow rates) resulting in increased resolution. All of the same types of chromatographic media

are available for HPLC as in conventional chromatography. HPLC is more widely used for the separation of small molecules but can be applied to the separation of proteins in some applications.

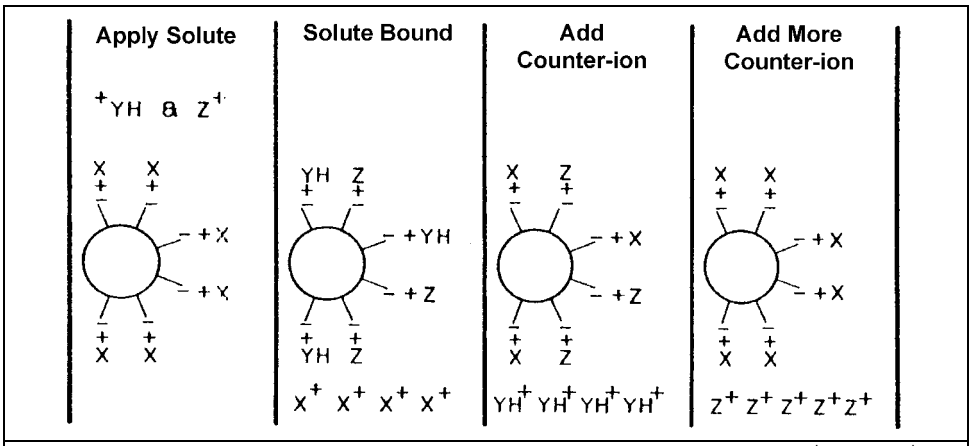
Fast protein liquid chromatography (FPLC) is a similar concept as HPLC, but specificially designed for protein separations. FPLC also uses special columns and pumps to achieve high flow rates and therefore faster separations. In general the flow rates obtained with FPLC are not as great as those achieved with HPLC.

ADSORPTION CHROMATOGRAPHY

The basic principal of chromatography discussed above refers specifically to adsorption chromatography. Thin-layer chromatography (TLC) and paper chromatography are examples of adsorption chromatography. These techniques are usually used to separate small molecules, such as nucleotides, amino acids, lipids, simple carbohydrates, for analytical work. However, in many applications HPLC in a column format is used. Adsorption chromatography is not widely used in protein purification. Hydroxyapatite, a crystalline form of calcium phosphate, is an example of a stationary phase medium that is sometimes used to analyze either proteins or nucleic acids. The basis of protein separation on hydroxyapatite columns is not understood and is difficult to predict based upon a protein's chemical and physical properties. It may involve non-specific interations between the positive calcium and negative carboxyl groups on the protein and the negative phosphate and positive amino groups on the protein. Chromatography on hydroxyapatite columns has been used successfully in some cases to separate proteins that were not readily separable by other techniques.

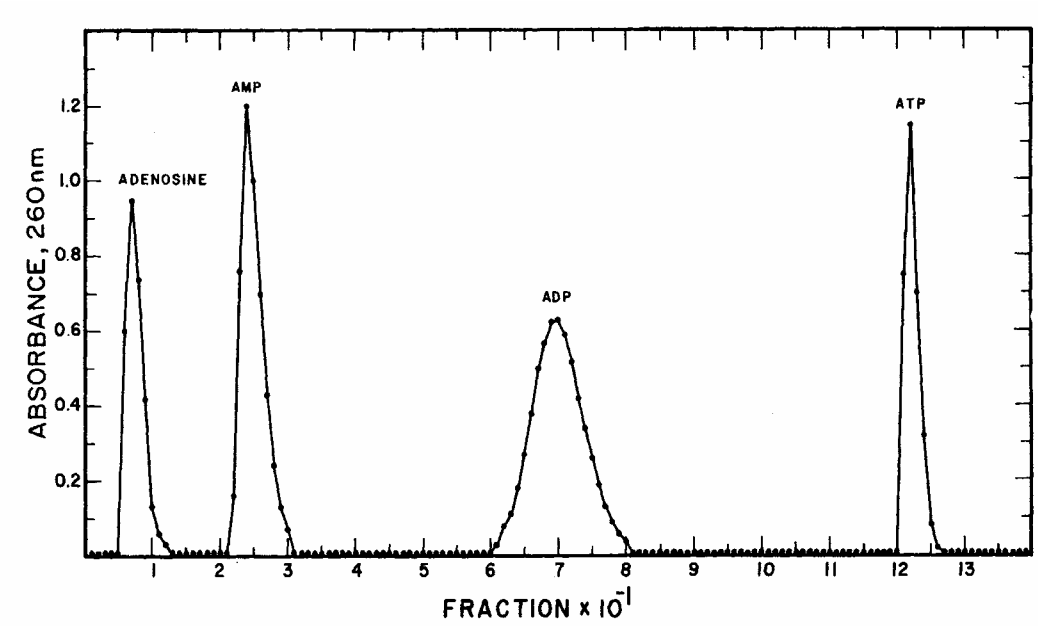
ION EXCHANGE CHROMATOGRAPHY

Ion exchange chromatography (IEC) is a specific type of adsorption chromatography based upon charge-charge interactions. The stationary phase consists of fixed charges on a solid support. The fixed charges on the stationary phase can be either negative or positive and are respectively referred to as **cation exchange** or **anion exchange** chromatography. Counter ions will interact with the fixed charged groups and can 'exchange' with solute molecules. In other words, substances to be separated will replace the counter ions associated with the chromatography medium and stably bind to the exchanger via electrostactic interactions. Conditions in which some solute molecules are electrostatically bound to the exchanger and other solute molecules are not bound can be used to separate solutes (Figure).



Basic Principal of IEC. A solution containing the solutes of interest ($^{+}$YH & Z^{+}) is applied to the ion exchanger. The solutes bind to the exchanger and displace the counter-ions (X^{+}) already bound to the column. The solute is removed from the stationary phase by increasing the concentration of counter-ions or by changing the pH to affect the charge of the solute. Weakly charged solutes ($^{+}$YH) will be displaced from the stationary phase with lower concentrations of counter-ions than more highly charged solutes (Z^{+}). This results in separation of solutes based upon their net charges

For example, mixtures of adenine nucleotides (i.e., adenosine, AMP, ADP and ATP) can be separated by IEC. Adenosine, which is uncharged does not bind to the anion exchanger. As an increasing concentration of formate is applied to the column, AMP is the first to elute followed by ADP and then ATP. The order of elution corresponds to the overall negative charge of the nucleotides (i.e., number of phosphate groups).

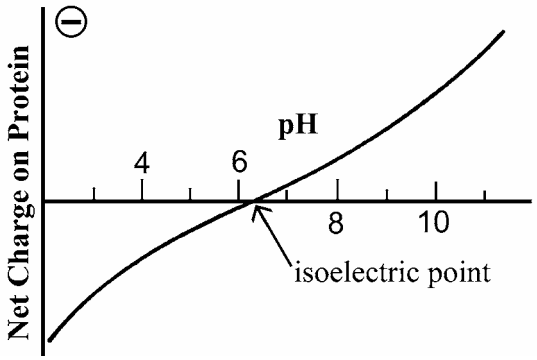


Proteins are complex **ampholytes** (have both - and + charges). Negative charges are due to aspartic acid and glutamic

Amino a.	pK_a
Asp, Glu	4.3-4.7
His	≈ 7
Lys, Arg	> 10

acid residues and positive charges are due arginine, lysine and histidine residues. (See Appendix of Introduction to Proteins for structure of side-chain groups.) Each of these side-chains functions as either a weak acid or weak base and has a pK_a value (see Table). The exact pK_a for the side-chain groups will depend on the position of the residue within the protein. The side-chains of aspartate and glutamate have a carboxylic acid group that will either be negatively charged or uncharged depending upon the pH. At pH values greater than the pK_a the carboxylic acid group will be deprotonated, or negatively charged, whereas below the pK_a the carboxylic acid will be protonated and thus have no charge. Similarly, the side-chains of the basic amino acids have an amine group that will be positively charged (i.e., protonated) at pH values below the pK_a or uncharged (i.e. deprotonated) at pH values above the pK_a.

The combination of all of the charged side chains will give the protein a net charge which depends on the pH. The **isoelectric point** (Figure) of a protein is the pH at which the net charge is zero (i.e., the number of positive and negative charges are equal). Therefore, proteins will have either a net negative charge or net positive charge depending upon the isoelectric point of the protein and the pH of the solution, and thus, it is possible to use either anion exchange or cation exchange chromatography. Anion exchange chromatography is generally carried out above the isoelectric point of the protein of interest and cation exchange chromatography is carried out below the isoelectric point.

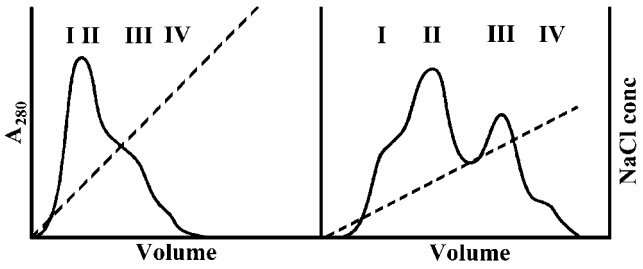


The functional groups in IEC can be either strong or weak acids for cation exchange or strong or weak bases for anion exchange. A strong exchanger is used for the separation of weakly ionizable groups and a weak exchanger is usually used for more highly charged solutes. In addition, weak exchangers result in better resolution when the charge differences between solutes is small. IEC of proteins is most often carried out with weak exchangers. The matrix should not have ionizable groups or bind proteins. Common matrices used in IEC are dextran (trade name Sephadex), agarose (trade name Sepharose), and cellulose are common matrices used in protein chromatography.

The basic steps in column chromatography are to prepare the column, apply the sample to the column, and to elute the solute from the column. Prepared IEC columns can be purchased or alternatively the ion exchange media is purchased and prepared. The amount of preparation necessary will depend upon the form of the media and the manufacturer's instructions should be followed. In general, short wide columns work best for IEC and that the total column volume should be such that all of the protein binds in the top 1-2 cm of the column. The sample should be loaded under conditions (i.e., appropriate pH and low ionic strength) which promote the binding of the protein of interest. Proteins that do not bind to the column are washed away using the same buffer.

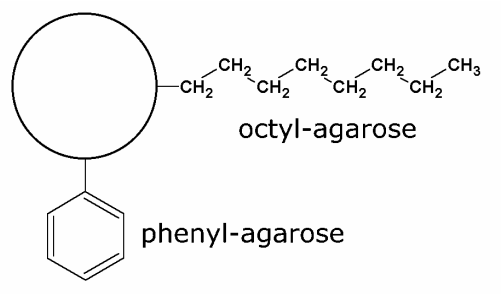
Proteins are eluted by increasing the ionic strength or changing the pH. Increasing the ionic strength can be done batchwise by sequentially washing with buffers containing higher

concentrations of salt or gradually using a gradient maker. Eluting in a 'steps' is generally simpler, but of lower resolution. A gradient maker will mix two buffers of different ionic strengths resulting in controlled increase in the ionic strength of the elution buffer. The slope and shape of the gradient will determine the resolution in regards to separating solutes. For example, decreasing the slope of the gradient will lead to a greater separation of the solutes (see Figure). However, as the slope decreases the proteins will be more dilute.



Chromatofocusing is a specialized form of ion-exchange chromatography which separates proteins according to their isoelectric points. This method takes advantage of the buffering capacity of the exchanger and generates a pH gradient across the column. The column is developed with a mixture of buffers called the 'polybuffer'. As these migrate down the column, the most acidic components will bind to the basic groups of the exchanger and lead to an increase in the local $[H^+]$. As elution progresses the pH as each point in the column is gradually lowered due to the addition of more buffer components. The more acidic components will elute, migrate further down the column and bind to again to the basic exchanger groups. This will result in a pH gradient across the length of the column. The sample is loaded onto the column at a pH > than the pI of the solute of interest. As the pH drops the protein elutes and migrates down the column until the pH is greater than the isoelectric point and then reabsorbs. This will continue until the proteins elute at their isoelectric points.

HYDROPHOBIC CHROMATOGRAPHY



Proteins can be separated according to differences in their hydrophobicities. Hydrophobicity is a chemical property which promotes the aggregation of nonpolar compounds with each other in an aqueous environment. These hydrophobic interactions are not an attractive force per se, but are forced upon nonpolar compounds in a polar environment. The media for hydrophobic chromatography is a support matrix such as agarose with long chain hydrocarbons covalently bound. Two common examples are octyl-agarose (8 contiguous methyl groups) and phenyl-agarose (Figure). This will provide a hydrophobic surface for proteins to interact with instead of aggregating with each other. Highly hydrophobic resins like octyl-agarose are best for weakly hydrophobic proteins, whereas less hydrophobic resins like phenyl-agarose are better for proteins of intermediate hydrophobicity.

Proteins bind to hydrophobic columns under conditions that promote hydrophobic interactions and these conditions will determine the extent of binding. For example, raising the ionic strength increases hydrophobic interactions, or the 'salting out' effect (see chapter on Differential Solubility). Both anions and cations can be listed in a series in terms of either promoting hydrophobic interactions or increasing the chaotropic effect. **Chaotropic agents** disrupt the structure of H_2O by decreasing H-bonding and therefore decrease hydrophobic interactions. The most common salt used in hydrophobic chromatography is $(NH_4)_2SO_4$. The $(NH_4)_2SO_4$ concentration should not promote protein precipitation (eg., 20% saturated or approximately 1 M). Other salts can be substituted, but higher concentrations may be required (eg., 4 M NaCl).

← increasing salting-out effect

anions: PO_4, SO_4, Cl, Br, NO_3, ClO_4, I, SCN

cations: NH_4, Rb, K, Na, Li, Mg, Ca, B

increasing chaotropic effect →

- decrease ionic strength
- decrease solvent polarity
- chaotropic agents
- detergents

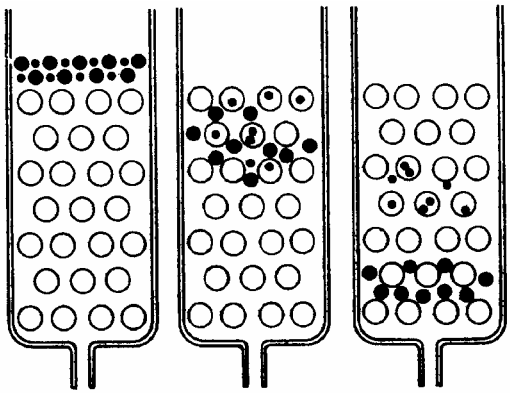
Conditions which decrease hydrophobic interactions are used to elute proteins from hydrophobic columns (Box). Decreasing the ionic strength is generally the preferred method since the other methods introduce substances which may be difficult to remove and/or denature the protein. Decreasing the ionic strength can be done in a step-wise fashion or with gradients as discussed for ion exchange chromatography. Substances which affect the polarity of the solvent (i.e., water) will also affect hydrophobic interactions. For example, a gradient of ethylene glycol will lead to the differential elution of proteins. Similarly, chaotropic agents can be use to elute proteins from hydrophobic columns. Detergents are more hydrophobic than proteins and will complete with the proteins for binding to the hydrophobic matrix.

Reverse phase chromatography (RPC) also separates compounds based upon differences in hydrophobicities. RPC differs from hydrophobic interaction chromatography (HIC) in that the mobile phase is a nonpolar solvent such as hexane instead of an aqueous salt solution. HIC is usually performed under non-denaturing conditions and separates proteins according to differences in surface hydrophobicity. RPC is carried out under denaturing conditions and separates according to differences in the total hydrophobicity, since all of the amino acid residues are available for interaction with the stationary phase. The separation of small polypeptides and proteolytic fragments is a common application of RPC.

HIC vs. RPC

	Hydrophobic	Reverse Phase
Mobile Phase	Polar Solvent	Nonpolar Solvent
Conditions	Native	Denatured
Solute Properties	Surface Residues	Total Residues

GEL FILTRATION CHROMATOGRAPHY



Gel filtration, also called molecular sieve chromatography or size exclusion chromatography, separates proteins on the basis of molecular size. A protein solution is passed over a column made up of small beads composed of cross-linked polymers. The degree of cross-linking will defined a pore size. Solutes larger than this pore size are excluded from the matrix and pass through the column unimpeded. In other words they do not enter the beads but flow around them. Smaller solutes will enter the gel matrix and are retained on the column longer. The retention time is inversely proportional to the size of the solute.

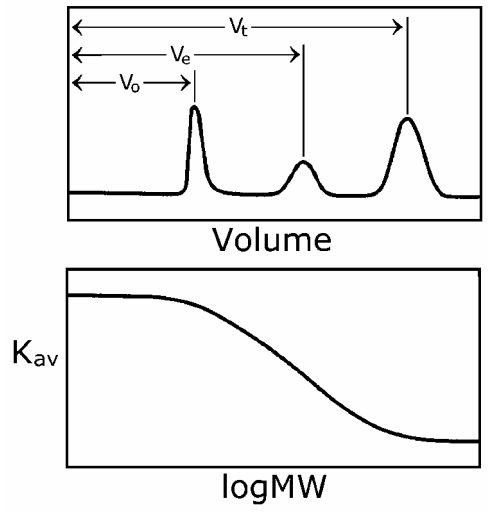
- Dextran (=Sephadex)
- Agarose (=Sepharose)
- Polyacrylamide

Several different polymers have been used as support matrices for gel filtration chromatography (Box). Different grades of gel filtration media with different pore sizes are available. The different pore sizes allow for separation of macromolecules with different size ranges. Gel filtration chromatography is generally carried out in buffers containing 0.15-1.0 M salt to prevent interactions of proteins with the support matrix. Unlike the other chromatographic methods the solute (eg., protein) does not bind to the stationary phase during chromatography. Therefore, gel filtration is a gentle technique and fragile proteins are not damaged by adsorption to the chromatographic support. Since proteins do not bind to the stationary phase, the resolution is dependent upon loading the smallest possible volume of sample.

Gel filtration can also be used to determine the molecular weight of a protein if the columns are calibrated by using molecular weight standards. Proteins of known size are passed over the column and the K_{av} is determined for each protein according to the following equation:

V_o = void volume
V_t = total volume
V_e = elution volume

$$K_{av} = V_e - V_o / V_t - V_o .$$

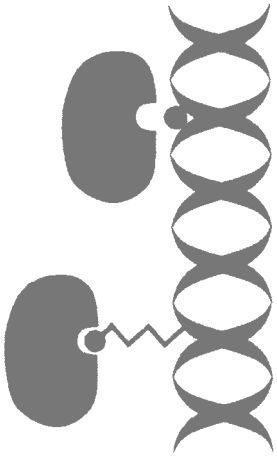


The void volume (V_o), also called the excluded volume, is the elution volume of a substance which is two large to enter the matrix of the support medium. This is experimentally determined with a standand known to be completely excluded and represents the solvent between the beads. The total volume is calculated from the volume of the column bed (πr^2 x length) and represents the elution volume of molecules completely included in the matrix. All substances, assuming there are no interactions with the matrix, will elute in this volume. The K_{av} values are plotted against the log of the molecular weight for each protein standard. The molecular weight of the unknown protein can then be determined from its K_{av} and the standard curve.

Shape also affects the retention of macromolecules. For example, long rod shaped proteins elute from gel filtration columns at apparent molecular weights greater than their actual molecular weights.

Desalting. Gel filtration is often employed to remove salts or other small molecular weight solutes from protein solutions or to carry out buffer exchanges. The sample is loaded onto a gel filtration column that will exclude proteins (eg., G-10 or G-25) and the void volume collected. Desalting columns are much faster than dialysis. However, the sample is usually diluted 2-4 fold. Spin columns, widely used in DNA isolations, are another example of desalting columns.

AFFINITY CHROMATOGRAPHY



The biological function of proteins often involves binding to or interacting with specific ligands, substrates, co-factors, inhibitors, other proteins, etc. Affinity chromatography takes advantage of these specific interactions between proteins and ligands. A protein mixture is passed over the column with an appropriate moiety covalently attached to a solid support. Only the protein which recognizes the ligand of interest will bind to the column and the other proteins will be washed away. Affinity chromatography can be very specific and may allow the isolation of a protein in a single step.

Affinity columns are usually prepared from cyanogen bromide (CNBr)-activated agarose. The CNBr group reacts with primary amines and covalently attaches the ligand to the solid support. In some cases, the ligand already bound to the matrix is commercially available. In addition, linker arms are often used to extend the ligand away from the matrix of the solid support. The matrix should not adsorb contaminating proteins and covalent attachment of ligand to matrix should not alter its binding to protein of interest. Binding of the protein to the ligand should be relatively tight (i.e., high affinity), but at the same time the binding should not preclude the ability to elute the solute from the column.

- affinity elution
- pH changes
- ionic strength
- chaotropic salts
- denaturing agents

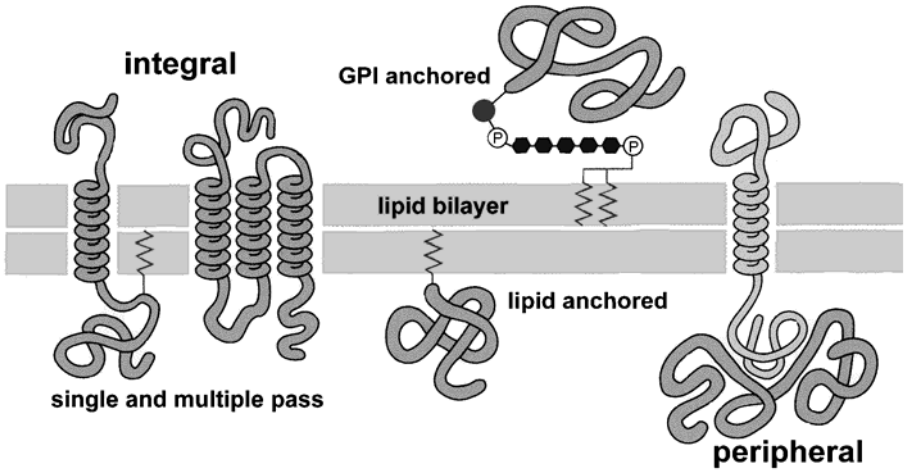
Conditions which promote the dissociation of the protein and ligand are used to elute proteins from affinity columns (box). For example, high concentrations of the free ligand can compete with the bound ligand resulting in elution. In cases where affinity elution is not possible or does not work, it may be possible to change the pH or increase the ionic strength resulting in a destabilization of the protein-ligand interactions. Another possibility is to use denaturing or chaotropic agents to elute the protein which will result in the protein unfolding. However, these later methods can only be used if it is not important to recover protein in the native state.

http://www.science.uts.edu.au/subjects/91326/Section3/section3.html (good summary of chromatography)

CHAPTER 10--MEMBRANES AND DETERGENTS

Membranes are composed of a lipid bilayer and associated proteins. Lipids are amphipathic molecules in that they contain a polar head group and hydrophobic tails. In aqueous solutions lipids will aggregate such that the hydrophobic tails interact with other hydrophobic tails and the polar head groups are exposed to water. The two possible configurations are: 1) a spherical micelles with the hydrophobic tails pointed inward, or 2) a bilayer with the hydrophobic tails sandwiched between the polar head groups. The shape of the lipid and its amphipathic nature cause them to spontaneously form bilayers in aqueous solutions and accounts for the stability of membranes.

Proteins interact with this layer bilayer in several different fashions. Transmembrane proteins pass through the bilayer. A hydrophobic domain, typically an α-helix composed of amino acids with hydrophobic side-chains, interacts with the hydrophobic tails of the lipids and anchors the protein to the bilayer. Some transmembrane proteins will have multiple membrane spanning domains. Other membrane proteins are anchored to the lipid bilayer via fatty acids or lipids that are covalently attached to the protein. Other membrane are attached to the membrane by non-covalent association with other membrane proteins.



Many cellular processes occur on membranes or in membrane-bound subcellular compartments. Approximately half of a cell's total protein is associated with membranes or found in membrane-bound compartments. The study of these membrane associated processes may require the isolation of membranes or membrane proteins. The choice of tissue and membrane fraction (i.e., organelles) will depend in part on the phenomenon being studied. Some cell types are better sources for certain types of membranes. The cells need to be disrupted by procedures which preserved the activity of interest (see Appendix). Disruption of cells or subcellular compartments will usually result in the membranes forming vesicles. The different types of membranes will have different sizes and densities based upon their lipid and protein composition and can be prepared by differential centrifugation, density gradient centrifugation or a combination of the two.

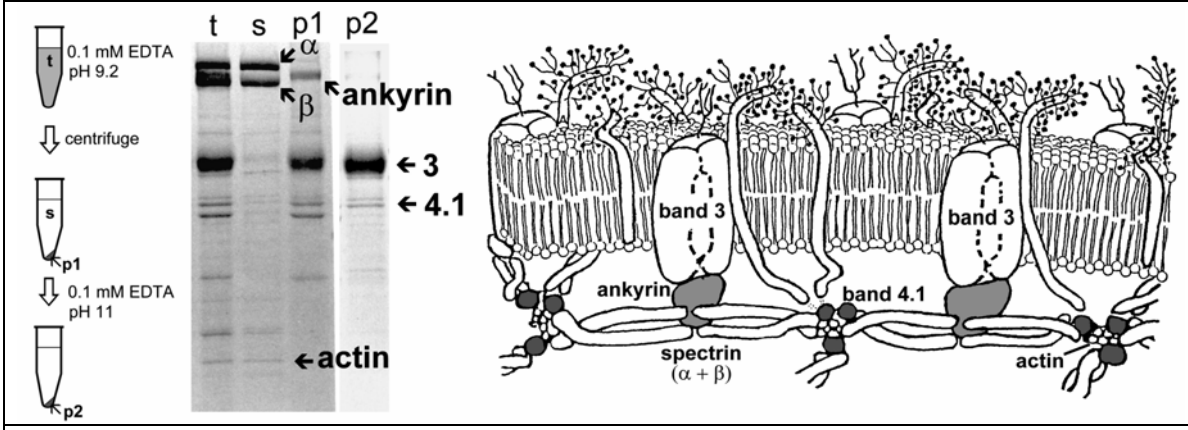
- choice of cells or tissue
- choice of membrane fraction
- homogenization conditions
- preparation of membranes
  - ♦ differential centrifugation
  - ♦ density gradient centrifugation
- solubilization of membranes
- isolation of proteins

The exact isolation technique will depend upon the source and type of membrane being isolated as well as the desired purity.

TYPES OF MEMBRANE PROTEINS

It is possible to remove proteins from the lipid bilayer and subject membrane proteins to further analysis. The method for solubilization will depend on the nature of the proteins association with the lipid bilayer. In this regard, membrane proteins can be viewed as either integral or peripheral. **Integral membrane proteins** include the transmembrane proteins and lipid anchored proteins. Such proteins are solubilized under conditions which disrupt the lipid bilayer. **Peripheral membrane proteins** are non-covalently associated with other membrane proteins and can be differentially extracted under conditions which disrupt protein-protein interactions, but leave the bilayer intact (eg. high salt, low salt, alkaline pH, urea).

Differential Solubility of Erythrocyte Membrane Proteins



Erythrocyte membranes were suspended in 0.1 mM EDTA (pH 9.2) and centrifuged. The pellet was resuspended in 0.1 mM EDTA (pH 11) and centrifuged again. The total extract (t), the first supernatant (s), and the first (p1) and second (p2) pellets were examined by SDS gel electrophoresis (see chapter on Electrophoresis). Integral membrane proteins such as band 3 and glycophorins are not extracted from the lipid bilayer under either condition. Ankyrin and band 4.1, which bind directly to band 3 and glycophorin respectively, are extracted at pH 11 but not pH 9.2. Spectrin and actin, which form a two-dimensional network connected to the membrane via interactions with ankyrin and band 4.1, are extracted at pH 9.2.

Membrane proteins can be selectively extracted from the membrane depending on the nature of their association with the lipid bilayer. For example, some peripheral membrane proteins can be extracted from the membrane by treating with high salt (0.5-1.0 M) solutions. This will remove proteins that interact weakly with the bilayer via electrostatic interactions. The integral membrane proteins and other peripheral membrane proteins remain associated with the lipid bilayer and are removed by centrifugation. Most peripheral proteins will require low ionic strength and alkali pH (< 9) for solubilization. The exact pH will depend on the particular protein (see figure for example). Chaotropic agents, such as urea, can also be used to solubilize peripheral membrane proteins. However, such agents will not exhibit selectivity and many will disrupt the bilayer and extract integral membrane proteins. Integral membrane proteins can be selectively extracted through the use of detergents (see below).

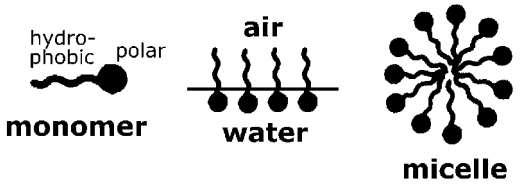
The optimal conditions for solubilization will need to be determined for each individual protein. Isolated membranes are mixed with the extraction buffer. The temperature and time can also be varied. Following the incubation, the particulate (i.e., pellet) and soluble (i.e, supernatant) fractions are separated by centrifugation and the two fractions analyzed for the protein of interest. In general, extraction with detergents will disrupt the bilayer resulting in a particulate fraction that represents cytoskeletal elements and associated proteins. However, some detergents, especially ionic detergents, will also disrupt protein-protein interactions involved in formation of cytoskeletons. Extraction with low ionic strength buffers at alkaline pH will result in a pellet consisting of the bilayer still containing the integral membrane proteins. Sequential extractions can be used to remove proteins that are not of interest before solubilizing the protein(s) of interest.

The extraction conditions may preclude the analysis of enzyme or other functional activities. It may be possible to restore the functional activity by restoring the samples to appropriate conditions or removing the detergents.

DETERGENTS

Detergents are amphiphilic compounds, meaning that they contain both hydrophilic and hydrophobic moieties. A wide variety of chemicals can be classified as detergents. These can be subgrouped based upon the natures of their hydrophobic and hydrophilic regions. Two types of hydrophobic regions found in biologically important detergents are alkyl chains and steroid skeletons (eg., deoxycholate). The alkyl chains can be straight or branched and some also contain phenyl (eg., NP-40) or cyclohexyl groups. The polar groups can be classified as anionic, cationic, amphoteric (i.e., zwitterionic), and non-ionic. Common non-ionic polar groups are glucosides and polyoxyethyl groups. In general, the non-ionic detergents are less denaturing than the ionic detergents.

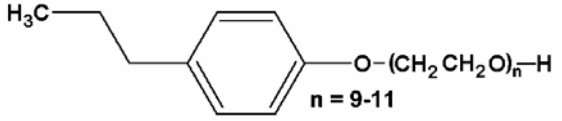
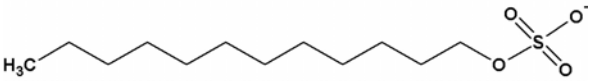
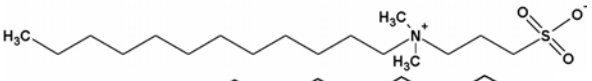
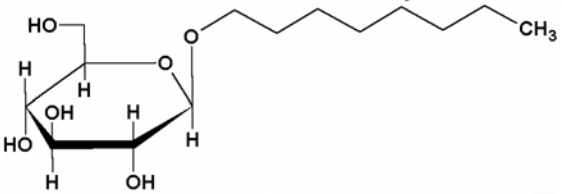
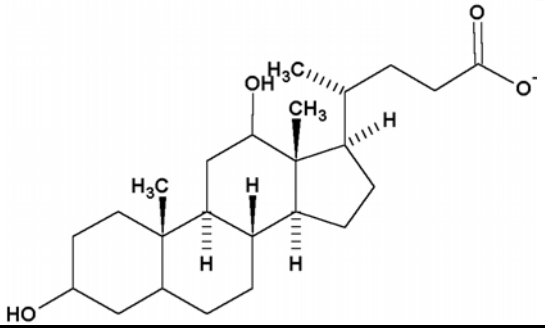
Detergents have surface activity and readily form micelles that are soluble in H_20 (unlike lipids). Micelles are aggregates of detergent molecules in which the polar groups are exposed to the aqueous environment and the hydrophobic groups form an inner core.



Terms defining the properties of detergents are: critical micellar concentration, micellar molecular weight, critical micellar temperature, and cloud point. The **critical micellar concentration** (CMC) is defined as the minimum concentration at which the detergent forms micelles. The **micellar molecular weight** is the average micelle size of a pure detergent and represents the number of detergent monomers within a micelle. The micellar M_r usually expressed as an **aggregation number** (N) which is the average number of monomers in one micelle. In general, detergents with a low CMC have a high micellar M_r. The CMC and micellar M_r are affected by temperature, pH and ionic strength. Increasing the ionic strength tends to lower the CMC and raise the micelle size. The **critical micellar temperature** is the minimum temperature at which a detergent will form micelles. The **cloud point** is the temperature above

which detergent micelles will form super aggregates. Detergents should be used under conditions which are optimal for micelle formation since micelles are critical for detergent function.

Examples of Detergents and Their Properties

Name	Structure	CMC	N (MW)
Nonidet® (N) P-40		0.3 mM	100-155 (647)
sodium dodecyl sulfate (SDS)		8.3 mM	62 (288)
sulfobetaine (SB12)		3.6 mM	55 (336)
n-octylglucoside		14.5 mM	20-25 (292)
deoxycholate		20 mM	3-12 (417)

A few examples of detergents illustrating the different types of hydrophillic and hydrophobic groups. Critical micellar concentrations (CMC) and aggregation numbers (N) are also shown. MW is the average molecular weight of the monomer in daltons.

Detergents interact with both membrane lipids and proteins. At low detergent to lipid ratios, detergent monomers incorporate into the lipid bilayer without disrupting the membrane. As the detergent:lipid ratio increases, lipids, because of their similar properties as detergents, will form mixed micelles with the detergent. Therefore, the detergent concentration needs to be above the CMC. Detergent molecules also interact with the hydrophobic portions of the proteins and in effect replace the lipids. This will lead to the proteins being coated with detergents and prevent protein-protein interactions that would normally result in protein precipitation. Protein solubilization occurs at or near the CMC for most detergents.

There are no general rules for choosing among the various types of detergents. It is not possible to predict which detergent will be most useful for any particular application. Pilot experiments to determine the optimal detergent as well as the optimal conditions for maximal protein solubilization are carried out. Membrane samples are incubated with various

Possible Detergent Effects
- protein structure
- protein activity
- interference with assays
- separation techniques

concentrations of detergents for appropriate times and detergent insoluble material is removed by centrifugation. The soluble and particulate fractions are analyzed for the protein of interest. Ideally, detergents should not affect the native structures or activities of the protein(s) being characterized. In addition the detergent should not affect the measurement of that protein (eg., assay) or chromatography systems that will be utilized in the isolation of that protein.

- Dialysis
- Chromatography
- Replacement

DETERGENT REMOVAL

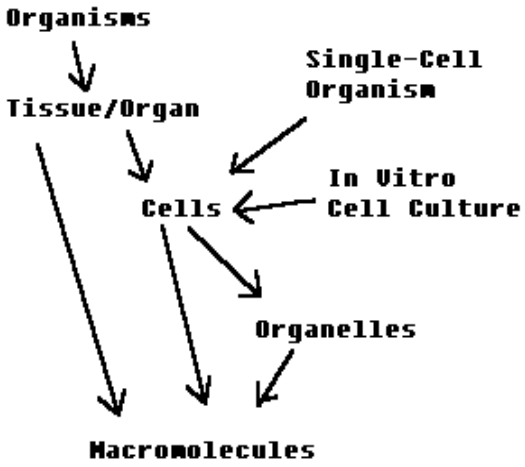
It is often necessary to use a detergent to solubilize the protein of interest. However, later it may be necessary to remove the detergent. In general, detergents are difficult to remove. Therefore, the ease at which a detergent can be removed is also an important criteria to consider when choosing a detergent.

Various methods for detergent removal have been described (Box). The choice of method will depend somewhat on the class of detergent being removed. Dialysis or desalting columns can be used for ionic detergents with a small micelle size and high CMC. However, detergents with a low CMC tend to have a large micelles size and will not pass through the dialysis membranes. Dialysis would require substantial dilution for the removal of the detergent monomers. Gel filtration can be used to remove the large micelles from the smaller proteins. The proteins will still be coated with detergent though.

Hydrophobic chromatography can be also used to remove detergent. Similarly, in the case of ionic detergents, urea can be added and ion exchange chromatography can be carried out. In the presence of urea proteins will not bind to the column. Conversely, ion exchange chromatography or affinity chromatography can be used to bind proteins of interest. The column is washed extensively and the proteins are eluted in detergent free buffers.

In many cases complete removal of the detergent will result in protein precipitation and is undesirable. Detergent replacement can be used in situations where a detergent which is optimal for solubilization interferes with the analysis or further purification of the protein. The sample is subjected to dialysis or chromatography in the presence of the second detergent. Another possibility is to replace the detergent with urea.

APPENDIX 1. CELL DISRUPTION TECHNIQUES



Organisms can be studied on many levels. The whole organism can be studied (eg., behavior, toxicity, etc.). Tissues or organs can be removed by dissection. Tissues can be broken down to single cells by grinding or enzymatic digestions. In addition, single cells can be obtained through *in vitro* cell culture or from single-celled organisms. The choice of organism, tissue, or cell type will depend on the research topic or the protein of interest. A particular organism, tissue or cell type is often the subject of the research. In other cases, though, a particular phenomenon is the object of research. Therefore, the choice of organism, tissue, etc. will depend upon the biological phenomenon being studied. Different tissues and cells will express different amounts of macromolecules or organelles. For example, liver is good source of mitochondria and erythrocytes are a good source of purified plasma membrane.

To study biology at the subcellular and molecular levels it is necessary to disrupt or lyse the cells. Many different methods can be used to disrupt cells (Table). The various techniques all have advantages and disadvantages and the method used will depend on the type of cell and the application. Cell disruption/lysis may destroy the organelle or denature the protein of interest. Therefore, it is important to have an assay to evaluate the disruption techniques in terms of preserving the structures or activities of interest. In general, one needs to empirically determine the optimal conditions for cell disruption.

Subcellular organelles are isolated after cell lysis by differential and/or density gradient centrifugation. Generally homogenation, presses or nitrogen cavitation are the appropriate methods for cell lysis if intact organelles need to be isolated. The other methods are most appropriate in cases were macromolecules are being evaluated without prior isolation of organelles.

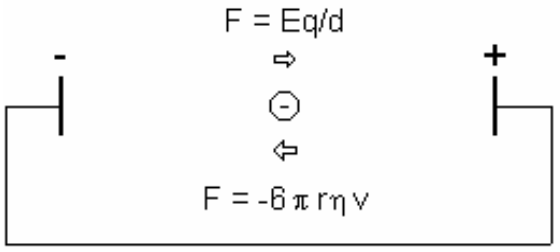
Cellular Disruption Methods

TECHNIQUE	COMMENTS
homogenation	Cells are placed in homogenizer or blender and disrupted by shear forces or grinding. Many different types of homogenizers are available depending on the application. Most consist of a pestle with a defined clearance and can be hand or motor driven. Homogenation can be combined with osmotic methods. Heating can be a problem.
presses	Cells are placed under high pressure in a stainless steel chamber. After reaching a defined pressure the chamber is opened and the cells exit through a small orifice at a controlled rate. The rapid exposure to atmospheric pressure causes the cells to lyse. Optimal pressure for cell disruption needs to be determined empirically for each type of cell and experimental condition.
N_2-cavitation	Cells are placed under high pressure in a N_2 atmosphere. The pressure is suddenly released, causing the N_2 dissolved within cells to boil off and rupture the cells. Both the pressure and rate of pressure release can be controlled.
osmotic, hypotonic	Cells are placed in an isomolar solution of a permeable solute or a hypotonic solution. Water will enter cell which results in swelling until the membrane ruptures. The technique is gentle, but may not lyse cells with walls or rigid cytoskeletons (eg., bacteria, plants, etc.).
sonication	Ultrasonic waves are used to disrupt cells. Heating is a problem and the process tends to vesiculate membranes and subcellular compartments.
freeze-thaw	Ice-crystals formed during freezing will rupture cell membranes. It will generally take multiple cycles of freeze-thaw to disrupt all of the cells.
detergents	Detergents will disrupt the lipid bilayer allowing the contents to be released and will solubilize membrane proteins. Many detergents denature proteins and they are often difficult to remove.
chaotropic agents	Chaotropic agents solubilize proteins by disrupting the structure of water and minimizing hydrophobic interactions. Protein denaturation is a problem.
enzymatic	Bacteria, yeasts and plant cells are often treated with enzymes to remove the cell wall. Enzymatic treatment is often combined with other disruption methods.

CHAPTER 11--ELECTROPHORESIS

Electrophoresis, like centrifugation, is a hydrodynamic technique. A charged particle (i.e., molecule) in an electric field experiences a force that is proportional to the potential difference (E), or voltage, of the electric field and inversely proportional to the distance (d) between the electrodes. (The potential difference divided by the distance (E/d) is referred to as the field strength.) The force will also be proportional to the net charge of the molecule (q). Therefore, the force experienced by the molecule can be expressed by the following equation:

$$F = Eq/d$$



This force will by opposed by a frictional force (= fv) where f is a frictional coefficient and v is the velocity of the particle. The frictional coefficient depends on the size (eg., r = radius) and shape of the molecule and the viscosity (η) of the medium. For example, in the case of a sphere the frictional force is:

$$F_f = 6\pi r\eta v$$

A particle will move at a velocity (v) so that these two forces are equal, therefore:

$6\pi r\eta v = Eq/d$ or solving for v

$$v = Eq/d6\pi r\eta$$

This equation indicates that the mobility (i.e., velocity) of a molecule in an electric field is proportional to the electric field (E/d), or more simply the applied voltage, and the net charge of the molecule. The mobility is inversely proportional to a frictional coefficient (i.e., size and shape of the molecule and the viscosity of the medium), as indicated by the following equation:

mobility = (applied voltage)(net charge)/(friction coefficient)

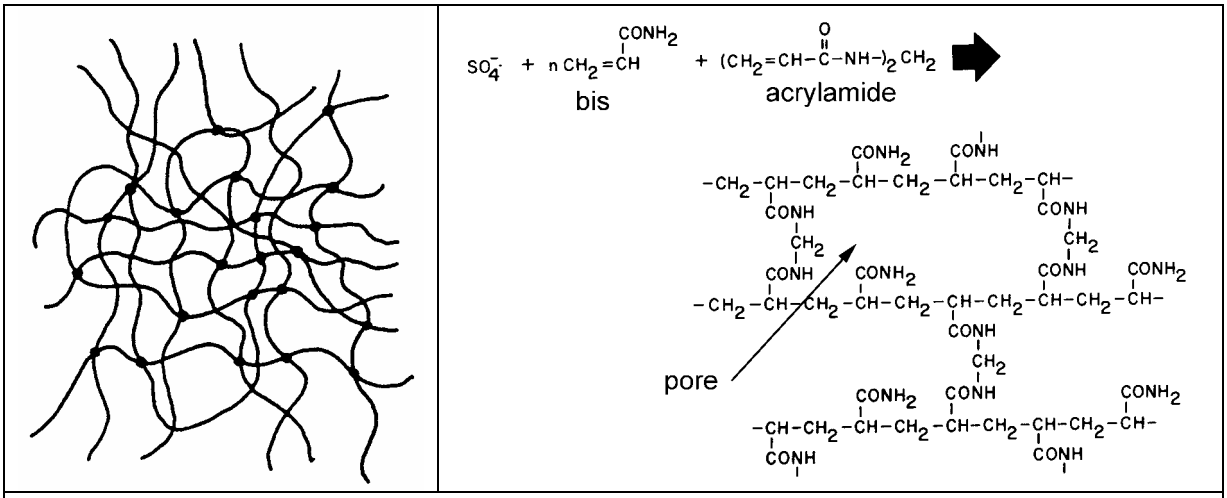
Therefore, it is possible to derive information about the charge, size and shape of a molecule by its mobility in an electric field.

GEL ELECTROPHORESIS

Electrophoresis of macromolecules can be carried out in solution. However, the ability to separate molecules is compromised by their diffusion. Greater resolution is achieved if electrophoresis is carried out on semi-solid supports such as polyacrylamide or agarose gels. Gels are formed by cross-linking polymers in aqueous medium. This will form a 3-dimensional meshwork which the molecules must pass through. Polyacrylamide is a common gel for protein electrophoresis whereas agarose is more commonly used for nucleic acids (see Chapter 18). Agarose gels have a larger pore size than acrylamide gels and are better suited for larger macromolecules. However, either type of gel can be applied to either nucleic acids or proteins

depending on the application.

Gels are formed from long polymers in a cross-linked lattice (Figure). The space between the polymers are the pores. Higher concentrations of the polymer will result in smaller average pore sizes. Polyacrylamide gels are formed by covalently cross-linking acrylamide monomers with bis-acrylamide with a free radical like persulfate ($SO_4\cdot$). The cross-linking of the acrylamide polymers results in 'pores' of a defined size. The total acrylamide concentration and the ratio of bis-acrylamide to acrylamide will determine the average pore size. The polyacrylamide solution is poured into a mold and polymerized. This mold can be a cylindrical tube, but is usually a 'slab' poured between two glass plates (see below).



Diagramatic representation of gels. Gels are form by cross-linking polymers into a 3-dimensional mesh-work or lattice (left). This results in a pourous semisolid material that solutions can pass through. In other words the open spaces are filled with solvent. Increasing the concentration of the polymer will resulting in smaller 'pores'. Agarose gels are formed by heating the desired concentration of agarose in an aqueous buffer. After the agarose dissolves, the solution is cooled and the gel forms. Polyacrylamide gels are formed by chemically crossing acrylamide and bis-acrylamide using a free-radical such as persulfate (right).

Since the gel is solid with respect to the mold, all molecules are forced through the gel. Smaller molecules will be able to pass through this lattice more easily resulting in larger molecules having a lower mobility than smaller molecules. In other words, the gel acts like a molecular sieve and retains the larger molecules while letting the smaller ones pass through. (This is opposite of gel filtration where the larger molecules have a higher mobility because they to not enter the gel.) Therefore, the frictional coefficient is related to how easily a protein passes through the pores of the gel and size will be the major determinant of the mobility of molecules in a gel matrix. Protein shape and other factors will still affect mobility, but to a lesser extent. Substituting size for the frictional coefficient results in:

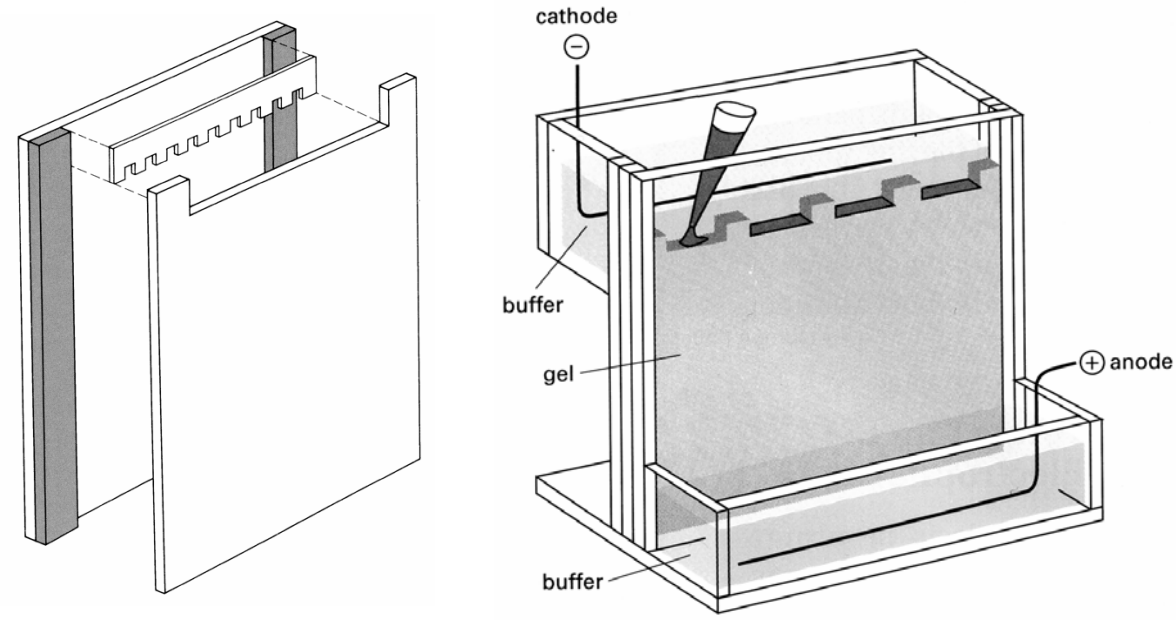
mobility ≅ (voltage)(charge)/(size)

In other words, the mobility of a protein during gel electrophoresis is primarily a function of its charge/mass ratio.

EQUIPMENT

Equipment to conduct gel electrophoresis is relatively simple. They consist of a mold to form the gels, an apparatus to hold the gel and contain buffers, and a power supply capable of delivering the required voltage or current. There are many types of apparati for carrying out electrophoresis depending on the application. Gels can be either in a vertical or horizontal configuration. Polyacrylamide gels are run in a vertical fashion and agarose gels tend to be run in a horizontal position.

Gels can either be formed as cylinders by using glass tubing as a mold (often called tube gels) or formed as rectangular slabs. These slab gels are formed by polymerizing the acryamide solution between glass plates separated by spacers (Figure). Typically the gel is 0.75-1.5 mm thick. At the top a 'comb' is used to form sample wells. Slab gels allow multiple samples to be compared on the same gel, thus eliminating gel-to-gel variations.

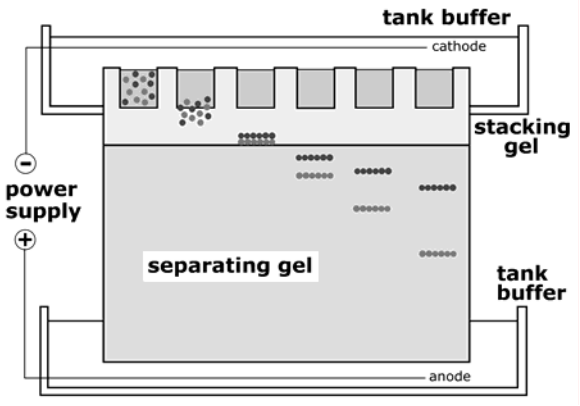


The formed gel is placed into the apparatus to that the top and bottom of the gel are in contact with chambers containing buffer (Figure). These chambers contain electrodes which are connected to a power supply. Thus an electric field is generated across the gel when a voltage is applied. The buffer in the chambers is generally different that the buffer making up the gel for protein electrophoresis and in some applications the buffers in the lower and upper chambers may be different. In most applications the buffers are such that the protein has a negative charge and therefore the anode (positve pole) will be in the lower chamber and the cathode (negative pole) will be in the upper chamber. However, there are applications in which the proteins of interest may be positively charged and therefore the electrodes will be reversed.

Discontinuous or "disc" electrophoresis. The Laemmli discontinuous buffers are extensively used in gel electrophoresis. Discontinuous gels consist of two distinct gel regions referred to as stacking gel and separating gel (see Table) and a Tris-glycine tank buffer. The stacking gel

has a lower acrylamide concentration, a lower pH and a lower ionic strength than the separating gel.

Composition of Laemmli Gels		
	Stacking Gel	Separating Gel
Acrylamide	3-4.5%	6-20%
pH	6.8	8.8
Ionic Strength	0.125 M Tris	0.375 M Tris



The lower ionic strength of the stacking gel results in a greater local electric field strength than in the separating gel. The field strength difference combined with the lower acrylamide concentration results in proteins having a higher mobility in the stacking gel than in the separating gel. In addition, the glycine in the tank buffer has a higher mobility in the separating gel than in the stacking gel because of the pH differences. Therefore, proteins will migrate faster than the glycine in the stacking gel. When proteins reach the separating gel their mobility is decreased because of the increased acrylamide concentration and decreased field strength, whereas the increase in pH results in glycine having a higher mobility. All of these factors result in the proteins becoming compressed at the interface between the two gels and thus increasing resolution (Figure). Resolution in non-discontinuous electrophoresis depends partially on the volume of the sample. However, stacking also occurs at the interface of the sample and gel, especially if a high voltage is applied.

SDS-PAGE

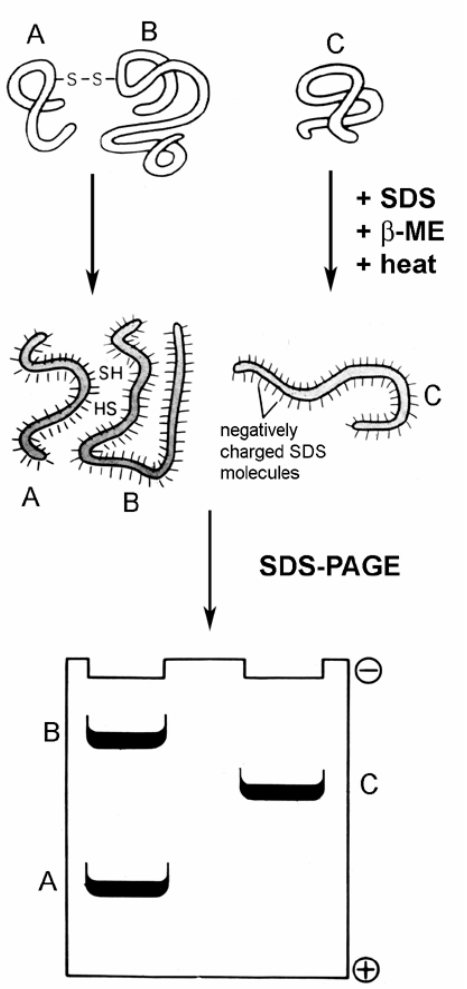
Polyacrylamide gel electrophoresis in the presence of SDS (sodium dodecyl sulfate) is the most common form of protein gel electrophoresis. SDS completely disrupts protein-protein interactions and denatures almost all proteins resulting in a complete unfolding of proteins. In addition, ß-mercaptoethanol (or other reducing agents) is often used to break disulfide bonds. The SDS binds to the unfolded proteins giving all proteins a similar shape (i.e., random coil or extend conformation) and an uniform charge-to-mass ratio. In other words, coating proteins with a negatively charged detergent minimizes the effects of a protein's net charge. Therefore, during electrophoresis in the presence of SDS the mobility of a protein now depends primarily upon its size (i.e., mobility is inversely proportional to protein mass).

Mobility in SDS gel electrophoresis is expressed as a relative mobility (R_f). The distance the protein migrated is compared to the length of the gel, or:

$$R_f = \text{distance protein migrated} \div \text{gel length}$$

The length of the gel is often defined by the migration of a substance which is not impeded by the matrix such a small molecular weight tracking dye (eg., bromophenol blue). This mobility can then be used to calculate the size of proteins. Protein standards of known size are used to

generate a standard curve by plotting the log of the molecular weight against the R_f values. The molecular weight of an unknown protein can be extrapolated from its R_f value (see Appendix 1). Such a calculated molecular weight is designated as M_r to indicate that it is a relative molecular weight based on comparisons to other proteins. For some proteins, though, this estimated molecular weight can differ from the actual molecular weight. In particular, highly charged proteins behave anomalously during SDS gel electrophoresis. In addition, some proteins are not completely denatured by SDS and this retention of some structure will lower the mobility.



Practical considerations. The first step in electrophoresis is to pour the separating gel (Box). Pre-poured gels are also commercially available. Separating gels will typically contain 6-20% acrylamide. The size range of the proteins being separated, the desired resolution and the amount of sample being applied are factors to consider when choosing an acrylamide concentration. Gradient gels can be used in situations where it is necessary to examine both high and low molecular weight proteins on the same gel. The stacking gel is poured after the separating gel polymerizes and just before electrophoresis to minimize diffusion between the two gels.

1. Pour separating gel.
2. Pour stacking gel.
3. Load samples.
4. Apply electric field.
5. Stain or process gel.

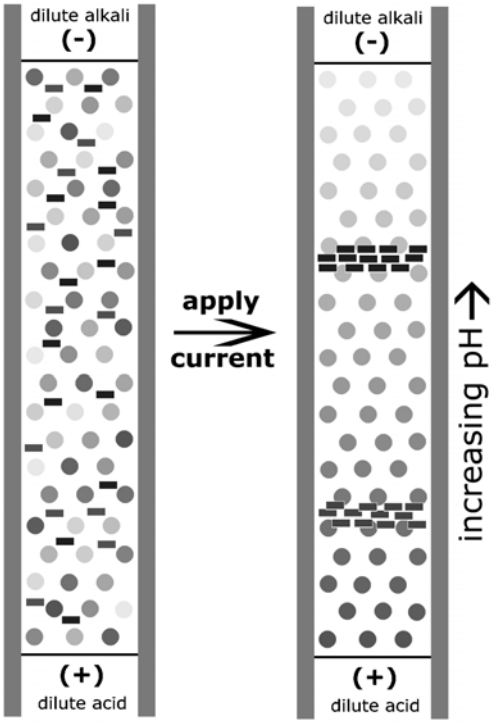
Proteins to be analyzed by SDS-PAGE are solubilized in a sample buffer that typically contains 2% SDS and 5% β-mercaptoethanol and then boiled. The reducing agent is omitted in situations where disulfide bonds need to be preserved. In situations where an enzyme activity will be measure following electrophoresis (see below) a lower SDS concentration is used and the sample is not boiled. The amount of protein that can be loaded onto a gel is limited. Overloading the gels results in the pores becoming plugged and has an adverse effect on the electrophoresis.

After loading the samples into the wells of the gel an electric field is applied across the gel. The mobility of proteins in an electric field is proportional to field strength (E/d), or simply the voltage (E) since the distance (d) is determined by the electrophoresis apparatus. Electrophoresis will proceed faster, and therefore finish sooner, at higher voltages. However, electrophoresis generates heat in proportional to the amount of power, which is the product of voltage and current (P = EI). Excessive heating may result in proteins precipitating within the gel or have other deleterious effects on proteins depending on the nature of the protein sample and total protein concentration. These electrical parameters change during electrophoresis because the ions migrate to the anode and cathode buffers, and therefore lead to an increase in

resistance (R). Resistance affects the voltage and current (E = IR) depending upon which variable is held constant. Most power supplies are capable of delivering constant voltage (E), constant current (I), or constant power (P). Electrophoresis is usually carried out under constant voltage or constant power to minimize the resulting increase in heating that occurs during electrophoresis.

A tracking dye (bromophenol blue) is included in the sample. When this dye reaches the bottom of the gel or some predetermined time afterwards the power is turned off and the proteins detected. A common way to detect proteins after electrophoresis is to stain the gel with Coomassie blue, a dye that binds proteins. Gels are usually 'fixed' before staining with an acetic acid and methanol solution which precipitates proteins into the acrylamide matrix.

ISOELECTRIC FOCUSING



Isoelectric focusing (IEF) separates proteins based on their isoelectric points. The isoelectric point is defined as the pH at which a protein has no net charge (i.e., the number of negative and positive charges are equal) and is a measure of the protein's net charge. Separating proteins according to their net charge is accomplished by generating a pH gradient in an electric field. The effect of protein size on mobility is minimized by carrying out the electrophoresis gels with large pore sizes such as low acrylamide concentrations (eg., 3.5%) or agarose. This large pore size minimizes the molecular sieving.

A pH gradient is generated with **carrier ampholytes**. These ampholytes are a mixture of aliphatic amines and either carboxylic or sulfonic acid. They have a high buffering capacity, low molecular weight (300-600 Da) and a range of pK_a values. Initially the pH of an ampholyte solution will be the average of the pK_a values of the mixture. Application of an electric current will cause the ampholytes to migrate toward the electrodes according to their charges. Ampholytes that have pK_a values above the pH will be positively charged and those with pK_a values below the pH will be negatively charged. As the ampholytes migrate this will result in changes in the local pH due to the buffering action of the ampholytes. This change in the local pH will affect the charge on the ampholytes depending upon the pK_a. The ampholytes will continue to migrate until they reach a position in which the local pH equals their pK_a (i.e., no net charge). The end result is a pH gradient in which the most basic ampholytes are found at the cathode, a dilute alkali solution (eg., NaOH), and the most acidic ampholytes are at the anode, a dilute acid solution (eg., H_3PO_4). Carrier ampholytes with defined pH ranges can be purchased or prepared by isoelectric focusing.

Proteins are also ampholytes and will migrate within the pH gradient until they reach a

pH equal to their isoelectric point. The carrier ampholytes are needed since the protein concentration is generally not high enough to establish a stable pH gradient and the isoelectric points of the proteins may not be uniformly distributed along a pH gradient.

Practical Considerations
- low ionic strength
- protein precipitation
- heating
- gradient breakdown

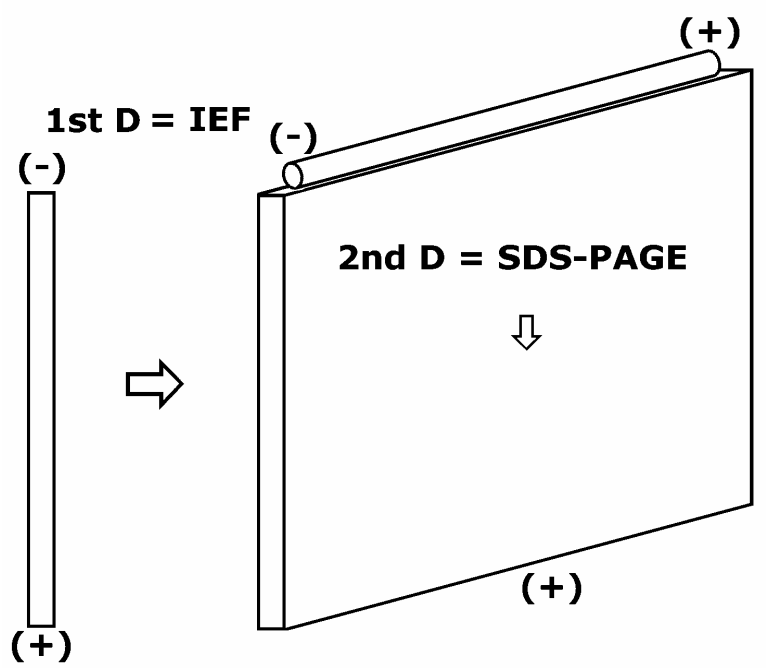
Sample preparation is important for IEF in that many reagents can adversely affect isoelectric focusing. In particular, the ionic strength should be as low as possible. Precipitation of proteins with acetone is a method for removal of excess salts, as well as concentrating the protein. Separations can be performed under either native or denaturing conditions. Urea is the preferred denaturing agent since it is uncharged. Similarly, non-ionic detergents, such as Triton X-100 or NP-40, are less likely to interfere with the formation of pH gradients. Protein precipitation is sometimes a problem in that proteins tend to be less soluble at their isoelectric points and their local concentrations can be quite high. In addition, the high voltages and high resistance (see below) associated with IEF generates substantial heat which increases protein precipitation. Many apparatuses for IEF will have a cooling mechanism to disperse the excess heat. Inclusion of urea and NP-40 in the gels will also minimize protein precipitation.

IEF is an equilibrium phenomenon since the components of the system migrate until they have no net charge. As the system approaches equilibrium the resistance approaches infinity since there are no ions to conduct the current. However, the pH gradient will start to break down before true equilibrium is reached and the ampholytes will migrate into the anode and cathode buffers. This gradient breakdown is accompanied by a lowering of the resistance. Therefore, the progress of IEF can be followed by performing the electrophoresis under constant voltage and monitoring the current. Initially the current will rapidly drop in concordance with the rapid migration of the ampholytes. As the ampholytes lose their net charge, the resistance increases and the current decreases ($E = IR$). The rate at which the current decreases levels off as the system approaches equilibrium. The current will start to rise again when the pH gradient starts to break down. IEF needs to be discontinued before this point.

The pH gradient can be determined with marker proteins with known isoelectric points or by measuring the pH along the gel. This is accomplished by slicing the gel into pieces, eluting the ampholytes into distilled water and measuring the pH.

TWO-DIMENSIONAL GEL ELECTROPHORESIS

Conventional electrophoresis separates proteins according to their charge/mass ratios. SDS-PAGE separates proteins according to subunit, or polypeptide, mass. IEF separates proteins according to isoelectric point (or charge). It is possible to sequentially combine the different types of electrophoresis and run two-dimensional (2-D) gels. A common form of 2-D gel electrophoresis is to first separate proteins by IEF in 'tube' gels. These gels are then equilibrated in SDS gel electrophoresis sample buffer and subjected to SDS-PAGE in a 'slab' gel. (It is necessary to carry out the IEF first since SDS interfere with the isoelectric focusing.) The 2-D separation results in higher resolution since proteins are being separated according to two distinct properties (i.e., charge and size).



PROTEIN DETECTION FOLLOWING ELECTROPHORESIS

- Coomassie Blue Stain
- Silver Stain
- Fluorescence
- Autoradiography/Fluorography
- Enzymatic

Total protein stains. Coomassie-blue staining is a popular method for the detection of proteins following electrophoresis. The procedure is to 'fix' the proteins into the gel, usually in methanol and acetic acid, and then to incubate the gel in a solution containing the dye (see Appendix 2). Excess dye is removed by destaining in the acetic acid and methanol solution. The method is cheap and easy to carry out, but of limited sensitivity.

Silver staining is generally 10-100X more sensitive that Coomassie-blue staining. Three basic types of silver staining methods have been described. All three methods are based upon binding silver ions to proteins and reducing the Ag^{2+} to metallic silver. The silver interacts with primary amines or sulfhydryls. In the diamine method, silver diamine complexes are stabilized with ammonium ions and the free amine groups of proteins. The silver bound to protein is then reduced to metallic silver with formaldehyde. In the non-diamine method, $AgNO_3$ reacts with protein under acidic conditions, followed by reduction to metallic silver with formaldehyde under alkaline conditions. Photodevelopment relies on light (photons) to reduce silver ions to metallic silver. Some silver stain protocols result in certain types of proteins being stained particular colors, thus aiding in the identification of proteins.

- diamine (or ammonical)
- non-diamine
- photodevelopment

Fluorescence. Fluorescent dyes that bind non-covalently to proteins are also used for protein detection following electrophoresis. Following incubation with the dye and destaining the gels are examined under ultraviolet illumination. Polyacrylamide exhibits some background fluorescence. Proteins can also be convalently labeled with fluorescent probes prior to

electrophoresis. However, attaching fluorescent groups to proteins will change their charges.

Autoradiography and Fluorography. Radioactive proteins can be detected by exposing X-ray film to the gel. Radioactivity emitted from the proteins will activate silver grains in the film emulsion which will be converted into metallic silver during development. The reduction to metallic Ag will result in a dark spot or 'band' appearing on the film in positions which correspond to proteins. **Fluorography** refers to the exposure of film to secondary light emitted by either intensifying screens or fluors upon their exposure to radiation.

Proteins can be made radioactive by incubating cells in the presence of radioactive amino acids or other compounds that are naturally incorporated into proteins (eg., phosphate, fatty acids, glycosyl groups, etc.). Non-physiological means can also be used to incorporate radioisotopes into proteins. Iodination involves the reaction of radioactive iodine (^{125}I or ^{131}I) with proteins by one of several methods. The iodine in incorporated primarily into tyrosine residues and to a lesser extent into cysteine and histidine residues. Several reagents are available that will react with either free sulfhydryls (eg., cysteine) or primary amines (eg., lysine). Such alkylating agents can be used for the incorporation of either ^{3}H or ^{14}C. In most applications the proteins are radiolabeled before electrophoresis. It should be noted though that alkylating primary amines will affect the charge of the protein.

Autoradiography and Fluorography

Isotope Energy	Detection Method	Sensitivity (dpm/mm^2)
High (^{32}P, ^{125}I)	Screens	2-5
	Direct	0.5
Medium (^{14}C, ^{35}S)	Fluor	15-25
	Direct	2
Low (^{3}H)	Fluor only	10-20

Gels are dried before carrying out autoradiography or fluorography. A gel dryer removes all of the water and solvent from the acrylamide matrix by heating the gel while under vacuum. This results in the matrix collapsing into a thin layer. The dried gel is placed next to a sheet of X-ray film in a light tight cassette and exposed for an appropriate amount of time depending on the isotope and the amount of radioactivity. The exposed film is developed.

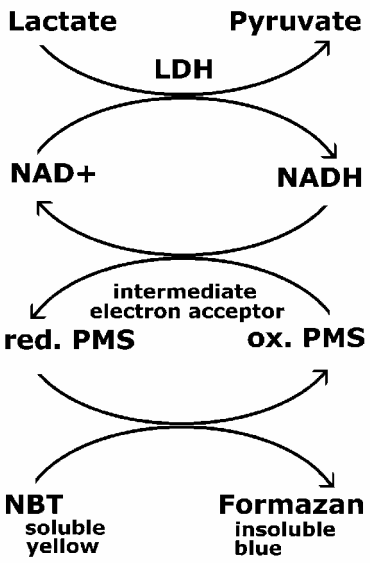
Some emissions of high energy isotopes (eg., ^{32}P and ^{125}I) will pass through film without exposing it. An **intensifying screen** is placed on the opposite side of the X-ray film from the dried gel. Emissions passing through the film will then interact with the intensifying screen which will fluoresce. The light from this fluorescence will then expose the X-ray film. Using an intensifying screen will usually decrease the exposure time by a factor of 7-10X. The efficiency of the intensifying screen is highest at -70^{o}.

Emissions from low energy isotopes (eg., ^{3}H) and some medium energy emissions (eg., ^{14}C and ^{35}S) are not energetic enough to escape the dried gel matrix. In this case a scintillation fluor is impregnated into the gel before drying. The radioactive emissions then interact with the fluors resulting in fluorescence. The fluorescence will then expose the X-ray film. Inclusion of fluors in the gel will not enhance the detection of high energy isotopes, and likewise, intensifying screens will not enhance the detection of low and medium level isotopes.

Phosphor-imager machines use screens containing storage-phosphors as a replacement for the use of film. These phosphor-imager screens trap the energy of radioactive emissions and are sensitive to both beta particles and gamma rays, but insensitive to visible light. The storage-

phosphor captures energy with an efficiency around 100% for any particle that strikes the screen. Subsequent scanning the screen with a laser beam releases the stored energy as blue light. This blue light is then converted into an image file for display and quantitation by the phosophor-imager's computer. Such high efficiency makes it possible to detect low levels of radioactivity after short exposure times. The phosphor screens can be used repeatedly by 'erasing' the screen following scanning.

Enzyme. It is possible to detect enzyme activity following gel electrophoresis of some proteins. Generally it is necessary to carry out non-denaturing gel electrophoresis or IEF since proteins will need to be in a native configuration for the expression of activity. However, some enzyme activities can still be detected following SDS-PAGE. This will require that activity is located in a single polypeptide chain (i.e., no multi-subunit enzymes) and that the protein is able to refold into an active configuration following electrophoresis. To promote this refolding the sample is initially solubilized in a lower concentration (eg., 0.1%) of SDS than normal and is not heated. Then following electrophoresis the SDS is removed from the gels by washing the gels in an appropriate buffer. The removal of the SDS and the refolding of the proteins can be facilitated in some situations by including a non-ionic detergent (eg., Triton X-100) in the wash buffer. This will lead to a replacement of the SDS by a more approiate detergent.



A common method to detect enzyme activity following electrophoresis is to use substrates which form insoluble products. These insoluble products will form a precipitate in the gel at the position of the enzyme activity. For example, tetrazolium salts are frequently used as color indicators for the detection of enzyme systems in which reduction equivalents are formed. The procedure is to incubate the gel with the appropriate substrates and the indicator molecules. For example, lactate dehydrogenase (see figure) can be detected by incubating the gel with lactate, NAD^+, phenazine methosufate (PMS) and nitro blue tetrazolium (NBT). The NADH formed by the dehydrogenease is not very efficient at transferring hydrogens and electrons to NBT and therefore an intermediate electron acceptor is needed (i.e., PMS). NBT is a soluble yellow substance which when reduced to a formazan results in the formation of an insoluble blue compound. Other dehydrogenase activities or other redox reactions are detected by using the appropriate primary substrate and changing the intermediate electron acceptor or tetrazolium salt as necessary.

It is also possible to detect enzymes which are generally not considered to be involved in redox reactions using a substrate that can be coupled to the reduction of tetrazolium salts to formazan dyes. For example, phosphatase activity can be detected by incubating with 5-bromo-4-chloro-3-indolyl phosphate (BCIP) and NBT. After dephosphorylation of BCIP by the phosphatase, the resulting indoxyl reduces the NBT to the foramazan dye. This subtrate pair is widely used for the detection of enzyme-linked antibodies in immunoblotting procedures (see chapter on immunoassays).

Proteases can also be detected after electrophoresis by co-polymerizing the gels with gelatin or other proteins. Following electrophoresis and removal of the SDS, the gel is incubated under conditions which promote proteolysis and then stained with Coomassie blue. Clear regions will denote the positions of proteases.

Quantification. The results obtained from gel electrophoresis tend to be qualitative and difficult to precisely quantify. One possible way to quantify the amount of a specific protein is to electrophorese known amounts of a protein and to subjectively compare the staining intensities with that of the unknown protein. The stained protein bands could also by excised, the dye extracted with a solvent and the amount dye determined with a spectrophotometer. The absorbance values could then be used to generate a standard curve. However, individual proteins bind different amounts of Coomassie blue and there will be gel-to-gel variations. Similarly radioactive proteins can be excised from the gel, the gel slice solubilized and the amount of radioactivity determined by liquid scintillation counting. This method can be rather laborious if all the proteins within a sample are being evaluated.

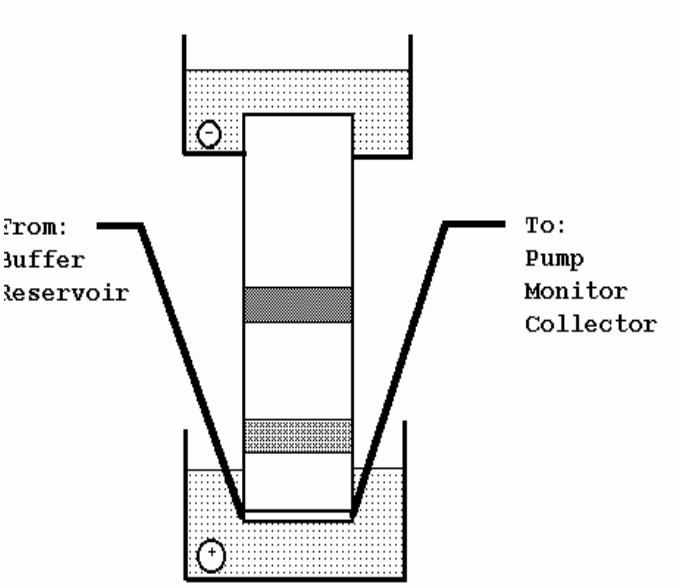
Scanning densitometry is usually a convenient method to quantify the amount of a particular protein. Gels, autoradiographs, or even photographs are scanned and the peak height and areas of bands are determined. The values are then used to calculate the amount of protein based on standard curves. The uniformity of the shape and width of the bands also needs to be considered.

PREPARATIVE ELECTROPHORESIS

Gel electrophoresis is a high resolution technique and can potentially be used for protein purification. One major problem, however, is the limited amount of protein that can loaded onto a gel. Another major problem is the recovery of proteins from gels, especially if the gels have been fixed and stained. One possible method is by simple diffusion. The excised gel slice is resuspended in a buffer and the protein allowed to diffuse out. This process is slow and the protein is excessively diluted. Alternatively, the protein can be recovered from the gel slice by electroelution using commercially available apparatuses. Both diffusion and electroelution are characterized by low yields. An alternative strategy is to transfer the protein to a membrane support and analyze the protein bound to the membrane as is done in immunoblotting or micro-sequencing.

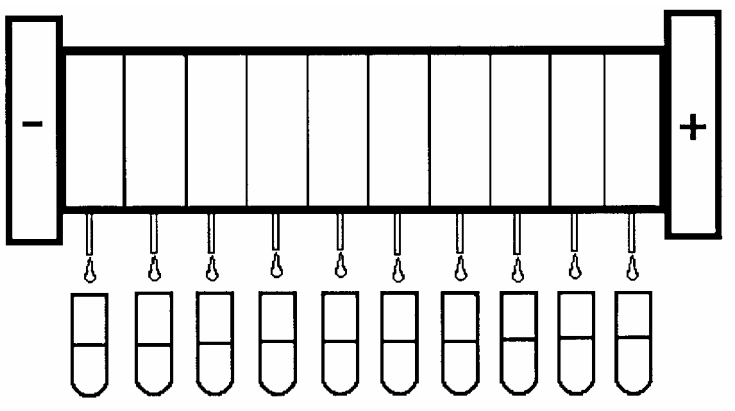
If the goal is to raise antibodies against the purified protein then it may not be necessary to remove the protein from the gel. The excise gel slice containing the protein of interest is frozen in liquid nitrogen and ground into a fine powder with a mortar and pestle. The ground gel is then resuspended in saline and used to directly immunize animals. The protein is slowly released from gel slices and actually enhances immunization through the 'depot' effect.

Commercial apparatuses for preparative electrophoresis are available. Most treat the gel as a column and continue electrophoresis until all of the proteins migrate off from the bottom of the gel. At the bottom of the gel is a chamber sealed off with a dialysis membrane to prevent the proteins from migrating into the tank buffer. The protein coming off from the bottom of the gel are continuously eluted and pumped to monitors and fraction collectors.



Preparative Electrophoresis

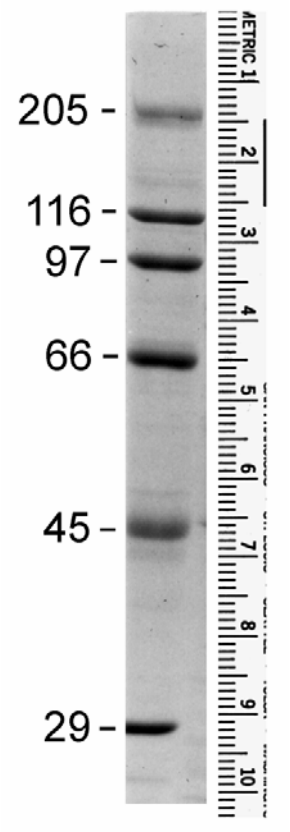
A commercially available preparative IEF apparatus, the Rotofor® (Bio-Rad), avoids the problems with gels by conducting the focusing in free solution. The apparatus consists of column (i.e., electrophoresis chamber) which is held in an horizontal position. The electrode solutions are separated from the electrophoresis chamber by ion exchange membranes and gaskets. A ceramic cooling core runs down the center of the electrophoresis chamber and the chamber is rotated around the horizontal axis during electrophoresis to dissipate heat, minimize convection, and maintain efficient electrical contact. The chamber is filled with the protein solution containing ampholytes and usually sucrose to the increase density. The mixture is subjected to electrofocusing by applying the appropriate voltage. The electrophoresis chamber is partitioned into 20 chamber each divided by a porous polyester screen with a pore size of approximately 6 μm. These screens offer some resistance to fluid convection by do not impede the flow or current or the movement of proteins. After the power is turned off the 20 fractions are collected with a special device that rapidly harvests the 20 fractions before substantial mixing occurs between them.



ELECTROPHORESIS APPENDIX 1. SIZE CALCULATION.

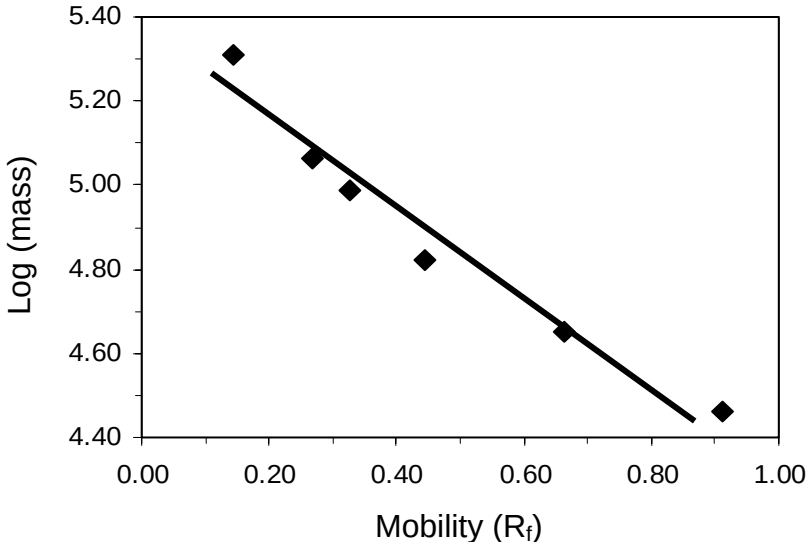
Caluculation of molecular mass from SDS gels.

Molecular masses of proteins can be estimated by comparing the migration of proteins of interest to standards of known size. The relative mobilities of the standards are plotted against the log of their molecular masses. The sizes of unknown proteins are then extrapolated from the standard curve.



The relative mobility (or R_f) is calculated by dividing the distance the protein migrate by the total length of the gel or the dye front as illustrated by the figures and table.

mass (daltons)	log (mass)	distance migraged (mm)	Rf (÷ 101 mm)
205,000	5.31	14.5	0.14
116,000	5.06	27	0.27
97,000	4.99	33	0.33
66,000	4.82	45	0.45
45,000	4.65	67	0.66
29,000	4.46	92	0.91



Coomassie Blue Staining Following Electrophoresis

G-250 Method (quick)

1. Fix the proteins by incubating the gel in the destain solution (10% methanol and 7% acetic acid) for at least 15 minutes. Gels can be left in fixative indefinitely.

2. Briefly rinse the gels with water.

3. Stain the gels for 20 minutes with 0.1% Coomassie-blue G-250 in 3.5% perchloric acid. Gels thicker than 0.75 mm may need to be stained longer.

4. Destain the gel with several changes of the destain solution (10% methanol and 7% acetic acid) over the next 30-60 minutes. (Works best if the first 1-2 washes carried out in the first 5-10 minutes.)

R-250 Method (slow)

1. Fix the proteins by incubating the gel in the destain solution (10% methanol and 7% acetic acid) for at least 15 minutes. Gels can be left in fixative indefinitely.

2. Stain the gels for 3 hours with 0.1% Coomassie-blue R-250 in 50% methanol + 10% acetic acid.

3. Destain the gel with several changes of the destain solution (10% methanol and 7% acetic acid) over the next 6-24 hours.

Notes

- For optimal results, the G-250 should be added to the perchloric acid. Water should then be slowly added while stirring. Continue stirring for at least one hour and then filter the solution over filter paper.

- The R-250 solution should also be filtered after its preparation.

- Other concentrations of methanol and acetic acid also work in the fixative and/or destaining solutions

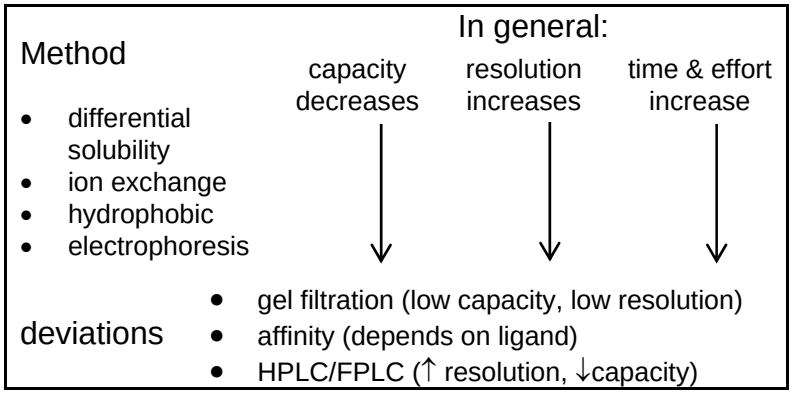
CHAPTER 12--PROTEIN PURIFICATION OVERVIEW

Why purify proteins?
- detailed studies on function
- determination of structure
- industrial/pharmaceutical applications
- generate antibodies
- amino acid sequence determination

There are various methods that separate proteins according to distinct chemical and physical properties. Proteins can be purified by sequentially applying these techniques. Before purifying a protein one needs to consider how much protein is needed, how pure it needs to be, whether the purified protein needs to be in its native configuration, or express an activity. These considerations will be determined by the ultimate use for the purified protein (Box). If the purified protein is going to be used in functional and structural studies, isolating the protein in its native and active configuration will be important. However, if the goal is to determine a partial amino acid sequence then concerns about denaturation are minimal.

The optimal method to purify a protein is empirically determined by carrying out small-scale pilot studies. After the optimal conditions for a particular purification step is worked out, then the procedures can be scaled up. The order in which the steps are carried out will also affect the final outcome. In general, though, it is best to start with high capacity methods that are easily and quickly carried out and then proceed to high resolution/low capacity methods.

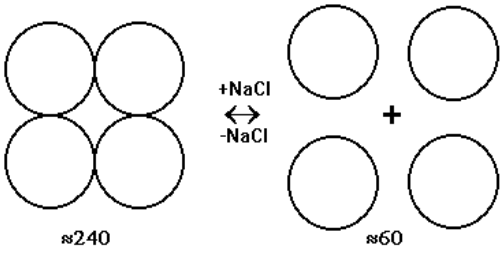
Capacity refers to how much total protein can be readily accommodated and resolution refers to the ability to separate the protein of interest from contaminating proteins. Generally, as the resolution of the technique increases, the time required to carry out the technique increases and the protein capacity decreases (Box). One exception is that gel filtration is not usually a high resolution technique, as well as being low capacity and relatively difficult to carry out. The capacity and resolution of affinity chromatography depends on the ligand. Some ligands afford both high capacity and high resolution. The resolution of any of the chromatographic technique can be substantially enhanced through the use of HPLC. However, the capacity of HPLC columns is much lower so that it may not be useful until later in the purification protocol.



The time and number of manipulations should be minimized in designing purification protocols. For example, the order of the separation techniques can be chosen as to minimize the number of buffer changes when other factors are equal. Another way to expedite purification is to exploit a unique feature of the protein. In this regard, affinity chromatography exploits the specificity of ligand binding to produce high resolution separations in a single step.

An example of exploiting a unique feature is the purification of calmodulin on hydrophobic columns in the presence of Ca^{2+}. Binding of Ca^{2+} causes a conformation change in calmodulin so that a hydrophobic domain is exposed on the protein's surface. The function of this hydrophobic domain is to bind to other proteins that are regulated by calmodulin. In the absence of Ca^{2+}, calmodulin is hydrophillic and does not bind to other proteins. Hydrophobic chromatography in the presence of Ca^{2+} results in absorption of calmodulin to the column. Specific elution of calmodulin is accomplished by the addition of EGTA which chelates the Ca^{2+} resulting in a conformation change in calmodulin and the loss of its surface hydrophobicity. The surface hydrophobicities of the other proteins are not affected by the presence or absence of Ca^{2+} and therefore do not elute with EGTA. the activity of the bound protein. binds to hydrophobic columns, but does not bind in the absence of Ca^{2+}. This unique feature permits a rapid and nearly complete purification of calmodulin in a single step.

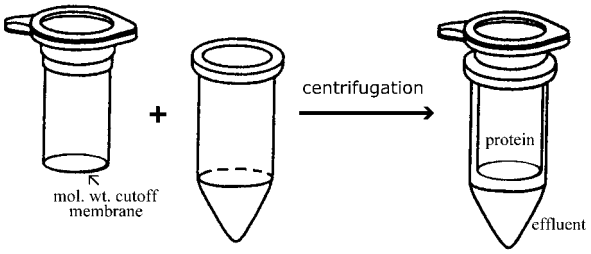
Another example of exploiting unique features is the use of gel filtration in the purification of proteins which can be reversibly dissociated subunits. For example, gel filtration can be carried out in the presence of low salt (typically 0.15 M NaCl) and the protein will elude from the column corresponding to the molecular weight of the complex. If the complex can be reversibly dissociated by high salt (eg., 1 M NaCl) and re-subjected to gel filtration chromatography, it will elude from the column corresponding to its subunit molecular weight. The intact protein can then be reformed by removing the salt. These sequential gel filtration columns in low and high salt will result in substantial purification with a normally low resolution technique.



- lyophilization (freeze drying)
- precipitation (eg., $(NH_4)_2SO_4$)
- dialysis against solid PEG
- dialysis against 50% glycerol
- ultrafiltration

Concentration. Many chromatographic procedures result in a dilution of the protein sample. To avoid the possible negative effects of dilute protein solutions, it is often necessary to re-concentrate the sample between chromatographic steps or after the purification is completed. Precipitation is a convenient method in cases where $(NH_4)_2SO_4$ does not adversely affect the protein or next chromatographic step. Water can be completely removed by lyophilization, or freeze-drying. However, many proteins are not stable following lyophilization and salt concentrations will be dramatically increased. Water can also be removed by placing the protein solution in a dialysis bag and surrounding it with solid high molecular weight polyethylene glycol. Water and buffer components will diffuse out of the dialysis and the ionic strength remains approximately constant. Similarly, dialysis against a buffer solution containing 50% glycerol will concentrate the solution 4-5 fold as the water

diffuses out of the dialysis bag. This may be especially useful as a terminal purification step in that glycerol is often added to purified proteins to increase the stability.



Ultrafiltration involves forcing a protein solution through a membrane with a defined molecular weight cutoff. The protein does not pass through the membrane, whereas the other buffer components do resulting in no change in the ionic strength. This method is usually relatively rapid and does not adversely affect the protein sample. Ultrafiltration is either carried out in a special apparatus or by centrifugation using a disposable filter cartridge that fits into the centrifuge tube. The protein sample is placed into the cartridge and centrifuged until the protein solution, which remains above the filter, is at the desired volume. The filtrate can also be recovered and evaluated. A wide range of sizes are available for both the ultrafiltration apparatus and the centrifuge cartridges.

Evaluation. It is important to monitor and evaluate the protein of interest during the purification procedure. Two important criteria to monitor are **fold-purification** and **percent yield** (see Appendix). The percent yield refers to the amount of the protein recovered following a purification step and is determined by dividing the amount of total activity recovered by the total activity in the starting material. The total activity of the starting material is determined by multiply the activity/ml by the total volume of the sample applied to the column or other separation method. The fractions containing the protein of interest are then pooled and the activity/ml is determined and multiplied by the total volume of the pooled fractions. Generally steps that result in low yields should be avoided unless the protein is available in large amounts.

The fold-purification is determined by dividing the specific activity following the purification procedure by the specific activity before the procedure. The specific activity is the activity per mg protein. The specific activity is determined by dividing the activity/ml by the protein concentretration (mg/ml) in both the starting material and the pool fractions following the separation technique. After each step the specific activity should go up since unwanted proteins are being removed. This increase in specific activity provides a measure of the degree of purification obtained by that method(s). Another way to view fold-purification is the removal of unwanted proteins. For example, a 4-fold purification means that 75% of the contaminating proteins were removed by that step. Both the percent yield and fold-purification can be determined after each step or with the original starting material to evaluate the entire purification process.

In addition, the qualitative aspects of the purification should be analyzed by gel electrophoresis. A protein purified to homogeneity should exhibit a single 'band' when analyzed by non-denaturing electrophoresis. The number of polypeptides (i.e., bands) observed after SDS-gel electrophoresis will depend upon the subunit composition of that protein.

MICROSEQUENCING AND PEPTIDE MAPPING

A partial amino acid sequence can provide information about a proteins identity or be used in the design of recombinant DNA probes. Through chemistry know as the Edman degradation cycle the N-terminal amino acids can be sequentially removed and then determined. Generally instruments which automate this sequential removal of amino acids and their analysis are utilized. Generally proteins or peptides to be sequence are submitted to a service provider and sequenced for a fee depending on the number of residues sequenced.

Edman degradation chemistry. Protein samples are usually absorbed or coupled to a membrane or glass fiber filter support which is placed in the reaction chamber of the instrument. Each cycle of the Edman degradation cycle consists of three steps: coupling, cleavage and conversion. The chemicals necessary for each of the steps are sequentially pumped into the reaction chamber. In the coupling step, the unmodified N-terminus of the peptide is reacted with phenyl isothiocyanate to form a phenylthiocarbamoyl-peptide. This is followed by addition of either liquid or gaseous trifluoroacetic acid which will cleave PTC derivatized N-terminal amino acid from the peptide resulting in the formation of an anilinothiazoline-amino acid derivitive. This unstable ATZ-amino acid is pumped from the reaction chamber into a conversion chamber where it is converted to a more stable phenylthiohydantoin derivative. The immobilized peptide remains in the reaction chamber and the N-terminus of the next residue is available for another round of Edman degradation. The PTH-amino acid is transferred from the conversion chamber to an HPLC and the specific amino acid is determined from its elution time. Depending on the quantity and quality of the sample up to 30 cycles can be carried. Generally about one pmole of amino acid can be detected although more sensitive instruments are available.

Microsequencing following gel electrophoresis. The N-termini of proteins can be sequenced following separation by gel electrophoresis. This obviates the need to purify the protein to near homogeneity by more laborious conventional methods and allows one to take advantage of the high resolution of SDS gel electrophoresis. The protein of interest should not significantly co-migrate with contaminating proteins and be at least 80% pure. Generally 10-100 pmoles of the protein, which corresponds to 0.5-5 μg of a 50 kDa protein, will be needed to make carry out the analysis. If no sequence data is obtained from 100 pmoles of protein then it is generally assumed that the N-terminus is blocked (see below). Whereas if less than 10 pmoles of sample are used it may be difficult to determine if the lack of sequence data are the result of a blocked N-terminus or due to insufficient material.

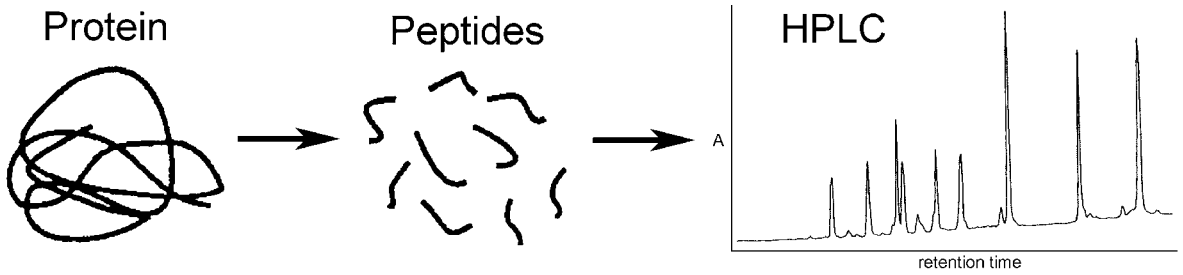
Following gel electrophoresis (either 1- or 2-dimensional) the proteins are electrophoretically transferred from the gels to a membrane support. The membrane of choice is a modified PVDF membrane which has good protein retention and is able to withstand the solvents, acids and bases used in the Edman degradation chemistry. The protein of interest is detected by a brief staining with Coomassie blue (or other dyes) and the band excised. Components of the electrophoresis and transfer buffers, such as salts, Tris, glycine, SDS and acrylamide, interfere with the sequencing reactions and must be completely removed by extensive washing in water. The membrane piece containing the protein to be sequence is submitted to the sequencing service.

Sequencing internal peptides. Edman chemistry is dependent on a free amino group at the N-terminus. Apporximately half of eukaryotic cellular protein are modified (eg., acylation) at the N-terminus and do not have a free amine in this position (i.e., 'blocked' N-terminus). In addition, sometimes during the purification and preparation of proteins for N-terminal sequencing the proteins can become blocked. There are some methods (chemically and enzymatically) to remove the N-terminal blocking group of a protein or peptide. These include deformylation, deacetylation and removal of the pyroglutamate group. The choice of a particular method depends on the protein or peptide and the nature of the blocking group. However, in most cases because of the uncertain nature of the blocking group an internal peptide sequence is determined.

Proteins can be treated with chemicals or site specific proteases to generate peptide fragments. For example, CNBr cleaves proteins on the C-terminal side of methionine residues. Many protease recognized specific amino acid residues and reproducibly cleave proteins at specific sites. The resulting polypeptides are then separated (typically on HPLC reverse phase columns). The peaks corresponding to peptides are collected and absorbed onto glass fiber filters as a solid support. Reverse phase chromatography is especially advantageous in this application since it is carried out in volatile solvents. The glass fiber filters containing the peptides are submitted to the service provider and the N-termini sequenced.

Site Specific Cleavage of Proteins

chemical	CNBr	Met-X
	skatole	Trp-X
proteolytic	trypsin	Arg-X
		Lys-X
	V8	Glu-X
		Asp-X



Protocols and discussions about preparing proteins for microsequencing can be found at http://www.biol.berkeley.edu/crl/microchemical/protein_sequencing.html.

Peptide mapping. It is also possible to use site specific proteases to compare proteins and determine if they are related. The general procedure is to digest a protein with a protease and characterize the pattern of peptides formed. Each protein will exhibit a unique pattern of peptides. Such procedures are known as 'protein fingerprinting' or 'peptide mapping'. The peptides can be analyzed by HPLC, thin-layer chromatography (TLC), or gel electrophoresis. In two-dimensional peptide mapping the peptides are separated by TLC followed by high voltage electrophoresis followed by separation by high-voltage electrophoresis in a second dimension.

Another form of peptide mapping involves excising a protein band from the gel and

loading it on a second gel of higher acrylamide concentration. A protease is added to the well and the protein of interest is partially digested during the re-electrophoresis. Proteins will exhibit a distinct pattern of smaller polypeptides. This method, often referred to as the Cleveland method, is rapid and easy to carry out and allows proteins to be directly compared in neighboring lanes. The pattern of polypeptides can be used to assess whether two proteins are similar or different.

Protein modification. Sometimes it is necessary to chemically modify proteins. and in particular to radiolabel proteins. Several methods are available to directly couple ^{125}I to proteins. Tyrosine residues and to a lesser extent histidine and cysteine, are labeled. Free amines, as found on lysine and arginine residues are quite reactive and numerous aldehydes, esters and imidoesters are available for modifying proteins. The free sulfhydryl group of cysteine is also reactive and easy to alkylate. Alkylation prevents the formation of disulfide bonds. Bifunctional reagents can also be used to cross-link proteins.

APPENDIX. FOLD PURIFICATION AND RECOVERY

Example of a form to record data on the purification of an enzyme.

Purification Procedure	Vol.	Protein				Activity			%yield	fold-purify
		µl	µg	mg/ml	total	units/ml	total	sp. act.		

The effectiveness of a purification procedure is evaluated by monitoring the recovery of the protein of interest and its relative purification after each step. The volume, protein concentration and amount of activity should be determined for the crude homogenate and all subsequent manipulations. The % yield for any single procedure (i.e., recovery) is determined by dividing the total activity recovered in the pooled fractions by the total activity in the preceding step. The overall % yield is determined by dividing the total activity by the total activity in the crude homogenate (i.e., the first step). The total activity is determined by multiplying the total volume by the units/ml. The specific activity (units/mg protein) represents the proportional amount of the protein of interest and should go up with each purification step since the total protein concentration is decreasing. The fold-purification is determined by dividing the specific activity after a particular step by dividing the specific activity by the specific activity of the preceding step or the specific activity of the crude homogenate. The maximum fold-purification that can be obtained is determined by the percentage of the protein of interest in the crude homogenate. For example, a protein that represents 0.1% of the total cellular protein should be homogeneous after a 1000-fold purification.

PART III

Immunological Methods

Topics covered:

- Antibody Production
- Monoclonal Antibodies
- Immunoassays

CHAPTER 13--IMMUNIZATION

Antibodies are proteins that specifically recognize and bind to other molecules with a high affinity and specificity. This high degree of specificity makes antibodies powerful reagents for the characterization of proteins and other macromolecules. Antibodies are generated against specific proteins or other macromolecules by immunizing animals and collecting the sera. The immune sera contains a mixture of antibodies that recognize various epitopes on the antigen. The immune system can also respond to foreign substances via a cellular immune response.

IMMUNOGENICITY

Immunogenicity (i.e., the level of the immune response) depends on the chemical nature of the immunogen and the ability of the animal to respond to the immunogen. An immunogen is a substance that induces a specific immune response, whereas an antigen is defined as a substance that reacts with the products of a specific immune response. An antigen can be an immunogen if it is correctly presented to the immune system. In addition, there is a genetic component to immunogenicity in that some substances are immunogenic in one species but not in another. Similarly, some substances are immunogenic in one individual but not in others (i.e. responders and non-responders). Immunogenicity can be manipulated, though, by either modifying the immunogen or stimulating the animal.

An immunogen generally needs to be a foreign substance since the immune system normally discriminates between self and non-self. In addition, immunogens need to have significant chemical complexity and be of a large size (eg., > 10 kDa). Proteins and polysaccharides tend to make good immunogens, whereas nucleic acids and lipids are usually poorly immunogenic. Lipids can function as haptens, though (see below).

- foreigness
- molecular complexity
- molecular size
- B-cell epitope
- T-cell epitope
- class II binding site
- degradation/presentation
- particulate/phagocytosis

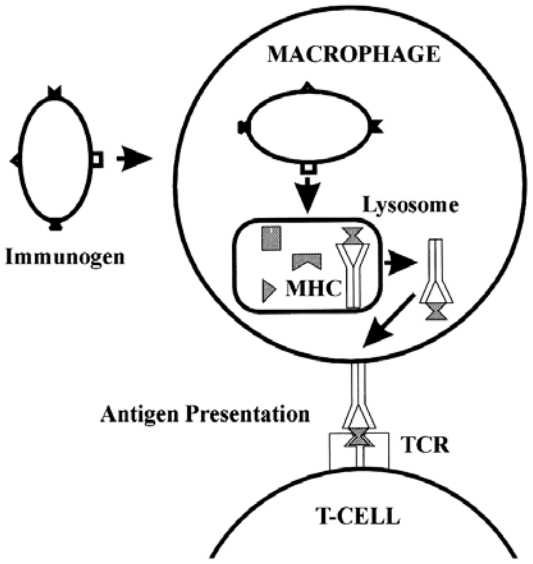
Optimal immunogens should also have B-cell epitopes, T-cell epitopes and class II binding sites (Box). Epitopes are the structural determinants recognized by either antibodies or T-cell receptors. B-cell epitopes bind to membrane bound antibodies on the surface of B-cells and stimulate the production and secretion of antibodies. This can occur independent of T-cells, but the antibody response (both in amount and affinity) tends to be lower. Polysaccharides are generally T-independent immunogens in that they are characterized by the repetition of the same antigenic determinant (i.e., the saccaride subunits) and are somewhat resistance to degradation. B-cell epitopes in proteins tend to be surface exposed hydrophilic residues and usually consist of 4-8 amino acid residues

T-dependent immunogens are characterized by a few copies of a complex mixture of epitopes as exhibited by proteins. T-dependent antigens are also processed into small peptides and expressed on the surface of antigen-presenting cells via the MHC class II proteins. For example, macrophages will take up foreign organisms or particulate matter and degrade these substances with hydrolytic enzymes in the lysosomes. Immunogens in particulate form are better since they are more efficiently phagocytosed by macrophages. Proteins will be

proteolytically processed into peptides, some of which will bind to the class II MHC molecules and transported to the cell surface. Thus a good immungen needs to contain peptide sequences which bind to the MHC protein. This also explains genetic restrictions in the immune response since individuals have different repertoires of MHC proteins and will react differently to immunogens.

Potential Problems Associated with Immunogens and Possible Remedies

Property	Defect	Result	Remedy
Lacks B-cell epitope	No B-cell recognition	No response	None
Lacks Class II site	No presentation	No response or only primary response	Conjugate with class II site or switch animals
Lacks T-cell epitope	No T-helper involvement	No response or only primary response	Conjugate with T-epitope or switch animals
Non-degradable	No presentation	No response or only primary response	None
Small size	No T-helper involvement	No response	Conjugate with carrier (hapten)
Non-particulate	Poor phagocytosis	Weak response	Self-polymerize or couple to beads



A macrophage with MHC associated peptides on its surface is an antigen presenting cell. Antigen presentation is not restricted to macrophages and other types of cells can also function as antigen-presenting cells. In the MCH-bound state the peptides (i.e., T-cell epitope) are recognized by a T-cell receptor found on the surface of T-lymphocytes. T-cell receptors exhibit a high degree of specificity in regards to the peptide sequence recognized and only those T-cells expressing the correct receptor will be activated. The exact immune response will depend in part on the nature of the activated T-cells. For example, type 2 T-helper cells (T_h2) will secrete cytokines that stimulate B-cells in the presence of antigen to replicate and then differentiate into the antibody-producing plasma cells. Some of the activated B-cells will differentiate into memory cells. Upon a second encounter with the immunogen memory cells more rapidly differentiate into plasma cells resulting in a faster and more intense antibody response. Thus, T-dependent immunogens are also characterized by 'immunologic memory' and a strong secondary response (i.e., boosting). Immunogens lacking class II sites or T-cell epitopes can be conjugated to good MHC-binding sites or T-cell epitopes.

IMMUNOGEN PREPARATION

The preparation of the immuogen and the immunization protocol can affect the nature of the antibody produced. How the antibody will be used in subsequent assays and the goals of the experiments need to be considered before purifying the antigen and immunizing animals. Typically a purified protein is used as an immunogen. A major decision concerns how pure the antigen needs to be before starting the immunization schedule. Ideally an immunogen should be purified to homogeneity for the production of monospecific antisera. Partially purified or crude immunogens can also cause problems with suppression and 'antigenic competition' in that some proteins are more immunogenic than others. This can result in a strong immune response against a minor contaminant in the preparation.

Synthetic peptides and recombinant proteins can be used in situations where it is not possible to purify sufficient protein of the desired purity. In addition, the use of recombinant proteins or synthetic peptides will permit the production of antibodies against particular epitopes or domains of the protein. Synthetic peptides or other small chemicals, often called haptens, need to be coupled to a carrier, or larger protein, before immunization. Conjugation to a carrier ensures an adequate immunogen size, increased haptenic epitope density, and source of heterologous T-cell epitopes. The haptens or peptides are chemically cross-linked to the carrier. The recombinant DNA approach (i.e., genetic engineering) is also amenable to the inclusion of good T-cell epitopes and class II MHC binding sites.

Native	**vs**	**Denatured**
• conformation epitopes		• sequential epitopes
• more immunogenic		• more highly purified

The structure of the immunogen is another factor to be considered since the nature of the epitope recognized by an antibody will affect its performance in the various immunoassays. Immunization with non-denatured (i.e., native) proteins tends to produce antibodies against conformation epitopes, whereas immunization with denatured proteins tends to produce antibodies against sequential or linear determinants. Native proteins tend to be more immunogenic than denatured proteins. But it is generally easier to obtain a higher degree of purity under denaturing conditions.

In cases where antibodies that recognize denatured protein are satisfactory, SDS gel electrophoresis can be used as the final step in the purification protocol. The protein band of interest is excised from the gel and minced into small pieces (~1 mm^3). The gel pieces are placed into a mortar, frozen with liquid nitrogen, and ground into a fine powder with the pestle. A slurry is formed by resuspending the gel in saline and used for immunization. Protein will be slowly released from the gel and result in a depot effect and therefore adjuvants are generally not needed.

IMMUNIZATION

Choice of Animal. A wide range of vertebrate species can be used for the production of antibodies (Table), but the two most often used animals are mice and rabbits. Rabbits react to most immunogens and relatively large quantities of antisera are obtained. The disadvantages of rabbits are that larger quantities of antigen are needed and rabbits often exhibit a high background response to heterologous antigens. In particular, high levels of antibodies to *E. coli* antigens will complicate the analysis of antibodies against recombinant proteins.

Animal	Advantages	Disadvantages
Rabbits	• large amounts of sera • strong antibody response • easy to maintain	• require more immunogen • high background
Mice/Rats	• inbred strains • monoclonal antibodies possible • easy to maintain	• small amounts of sera
Guinea pigs	• generally high titer antibodies • easy to maintain	• difficult to bleed • small amount of sera
Goats/Sheep	• very large amounts of sera	• difficult to maintain
Chickens	• good for some highly conserved mammalian antigens	• difficult to maintain

Mice are easy to handle and several different genetic strains are available. The different genetic strains provide a more uniform response to immunogens and also some control of the immune response since the different strains may respond differently to the same antigen. In addition, mice are used models for many infectious diseases. The major disadvantage of mice is that a small amount of antisera is obtained.

Adult animals should be used for immunizations (i.e., rabbits 6-10 lb. and mice > 7 weeks). If possible, 2-3 rabbits and 5-10 mice should be immunized with the same immunogen to control for animal-to-animal variation.

Main Classes of Adjuvants

Adjuvant	Comment
aluminum salts	The immunogen is absorbed to a gel-like precipitate of aluminium hydroxide, aluminium phosphate, or alum. Action is due to efficient uptake by macrophages and slow release.
water-in-oil emulsion	Aqueous microdroplets containing the immunogen and stabilized by a surfactant are in a continuous oil phase such as mineral oil or squalene. Immune modulators (eg., mycobacteria, MDP) are often incorporated into the emulsion. The emulsion provides a good short term depot effect and is especially useful for soluble immunogens.
oil-in-water emulsion	The immunogen is associated with the surface of microdroplets of oil (eg., sqalane) in a continuous aqueous phase and stabilized by surfactants (eg., Tween 80). Highly suited for lipophilic or amphipahtic immunogens since it is important for the immunogen to be incorporated into the oil phase.
saponin	Saponins form complexes with proteins and lipids resulting in an open cage-like matrix (i.e., micelles) which are sometimes referred to as immunostimulating complexes (ISCOM).
liposomes	Immunogens can be incorporated into or absorbed to liposomes (single or multilayered bilayer vesicles) composed of phospholipids and cholesterol. Immune modulators such as LPS and MDP can also be incorporated.

Adjuvants. Adjuvants are substances that enhance the overall immune response. Most adjuvants contain two components. One component will form a deposit that prevents the rapid breakdown of the antigen and promotes a slow release of the antigen, called the 'depot effect'. There are several different classes of adjuvants with distinct mechanisms by which the depot effect is evoked (Table).

The second component of many adjuvants non-specifically stimulates the immune system (mitogenic or polyclonal activation). For example, killed bacteria, especially mycobacteria, or bacteria components are included in some adjuvants. Common bacterial components included in adjuvants are lipopolysaccharide (LPS) or muramyl didpeptide (MDP), a synthetic analogue of the adjuvant-active component from the cell walls of mycobacteria. These substances stimulate macrophages and other immune effector cells to release cytokines which then enhance the antibody response. The immunomodulatory component may also influence the class of immunoglobin produced as well as the isotype of IgG.

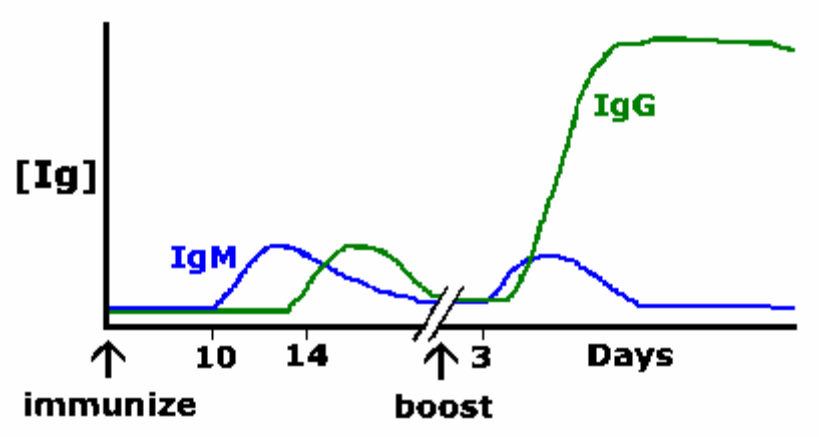
(Review of adjuvants: Cox and Coulter. 1997. Adjuvants – a classification and review of their modes of action. Vaccine 15, 248-256.)

Injection. It is possible to immunize by several different routes (Table). The route of injection can also affect the antibody response, both in terms of quantity of antibody produced and in the qualitative aspects of the antibody. The volume of immunogen and how quickly the immunogen should be released into the lymphatics or circulation will influence the choice of injection method. Rabbits are generally immunized subcutaneously at multiple sites to stimulate regional lymph nodes. Immunization of mice with large volumes can only be accomplished intraperitoneally. Intramuscular and intradermal injections are best for slow release. However, intradermal immunizations are difficult to do correctly. Intraveneous immunization is only used for boosting (see below) and immunogens containing adjuvants should not be given intravenously.

Route		Comments
subcutaneous	sc	easy injections
intramuscular	im	slow release
intradermal	id	difficult injections, slow release
intravenous	iv	only for boosting, no adjuvants
intraperitoneal	ip	easy injections, common in mice

The antigen dose will depend on the animal, the purity and the immunogenicity of the antigen. In general, 0.05-1 mg and 5-50 μg of pure antigen are needed for rabbits and mice, respectively. To obtain the optimal antibody response (i.e., highest titer and affinity) the animals will need to be reimmunized, or boosted. Boosting promotes class switching (eg., IgM $\rightarrow$ IgG), higher levels of antibody, and an increase in the affinity of the antibody for the antigen. Typically, the boosting is carried out at intervals of 2-6 weeks and 3-4 total immunizations are given.

Initial contact with a new antigen may evoke a primary response. The primary response is characterized by a initial rise in IgM followed by IgG (Figure). Antibodies following the primary response are generally transient and will disappear from the circulation. Subsequent exposure (eg., booster injection) to the same antigen results in a much faster antibody response as well as significantly higher levels of antibodies which are predominantly IgG. Furthermore, the antibodies associated with this secondary response persist for a much longer time in the serum. This additional exposure to the antigen also results in a selection for plasma cells producing antibodies with a higher affinity for the antigen. This affinity maturation results in antibodies which react more strongly with the antigen.



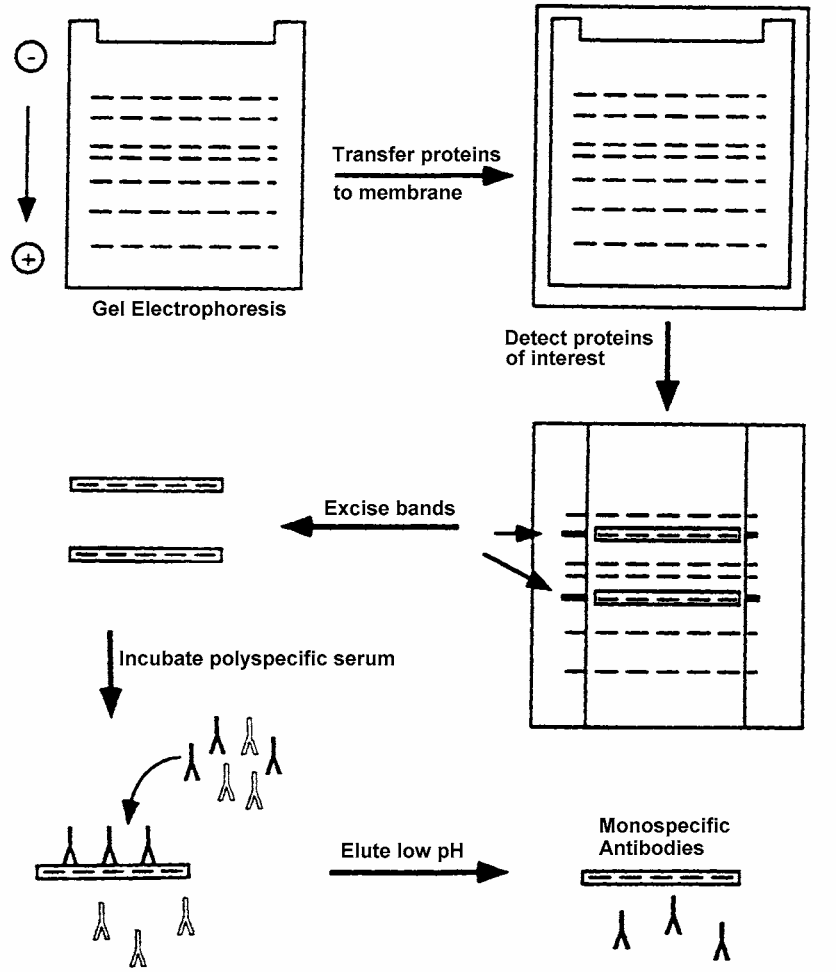
COLLECTING AND PROCESSING BLOOD

Serum samples are typically collected 4-10 days after an immunization and tested for antibody titer. Clotting factors interfere with many assays, and therefore, serum is usually preferable to plasma (total non-cellular component of blood). Serum is prepared by collecting the blood without anticoagulants and allowing the blood clot for at least 30-60 min at 37^{o}. Blood clotting can also be performed for 2-4 hours at room temperature or overnight at 4^{o}. Incubation overnight at 4^{o} will usually produce more hemolysis. The clot is removed and the sample centrifuged to remove any remaining blood cells. In some applications it is also beneficial to inactivate complement by heating the serum at 56^{o} for 30 minutes.

ANTIBODY PURIFICATION

Whole sera can be used for most immunological assays and further purification of antibodies is generally not necessary. Antibodies can be concentrated and partially purified by precipitation with 40-50% saturated $(NH_4)_2SO_4$. This will remove a substantial portion of serum albumin--the major component of serum. Higher levels of purity can be obtained after anion exchange chromatography on DEAE. The immunoglobulins elute from the column before serum albumin. Affinity chromatography using protein A or G can also be used. Protein A and G are bacterial proteins that bind IgG. The various IgG subclasses from different animals exhibit a range of affinities for these proteins and can be differential eluted from the affinity columns at different pH values.

Specific antibodies can also be purified from polyclonal sera. For example, affinity purification using the antigen as ligand can be used to isolate the antibodies specific for that protein. Monospecific antibodies can also be generated by taking advantage of the high resolution of gel electrophoresis (see Figure below for example). (See section on Immunoblotting for a detailed discussion of the methods.) It is also possible to preabsorb serum to remove unwanted antibodies and to reduce background. Examples include *E. coli* proteins or the fusion partner in recombinant proteins.



CHAPTER 14--MONOCLONAL ANTIBODIES

Serum contains a mixture of antibodies recognizing many different antigens. Furthermore, different epitopes of any particular antigen are also recognized. These non-specific reactions and cross-reactivity may limit the usefulness of antisera. This heterogeneity is the result of antibodies being the products of clonal lineages of mature B-cells (see appendix for B-cell differentiation). Each clonal lineage of a B-cell produces a specific antibody that recognized a specific epitope. It is not possible to isolate and grow B-cells in culture.

In 1975 Kohler and Milstien described a technique that allowed the growth of clonal populations of cells secreting antibody of defined specificity, for which they received the 1984 Nobel prize in medicine. The method consists of fusing B-cells with **myeloma cells** (a type of B-cell tumor). The fused cell is known as a **hybridoma**. This hybridoma now has the ability to grow in culture and it secretes an antibody of defined specificity (phenotype acquired from the B-cell). The antibody produced by the hybridoma is referred to as a **monoclonal antibody** (mAb).

- Advantages of mAbs
  - ♦ unlimited supply
  - ♦ defined reagent
- Disadvantages of mAbs
  - ♦ single epitope
  - ♦ time consuming

The major advantage of a mAb is its unlimited supply (Box). The hybridoma cell line has indefinite life whereas only a limited amount of serum can be obtained from immunized animals. In addition, the mAb is a defined reagent that recognizes a single epitope. Individual animals may respond differently to the same antigen resulting in potentially different polyclonal antibodies against the same antigen. Another advantage is the ability to produce specific antibody with impure antigen. A disadvantage of mAbs is that only single epitope recognized and the mAb may not have desire specificity or affinity. Secondly, mAb are time consuming (6-9 months) to produce and requires tissue culture facilities. Following immunization there are several steps (Box) involved in the production of monoclonal antibodies.

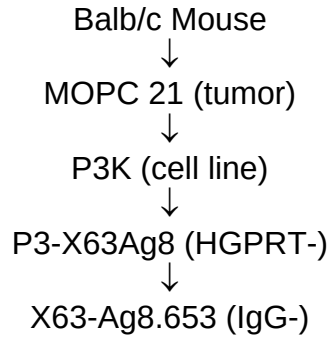
1. Immunization
2. Fusion
3. Selection
4. Screening
5. Cloning
6. Production

IMMUNIZATION

Immunized animals are needed for the first step in mAb production (Box). The same basic factors (i.e., immunogen preparation, adjuvants, etc.) for the generation of polyclonal sera (see chapter on Immunization) apply to the production of mAbs. However, rodents, and especially mice, are almost always the animal of choice. This is primarily due to the availability of mouse myeloma cell lines, but human and simian hybridomas have been successfully produced by transforming the lymphocytes with Epstein-Barr virus. If future plans include the production of ascites tumors then the mouse strain used for the immunization should match the mouse strain from which the myeloma cells were derived. In some situations it is not necessary to use highly purified immunogens. For example, it is also possible to use animals immune to an infectious disease or to immunize with crude antigens. Monoclonal antibodies with the desired specificities are then selected during the screening stage.

The mice are immunized, boosted and tested until an antibody response of the desired specificity is elicited. Generally the monoclonal antibodies obtained will be reflective of the antibodies found in the polyclonal sera. Therefore it is critical that the mouse exhibits antibodies of the desired specificity before proceeding with the fusion.

FUSION



Several myeloma cell lines are available. A common mouse cell line used for the production of mAbs is X63Ag8.653. The parental cell line was initially derived from a tumor induced by injecting mineral oil into a Balb/c mouse. The resulting cell line was subjected to mutagenesis to produce a hypoxanthine-guanine phosphoribosyl transferase (HGPRT) deficient cell line. Further cloning yielded X63Ag8.653 which no longer secretes its own antibody. Therefore hybridomas prepared with this cell line will only secrete antibodies obtained from the spleen cells.

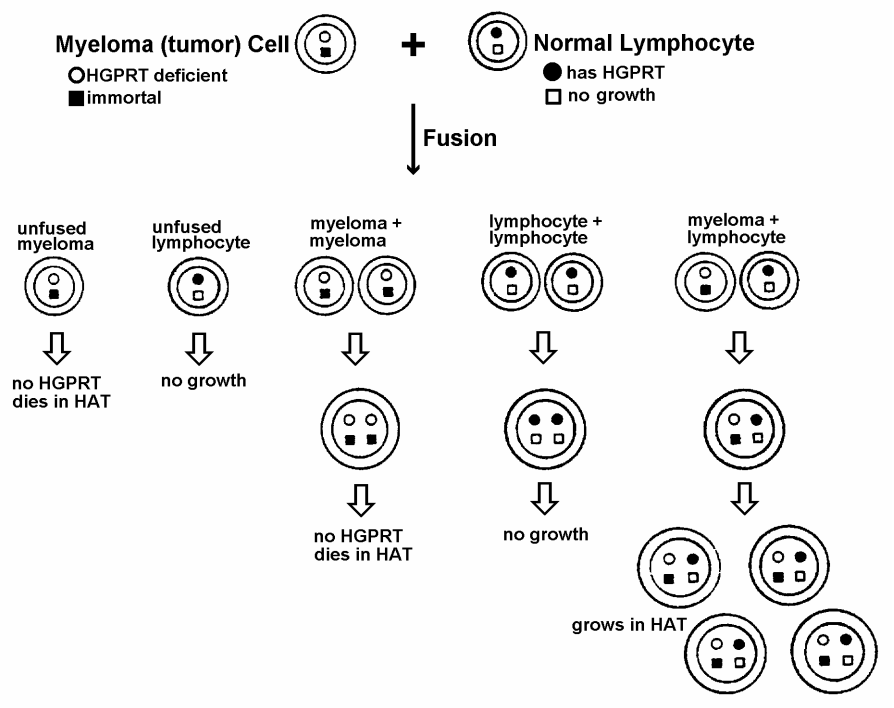
B-cells are isolated from the immunized mouse. The mouse exhibiting the desired antibody response is usually boosted 3-4 days before doing the fusion to stimulate the B-cells. The spleen is most common source of B-cells, but it is also possible to use lymph nodes. Spleens or lymph nodes are removed from immunized mice and teased apart to prepare a single cell suspension which are mixed with the myeloma cells derived from in cultures. It is important to have healthy and actively dividing (i.e., log phase) myeloma cells on the day of the fusion.

A fusogenic agent is added to the mixed myeloma and spleen cells. Polyethylene glycol (PEG) is the most common agent used for the fusion of cells. Other fusogenic agents will work, though. The spleen cells and myeloma cells are mixed and co-pelleted by centrifugation. A small volume of PEG solution in media is added and the cell pellet is gently dislodged. PEG fuses the plasma membranes of adjacent cell resulting in a heterokaryon. PEG is toxic and exposure needs to be minimized. The cells are washed to remove the PEG, resuspended in HAT medium (see below), and then generally plated in 24-well plates.

SELECTION

Not all of the cells will fuse following PEG treatment and many of fusions will be homotypic. The unfused myeloma cells will tend to overgrow the fused cells in culture, and therefore, it is necessary to select for fused cells. The use of hypoxanthine-guanine phosphor-ribosyl transferase (HGPRT) deficient myeloma cells is a commonly used genetic selection scheme for selecting fused cells. HGPRT is a crucial enzyme in the purine salvage pathway. Selection of fused cells is achieved by adding hypoxanthine, aminopterin and thymidine (HAT) to the culture medium. Aminopterin is an inhibitor of dihydrofolate reductase (DHFR) which is a crucial enzyme for both de novo purine and pyrimidine synthesis. Hypoxanthine and thymidine are nucleotides utilized in the purine and pyrimidine salvage pathways, respectively. HGPRT-deficient myeloma cells will die in the presence of aminopterin since they are and cannot salvage purines and de novo purine synthesis is inhibited. Spleen cells contain an intact HGPRT and can therefore salvage, but do not have the ability to grow in culture. Therefore,

only hybridomas (i.e., fusions between myelomas and spleen cells) are able to grow in the presence of HAT medium. The myeloma cells provide the ability to grow in culture and the spleen cells provide a functional HGPRT. The HAT medium is usually replaced with HT medium after 7-10 days after the fusion to minimize the aminopterin toxicity. Hypoxanthine and thymidine are still necessary, since aminopterin slowly decays and is still present in the medium. The plates are monitored for the next 2-4 weeks for the appearance of colonies.



SCREENING

The goal of the screening step is to identify wells containing the mAb of desired specificity. The wells may contain hybridomas that do not secrete antibody or secrete irrelevant antibody. Media from wells exhibiting cell growth are removed and tested for antibody. The immunoassay that is used for screening needs to be worked out before reaching this point. Cells producing mAb of desired specificity are expanded and stabilates frozen.

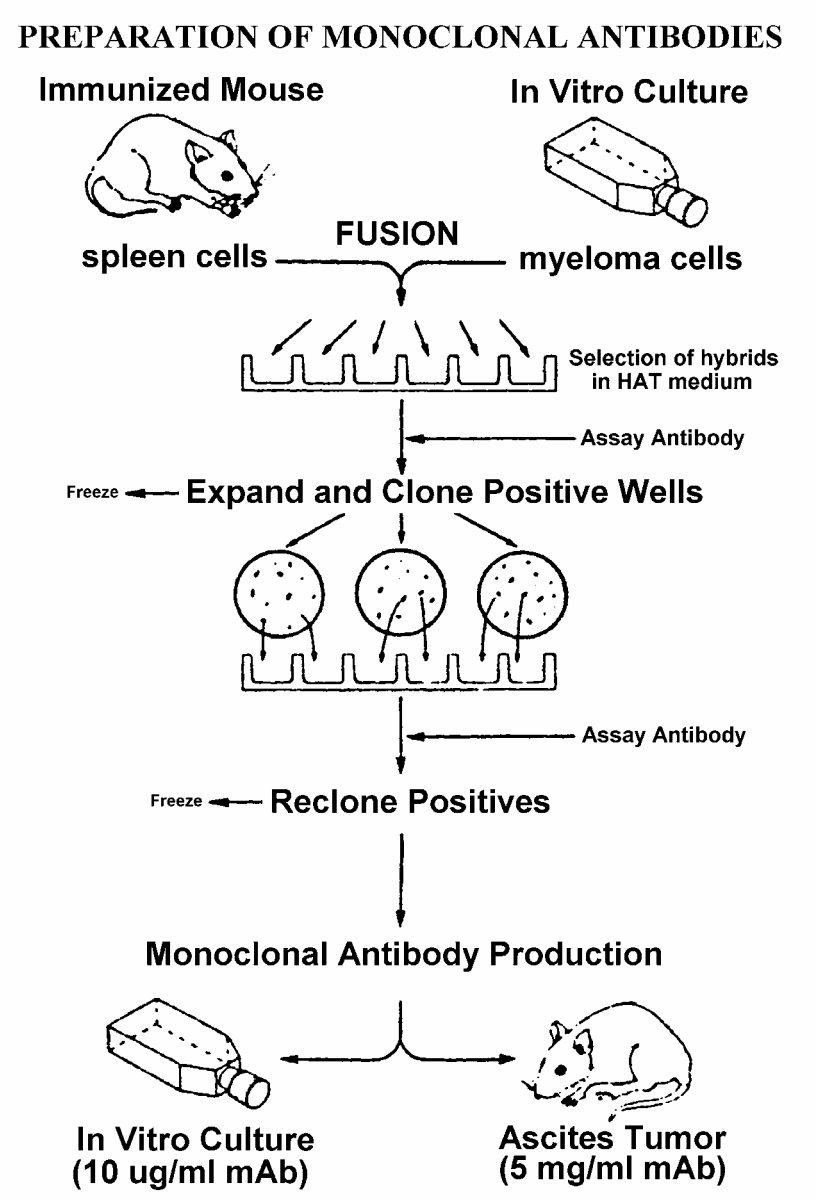
CLONING

Wells will sometimes contain mixtures of hybridomas. In addition, some of hybridomas will lose the ability to produce antibody due to chromosome loss or rearrangement. Cloning the hybridoma of interest will prevent irrelevant hybridomas from overgrowing the hybridoma of interest. Two common methods utilized for cloning hybridomas are soft agar and limiting dilution. In soft-agar cloning the hybridomas are plated in a semi-solid medium such as a low concentration of agar. Single cells will yield colonies that can be expanded. The advantage of soft-agar cloning is that a colony originating from a single cell is selected. However, many hybridomas will not grow in soft agar. In limiting dilution cells are plated in 96-well plates at a calculated concentration of one cell per well. With both methods it is recommended to subject

the hybridoma to several rounds of cloning until a stable hybridoma is obtained.

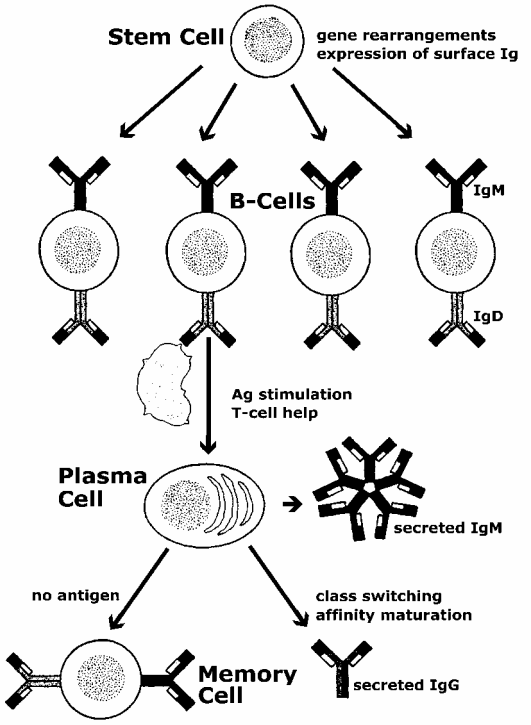
PRODUCTION

Large quantities of mAbs can be produced by either *in vitro* culture or through the generation of **ascites** tumors. mAb produced *in vitro* is sometimes too dilute and will generally be contaminated with bovine serum components. Therefore, it may be necessary to purify and/or concentrate the mAb. Ascites are produced by injecting the hybridoma into the peritoneal cavity of a mouse. The hybridoma cells will produce an ascites tumor and secrete mAb into the peritoneal cavity. The ascites fluid is then collected several weeks later. Ascites can produce very high concentrations of antibody. The ascites fluid will be slightly contaminated with other mouse IgG.



APPENDIX. B-CELL DIFFERENTIATION

Lymphoid stem cells found in the bone marrow undergo gene rearrangements in the variable regions of the Ig-heavy and the Ig-light chain genes. Cells undergoing heavy chain rearrangements are called progenitor (pro) B-cells and those undergoing rearrangements of the light chains are precursor (pre) B-cells. These rearranged genes will then be expressed as IgD and IgM on the surface of the mature B-cells. The mature B-cells are a heterogeneous population of cells which all express different rearrangements of the immunoglobulin (Ig) genes. Each rearrangement results in an immunoglobulin with a different epitope specificity.

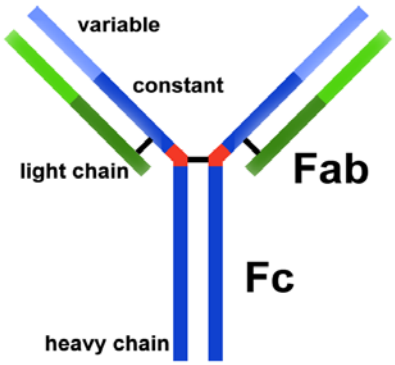


These naive B-cells (not yet exposed to antigen) are found in the secondary lymphoid tissues where they will be exposed to potential immunogens. Binding of antigen to the surface immunoglobulin will stimulate those particular cells to proliferate and differentiate into plasma cells. The activation of most B-lymphocytes will also require the participation of activated T-helper cells. In particular, Th_2 cells secreting IL-4, IL-5, IL-6 and IL-10 stimulate the proliferation and differentiation of the B-cells. Bound antigen is also taken up by receptor-mediated endocytosis, processed and presented on the B-cell surface. These B-cells function as antigen presenting cells and stimulate T-cells.

The plasma cells no longer express surface immunoglobulin and begin to secrete IgM. The switch from membrane IgM to secretory IgM involves an alternate mRNA splicing. Plasma cells can also undergo a class switching resulting in the production of IgG (primarily in blood), IgA (primarily associated with mucosa), or IgE (associated with allergenic cells). IL-2 and INF-γ are needed for class switching. Immunoglobulin class switching can be accomplished at either the DNA level (i.e., gene rearrangement) or at the mRNA level (i.e., differential splicing). An affinity maturation due to mutations in the hypervariable region is also observed.

The disappearance of the antigen leads to apoptosis of many of the plasma cells, or a conversion to memory cells. A second encounter with the same antigen leads to an activation of these memory cells and a more rapid response and a higher steady-state level of antibody. In addition, the secondary response is primarily IgG.

CHAPTER 15--IMMUNOASSAYS



Due to their specific and high affinity interactions with epitopes, antibodies are powerful reagents for the analysis of proteins or other antigens. Antibodies are composed of conserved regions and variable regions. The variable regions are responsible for recognizing epitopes within antigens. In the case of proteins, the epitopes are typically 4-6 amino acids and are usually specific for a particular amino acid sequence. The antibody molecule is made up of heavy chains and light chains which are joined by disulfide bonds. The heavy chain defines the type of immunoglobulin. For example, IgM and IgG contain μ and γ heavy chains respectively. In addition, different γ chains define subclasses of IgG. The light chains (κ and λ) are shared between the different immunoglobulin classes and subclasses. IgG is composed of two functional domains separable after digestion with papain: F_{ab} and F_c. The F_{ab} fragment contains the epitope binding domain or the antibody activity, and the F_c domain is characterized by complement binding in some IgG subclasses.

One possible application of antibodies is affinity chromatography. Antibodies are covalently attached to a matrix such as agarose to make affinity columns. More often, though, antibodies are used in various analytical assays. These include the enzyme-linked immuno-sorbent assay (ELISA), the radioimmunosorbent assay (RIA), immunofluorescence (IFA), immunoblotting (also called Western blotting) and immunoprecipitation. Immunoassays are used to detect, quantify and characterize antigens. The assays can also be used to detect and quantify antibodies.

- directly label 1^o antibody
- anti-antibody (2^o Ab)
- protein A or protein G

The general procedure used in all immunoassays is 1) to allow antibody-antigen complex formation, and 2) detection of the antibody-antigen complex. Antigen-antibody complexes are formed by incubating the antibody with an extract containing the antigen. Alternatively, the antigen is absorbed or fixed to a solid support and then antibody binds to the immobilized antigen. The detection of the complex will depend on the assay being perform (Box). One method is to directly label the antibody (or antigen) with a detectable marker (eg., radioactivity, fluorescence, or enzyme). Alternatively, a labeled **secondary antibody** (prepared in another species) that recognizes the **primary antibody** can be used to detect the complex. Similarly, either protein A or protein G labeled with a marker could be used to detect the antigen-antibody complex. Protein A and protein G are proteins isolated from *Staphylococcus areus* or *Streptococcus*, respectively, that bind to the F_c-region of IgG. Protein G recognizes more immunoglobulin subclasses, as well as a wider range of species, than protein A.

IMMUNOPRECIPITATION

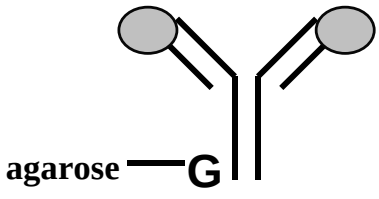
Immunoprecipitation involves the purification of antigen-antibody complexes, which are then analyzed by gel electrophoresis. This method provides a means to identify and characterize proteins in conjunction with gel electrophoresis. Initially, the technique involved the formation of large supramolecular antigen-antibody complexes due to polyclonal sera that recognized multiple epitopes on the protein. These large protein complexes are collected by centrifugation. Formation of the large complex is dependent on high concentrations of both antigen and antibody and the complexes will often contain contaminating proteins. The ability to purify the antigen-antibody complex with bacterial proteins that bind immunoglobulin has greatly increased the sensitivity and specificity of immunoprecipitation.

1. ± Label the Ag
2. Solubilize the Ag
3. Mix and incubate Ag and Ab
4. Isolate Ag-Ab complex
5. Electrophoresis/fluorography

A typical first step in immunoprecipitation is to radiolabel the protein(s). This is easily accomplished in situations where in vitro cell culture is possible by metabolically labeling the protein with radioactive amino acids. Proteins can also be labeled by chemical means such as iodination as long as the treatment does not adversely affect the epitopes recognized by the antibodies. It is not necessary to label the protein if it can be detected by other means such as protein staining or immunoblotting (see combined immunoprecipitation and immunoblotting below).

Proteins are then solubilized under conditions which do not affect the ability of the antibody to recognize the epitope. For example, some antibodies recognize 'conformational' epitopes and denaturing agents will destroy the epitope. Generally, non-ionic detergents (eg., Triton X-100) do not adversely affect the immunoprecipitation assay. It still may be possible to use immunoprecipitation for the analysis of proteins which are only soluble under denaturing conditions if a non-conformational dependent (i.e., linear) epitope is recognized by the antibody. However, the denaturing agent will have to be removed before adding the antibody since the structure of the antibody must be maintained for epitope recognition. For example, proteins can be solubilized with SDS followed by the addition of a 10-20 fold excess of Triton X-100 (or another non-ionic detergent). The excess SDS will be incorporated into mixed micelles and therefore will have little affect on the antibody when it is added. Enough antibody should be added to precipitate all of the antigen in a sample, especially if quantitative interpretations are going to be made.

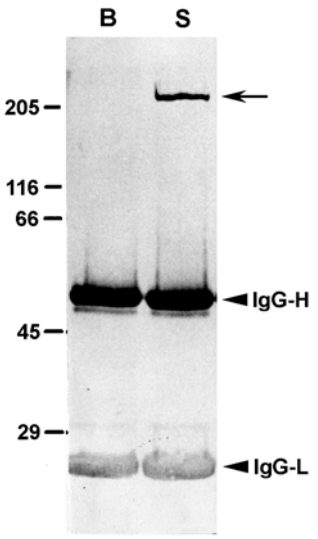
In general, the formation of antigen-antibody complexes is rapid and incubations of 30-60 minutes are sufficient for maximum recovery. Longer incubations may be necessary for low concentrations of the target protein. Proteolysis can also be a problem during the incubation step. Protease inhibitors should be added and the incubations carried out on ice to minimize losses.



The isolation of the antigen-antibody complex is accomplished by an affinity purification with protein A or protein G conjugated to agarose beads. Historically, fixed intact *Staphylococcus areus* bacteria were used to isolate Ag-

Ab complexes. Protein A and protein G bind to the F_c region of IgG with an high affinity. In general, protein G binds with a higher affinity to a wider range of immunoglobulin subclasses as well as recognizing immunoglobulins from more species than protein. These bacterial proteins have a higher affinity for antibody with bound antigen and are therefore the protein A (or G) agarose is usually added after the formation of the antigen-antibody complex. But all three components (antigen extract, antibody, and affinity matrix) can be mixed simultaneously. Generally, incubation for 30-60 minutes with continuous rocking is sufficient for complete recovery of the antigen-antibody complexes.

The antigen-antibody complexes bound to the protein A (or G) are collected by centrifugation and washed several times to remove contaminating proteins. Antigen-antibody complexes are then eluted from affinity matrix by boiling the agarose beads in electrophoresis sample buffer containing SDS and β-mercaptoethanol. Bound proteins can also be eluted with low pH and quickly neutralized if non-denatured protein is need for subsequent analyses. The sample is then subjected to SDS gel electrophoresis and immunoprecipitated proteins are detected by autoradiography or fluorography depending on the isotope. It may also be possible to detect abundant antigens by protein staining. Alternatively, it may also be possible to transfer the electrophoresed proteins to a membrane and carry out immunoblotting (see below).



IgG heavy and light chains will also be contained in the sample and can be observed as approximately 55 (IgG-H) and 25 (IgG-L) kDa bands (Figure). Most of the time this will not cause any problems unless the target protein is of a similar size and the immunoglobulin subunits interfere with the protein of interest.The immunoglobulin bands are easily control for by including an immunoprecipitation blank (B) which contains everything (i.e, antibody and buffers) except the sample (S) being analyzed. Bands appearing in the sample (arrow), but not the blank will be the protein(s) recognized by the antibody.

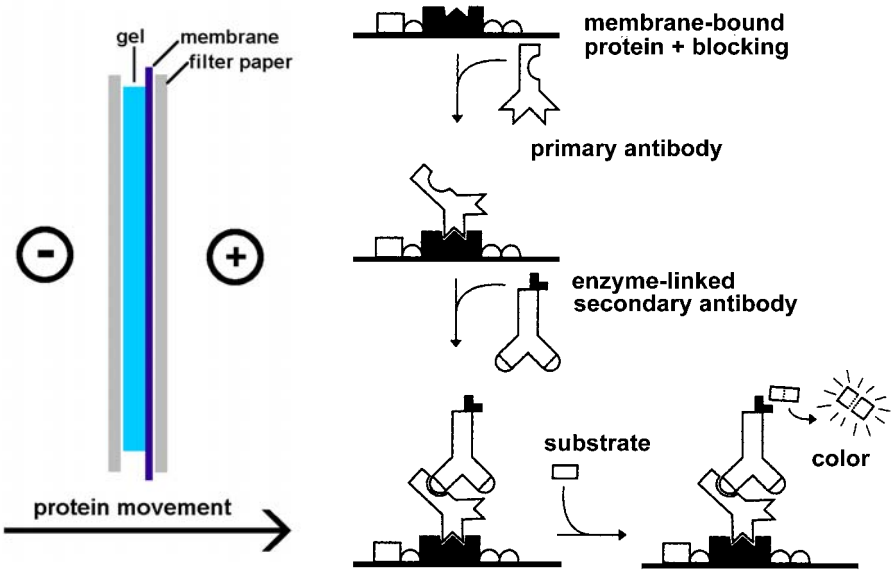
Immunoprecipitation can be used to detect a protein in a biological sample and in conjunction with SDS-PAGE to determine its molecular weight. Since the formation of the Ag-Ab complexes can be carried out under non-denaturing conditions it is also possible to determine the quarternary structure of a protein and to analyze proteins that bind to a target protein. Other proteins binding to the protein recognized by the antibody will be co-immunoprecipitated with the target protein if the protein-protein interactions are not disrupted by antibody binding (or visa versa). Generally the information obtained by immunoprecipitation is qualitative. To obtain reliable quantitative information is will be necessary to titrate the antibody and to run controls demonstrating that all of the available target protein has been immunoprecipitated.

IMMUNOBLOTTING

1. Antigen electrophoresis
2. Transfer to membrane
3. Blocking
4. Incubate with 1° antibody
5. Wash
6. Incubate 2° antibody
7. Wash
8. Develop with substrate

Immunoblotting, also called Western blotting, is another technique used in conjunction with gel electrophoresis. In this case, proteins are first subjected to gel electrophoresis, transferred to a membrane support and the protein of interest is detected with an antibody.

The first step in immunoblotting is antigen preparation and electrophoresis. SDS-PAGE is the most common electrophoretic system, but other gel electrophorectic systems can also be used. After electrophoresis the separated proteins are transferred to a membrane support such as nitrocellulose or PVDF (Immobilon-P). PVDF membranes bind more protein and are generally superior to nitrocellulose. Before the transfer the gel is usually equilibrated with the transfer buffer containing methanol to remove excess SDS. Excessive SDS can inhibit the binding of proteins to the membrane. The transfer of proteins from the gel to the membrane is carried out electrophoretically in a special apparatus. The gel is positioned between two electrodes that generate an uniform electric field across the thickness of the gel (Figure). A membrane is placed on the anode side of the gel. Filter papers are used to maintain buffer contact with the gel. Proteins migrate out of the gel and a replica of the gel is formed on the membrane. The completeness of the transfer can be assessed through the use of prestained proteins as size markers or by staining the gel after the transfer.



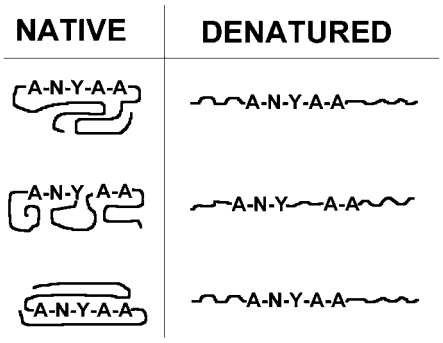
Following the transfer, the membranes are washed with a buffer which removes any loosely bound proteins and which will 'block' the non-specific binding of proteins (i.e., antibodies). Two common blocking agents are the detergent Tween-20 and a powdered-milk solution, called BLOTTO in lab jargon. The detergent (or milk proteins) will bind any free sites on the membrane that are not already bound with protein and prevent non-specific antibody binding. The membrane is then incubated with the primary antibody (diluted with the blocking buffer). The dilution and time of incubation will depend on antibody affinity and the amount of

the target protein on the membrane. Typically incubations are carried out for one hour at room temperature. Alternatively, overnight incubations at 4^{o} can be carried out.

The membranes are then washed extensively to remove any non-specifically bound antibody followed by an incubation with the secondary antibody conjugated with an enzyme. This secondary antibody recognizes immunoglobulin from the species of the primary antibody and are purchased. The secondary antibody will bind the primary antibody which is bound to the antigen. The membranes are then incubated with substrates of the enzyme which will form insoluble colored products (see Table in section on ELISA for examples). This results in a colored precipitate being formed on the membrane in the position of the target protein. Chemiluminescent substrates can also be used instead of the precipitating chromogenic substrates (see section on non-radioactive probes in chapter on nucleic acid blotting). Alternatively, ^{125}I-Protein A in conjunction with autoradiography can be used to detect the Ab-Ag complex.

Both immunoblotting and immunoprecipitation can identify proteins recognized by an antibody in conjunction with SDS gel electrophoresis. The choice of the assay will be affected by the experimental design, antigen concentration, antigen solubility and the ability of the antibody to recognize the antigen in the two assays. As an example of experimental design, immunoblotting will provide information about the total amount of antigen present, whereas immunoprecipitation of metabolically labeled proteins can provide information about the newly synthesized protein and/or its rate of synthesis. The limited amount of protein which can be electrophoresed may preclude the use of immunoblotting for the analysis of low abundance proteins, whereas immunoprecipitation can be used to concentrate low abundance proteins from relatively large volumes.

- Experimental Design
  - eg., synthesis (IP)
- Ag concentration
  - IP better for low abundance proteins
- Ag solubility
  - blot for insoluble proteins
- Ab recognition
  - conformational dependent epitopes
  - 4^{o} structure



Another factor involved in the success of either immunoprecipitation or immunoblotting may be whether the antibody recognizes the protein of interest. In general, native non-denatured proteins are recognized by the antibody in the immunoprecipitation assay, whereas denatured proteins are recognized by the antibody in the immunoblotting assay. The nature of the epitope will determine whether one, or both, assays will work. For example, if the antibody recognizes a linear epitope that is exposed on the surface of the protein, it will likely function in both assays. However, if the antibody recognizes a conformation epitope, then only immunoprecipitation can be used. In the situation where an antibody recognizes a linear epitope that is not exposed in the native protein it will be

necessary to denature the protein before carrying out the immunoprecipitation. The quarternary structure of a protein may also influence the choice of the assay. For example, an antibody that only recognizes one subunit of a multisubunit protein will only produce one band if analyzed by immunoblotting, whereas all of the protein subunits may be detected in the immunoprecipitation assay.

One problem with the immunoprecipitation assay is that it is not always possible to radiolabel the antigen of interest, and therefore detecting the target protein is difficult. Immunoblotting provides a very sensitive method of detecting proteins after gel electrophoresis, but is limited by the amount of protein which can be electrophoresesed. Affinity purification of the target protein by immunoprecipitation and detecting the protein by immunoblotting possibly overcomes these limitations. Following immunoprecipitation and gel electrophoresis, the proteins are transferred to a membrane and analyzed by immunoblotting. This does require that the antibody works in both assays, or that distinct antibodies for each assay are available. A blank immunoprecipitation which contains the antibody, but no antigen-containing extract, should also be prepared. This blank will serve to identify the IgG heavy and light chains which may be detected with the secondary antibodies.

1. Prepare cells or tissue
2. Incubate 1^{o} antibody
3. Wash
4. Incubate 2^{o} antibody
5. Wash
6. View under UV illumination

IMMUNOFLUORESCENCE

Immunofluorescence is used to localize antigens to specific cell types or subcellular compartments. The assay can also be adapted for the detection and quantification of antibody. The general procedure is to incubate cells or tissue with the primary antibody and detect the antigen-antibody complex with a secondary antibody conjugated with a fluorochrome. Fluorescein and rhodamine, which fluoresce green and red, respectively, are two common conjugates. The location of the protein recognized by the primary antibody is determined by examining the sample under UV illumination (see Section of Fluorescence Microscopy). It is also possible to use an enzyme-conjugated secondary antibody and develop with substrates that form precipitates. The location of the precipitates is then detected by conventional microscopy. This variation is not widely used, though.

The preparation of the cells or tissue for immunofluorescence is a crucial step that depends on the type of cells being analyzed and upon the application. As for microscopy in general, tissues need to be sectioned with a microtome and mounted onto microscope slides before immunofluorescence analysis. Adherent tissue culture cells are usually grown on cover slips or microscope slides. Cells grown in suspension can be smeared or dried onto microscope slides or the antibody incubations and washings can be carried out with the cells in suspension.

The fixation procedure should not affect the ability of the antibody to recognize the antigen. Organic solvents, such as ethanol, methanol and acetone, or paraformaldehyde are routinely used. Glutaraldehyde is not a good fixative since the bonds formed by the crosslinking of glutaraldehyde exhibit a strong endogenous fluorescence. Fixation also permeabilizes the cell so that the antibody has access to internal cellular structures. The permeability can be enhanced by treating with a detergent after the fixation. Intact and unfixed cells can be used to

demonstrate the surface localization of an antigen.

The subcellular structures recognized by the antibodies will appear as a bright image against a dark background. Comparing the epifluorescence with the bright field image is used to localize the protein of interest to a particular subcellular compartment or to a particular cell type within a tissue. Fluorescent dyes which bind to particular subcellular compartments can also be used to assist in the localization of the protein of interest. Dual-labeling experiments can also be used to demonstrate the extent of co-localization between two different proteins. The most convenient method is to used antibodies generated in different species (eg., mice and rabbits) as the primary antibody and then used species specific secondary antibodies conjugated with different fluorochromes. If only antibodies raised in the same species are available then it will be necessary to purify the two antibodies and conjugate them with different fluorochromes. The two different fluorochromes can be viewed separately through the use of filters specific for each fluorochrome. It is also possible to digitally merge the two images to demonstrate the extent of co-localization between the two proteins.

IMMUNOGOLD ELECTRON MICROSCOPY

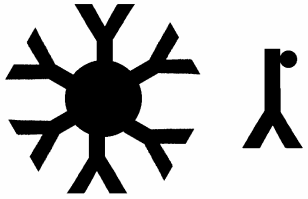
The resolution of immunofluorescence is essentially the same as that for light microscopy. Immunogold electron microscopy can more precisely localize proteins. The principle of immunogold electron microscopy is the same as immunofluorescence in that antibodies are bound to the protein of interest and then the location of these antibodies are determined through the use of secondary antibodies conjugated with colloidal gold. It is also possible to use to use enzyme-conjugated secondary antibodies in combination with substrates that form an electron dense precipitate. In practice, though, localizing antigens at the level of the electron microscope is more difficult and subject to more problems.

Samples need to be fixed and embedded for electron microscopy (see Section on Electron Microscopy). However, these fixation and embedding procedures are quite harsh and typically destroy epitopes. Gentler fixation procedures are generally needed and embedding is done in resins which do not require heating. In general, fixation conditions which result in good preservation of ultrastructure tend to lead to a loss in the ability of the antibody to recognize the antigen. Following fixation and embedding, the samples are sectioned and mounted onto grids. The incubations with primary or secondary antibodies and washes are carried out by floating the grids on drops of buffer. The secondary antibodies are conjugated with colloidal gold. Different sizes of gold particles are available for dual labeling experiments. The gold particles are electron dense and are visible when examined by electron microscopy.

Cryo-electron microscopy can be used to overcome some of the problems associated with the loss of antigenicity during the preparation of samples for electron microscopy. Special micromes and stages which keep the sample frozen in liquid nitrogen are used. This allows for lower concentrations of glutaraldehyde to be used. However, the resolution is much less in cryo-electron microscopy than conventional transmission electron microscopy.

The relatively large size of the gold particles (5-15 nm) poses some accessibility problems due both to the nature of the fixed cells and the embedding polymers. Furthermore,

many antibody molecules bind to a single gold particle resulting in a large complex (Figure). One method to overcome these accessibility problems is to use secondary antibodies conjugated with ultrasmall gold (<1 nm). These small gold particles are not readily visible under the electron microscope and need to be treated with silver which coats the gold particles (i.e., silver enhancement). In addition, the ultrasmall gold conjugated secondary antibodies are more similar to the fluorochrome conjugated secondary antibodies since one gold particle will tend to bind one antibody molecule. This allows for pre-embedding experiments in which the fixed and permeabilized cells are incubated with the primary and seconday antibodies in suspension before the sample is embedded. Therefore it is possible to carry out the assay under conditions simlar to that used for immunofluorescence and avoid the problems associated with embedding.



ELISA (AND RIA)

1. Antigen binding
2. Blocking
3. Incubate 1^{o} Ab
4. Wash
5. Incubate 2^{o} Ab
6. Wash
7. Develop with substrate

The enzyme-linked immunosorbent assay (ELISA) involves absorbing an antigen a solid support such as a membrane (as used in immunoblotting) or a 96-well microplate. Proteins which have not been separated by electrophoresis can be bound to membranes and analyzed with primary and secondary antibodies as in the immunoblotting procedure. Such analyses are often called dot blots. The more common format is to absorb the antigen to the wells of a 96-well microplate and to use substrates that produce a colored soluble product. The ELISA is particularly well suited for the analysis of large numbers of samples and most of the procedure can be automated.

The first step is to bind the antigen to the plate. The antigen range from a purified protein or peptide to a crude protein extract. It is also possible to use whole fixed cells. The choice of antigen will depend on the application and the nature of the antibodies being tested. Binding to 96-well microplates is accomplished by incubating the wells with a solution containing the antigen. Protein binding is due to hydrophobic interactions between the protein and the plastic. After antigen is bound to the plates, they are then blocked to eliminate non-specific binding. Common blocking agents are bovine serum albumin (relatively cheap protein), Tween-20 (detergent), or powdered milk (very cheap protein). The function of the blocking step is to coat any protein-binding sites on the plate that could bind antibody non-specifically.

The antigen-containing wells are then incubated with the primary antibody. The proper antibody dilution must be determined empirically. Serial dilutions are often carried out and the minimum titer giving a positive response will be a quantitative measure of antibody concentration. ELISAs are typically carried out in duplicate or triplicate. It is important to include negative (without antigen) controls on the same plate. Positive controls can also be included in situations where the antigen is being evaluated. After incubation with the primary antibody the wells are washed extensively to remove any unbound antibody.

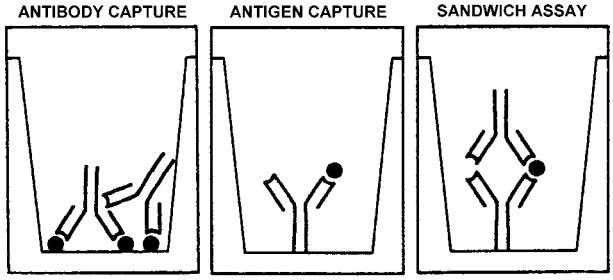
The wells are then incubated with a secondary antibody that is conjugated with an enzyme. Alkaline phosphatase and horse-radish peroxidase are two common enzymes. These reagents are obtained commercially and recommended dilutions will be supplied by the manufacturer. It is possible to buy secondary antibodies that are class or subclass specific (i.e., made against the heavy chain). Second antibodies are made against the whole molecule will usually recognize the light chains which are shared between IgG and IgM. The plates are again washed to remove any unbound antibody. The final step is to develop the plates with a chromogenic soluble substrate (Table). The intensity of the color will be proportional to the amount of secondary antibody which proportional to the amount of specific primary antibody in the sample. Alternatively a constant amount of a defined antibody can be used to quantify the amount of a specific protein in a sample. An ELISA reader is a spectrophotometer designed to read the individual wells in a 96-well microplate. Many ELISA readers are interfaced with a computer to assist in data management.

Examples of ELISA Substrates

	Substrates	
Enzymes	Soluble	Precipitating
AP	pNPP	BCIP/NBT
HRP	TMB	DAB

AP (alkaline phosphatase); HRP (horse radish peroxidase); pNPP (p-nitrophenyl phosphate); BCIP (5-bromo-4-chloro-3-indolyl phosphate); NBT (nitro blue tetrazolium); TMB (3,3',5,5'-tetramethylbenzidine); DAB (3,3-diaminobenzidine)

The RIA is essentially the same as the ELISA, except ^{125}I-protein A is used instead of enzyme-conjugate secondary antibody for the detection of the Antigen-antibody complex. The amount of radioactivity bound to the membrane filters or wells is determined instead of developing with substrates. The RIA was developed first, but the ELISA is more popular since radioactivity is not needed.



Variations to the ELISA can also be carried out. For example, the plates can be coated with the antibody instead of the antigen. The antibody will then 'capture' the antigen when extracts are added to the wells. In this case the antigen will need to be radioactive or have some type of measurable activity. If two antibodies with different specificities are available it may be possible to create a 'sandwich' type of assay. Monoclonal antibodies are especially useful for the development of antigen capture assays.

DIRECT VS. INDIRECT

Direct	vs.	Indirect
• less steps		• convenience
• less bkg (?)		• amplification (?)
• dual label		

In summary, ELISA, immunoblotting and IFA are all techniques in which antigen is immobilized, incubated with primary antibody and the bound antibody is then detected. The bound antibody is usually detected with a second antibody, which recognizes the first antibody and is conjugated with a detectable marker (eg., radioactivity, enzyme, fluorescence). The use of a secondary antibody is referred to as indirect detection.

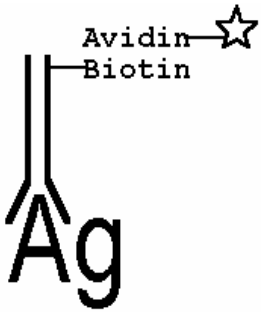
It is also possible to label the primary antibody with the detectable marker and use the primary antibody to directly detect antigen. Direct detection methods are quicker to carry out since an antibody incubation step and the associated washing steps are eliminated. In addition, there is sometimes less background associated with direct detection methods. However, the primary antibody must be first purified and conjugated with the appropriate label. The availability of secondary antibodies from commercial sources tends to make the indirect methods easier to carry out. In addition, some signal amplification may occur during the secondary antibody step.

- Radiolabeling
  - ♦ iodination
  - ♦ metabolically (Mabs)
- Fluorochromes
  - ♦ fluorescein
  - ♦ rhodamine
- Biotinylation
- Enzyme Crosslinking
  - ♦ alkaline phosphatase
  - ♦ horseradish peroxidase

The most common detectable markers used to label antibodies are: radioactivity fluorochromes, biotinylation and enzyme cross-linking. In all of these labeling procedures, it is important that the modification does not affect the antibody's ability to recognize the antigen. Iodination is the preferred method for radiolabeling antibodies. Many iodination methods are gentle and do not adversely affect antibody structure. In addition, mAbs can be metabolically labeled by incubating the hybridomas with radioactive amino acids and harvesting the mAb. Arginine is a good amino acid for the labeling of mAb.

Fluorescent antibodies can be made with commercially available isothiocyanate derivatives of fluorescein and rhodamine. These derivatives react with primary amines (i.e., primarily lysine). The efficiency of labeling can be conveniently determined by spectrophotometry. For example, the A_{495}/A_{280} ratio should be between 0.3-1.0 following conjugation with fluorescein and the A_{575}/A_{280} ratio should be 0.3-0.7 following conjugation with rhodamine. Biotinylation with a succimide ester of biotin also labels primary amines. The labeling of free amines is usually mild and does not affect antibody function. However, the labeling should be kept to a minimum.

Enzymes can also be crosslinked to antibodies. The three most common enzymes are horse radish peroxidase, alkaline phosphatase and β-galactosidase. Methods of cross-linking include glutaraldehyde, periodate treatment and bifunctional cross-linking agents.



Biotin-Avidin Detection Systems. Biotin-avidin based detection systems are also available. Biotin-conjugated secondary antibody or protein A is used to bind the primary antibody. This is followed by an incubation with avidin (or streptavidin) conjugated with detectable marker (i.e., fluorescein, enzyme, ^{125}I). One advantage of this detection system is that the same secondary antibody can be used in different immunoassays, thus reducing the need to purchase many different secondary antibodies. However, one still needs the appropriately labeled avidin for all of the different assays. Avidin has a high affinity for biotin and several biotin molecules can be incorporated into a single antibody molecule; thus increasing the sensitivity of the assay. However, the use of the biotin-avidin systems requires an additional incubation step and the associated wash steps.

OTHER METHODS

Some older immunological methods that are still occasionally use are Ouchterlony, immunoelectrophoresis and crossed immunoelectrophoresis. All of these methods depend on the formation of large supramolecular complexes of antigen and antibody (as discussed in the section on immunoprecipitation). In all of these methods the large antigen-antibody complexes will precipitate within a gel. These methods are insensitive and depend on high antigen and antibody concentrations, as well as a polyclonal sera. In Ouchterlony the antigen and antibody are placed in separate wells and diffuse towards each other. Antigen and antibody complexes will form within the gel at a position between the two wells and precipitate. Immunoelectrophoresis is a similar method, except that the antigen sample is first subjected to electrophoresis to separate proteins and then the antibody is allowed to diffuse into the gel. In crossed immunoelectrophoresis the sample is electrophoresed in a second dimensional gel containing the antibody. In both of the electrophorectic methods, the precipitates will form in association with the specific proteins recognized by the antisera.

Comparison of Immunoassays

TECHNIQUE	GENERAL PROCEDURE	TYPICAL APPLICATIONS
ELISA	• adsorb Ag to solid support • incubate with Ab • detect bound Ab	• quantify Ab • quantify Ag • process large # of samples
BLOTTING	• SDS-PAGE and transfer • incubate with Ab • detect bound Ab	• identify subunit MW • quantify Ag • quantify Ab?
IFA	• fix cells on slide • incubate with Ab • detect bound Ab	• subcellular localization • quantify Ab • quantify Ag?

PART IV

Nucleic Acids and Recombinant DNA

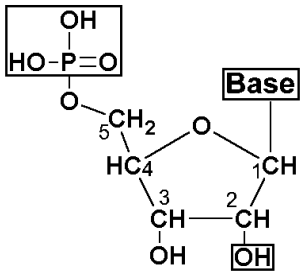
Topics covered:

- Nucleic Acid Structure and Isolation
- Electrophoresis and Blotting
- Polymerase Chain Reaction (PCR)
- DNA Sequencing and Bioinfomatics
- Recombinant DNA

CHAPTER 16--NUCLEIC ACID STRUCTURE AND ISOLATION

Nucleic acids encode information relating to cell structure and function. Cells have the ability to make exact copies of their DNA and pass this information to daughter cells. DNA also serves as a template for the synthesis of RNA, or **transcription**. The different types of RNA are known as ribosomal RNA (rRNA), transfer RNA (tRNA), and messenger RNA (mRNA). The mRNA is processed and **translated** into proteins. rRNA and tRNA are necessary components for protein translation. Manipulating and sequencing nucleic acids are often an easier approach to studying protein function and structure than isolating and characterizing proteins.

Nucleic acids are polymers of nucleotides. Nucleotides are composed of ribose (a 5-carbon sugar) and either a purine or a pyrimidine base at the 1'-position (Figure). The purine bases are adenine (A) and guanine (G) and the pyrimidine bases are cytosine (C), thymine (T) and uracil (U). Uracil is only found in RNA and thymine is only found in DNA. The 5'-hydroxyl of the ribose group can be phosphorylated. Unphosphorylated forms are called nucleosides and phosphorylated forms are called nucleotides (Table). Nucleotides exist as monophosphates, diphosphates or triphosphates and are designated by the nucleoside name and the number of phosphates (eg., adenosine monophosphate, or AMP; adenosine diphosphate, or ADP; etc). Nucleotides that make up DNA lack the 2' hydroxyl (hence the name deoxy-ribonucleic acid) and are indicated with a lower case 'd'.



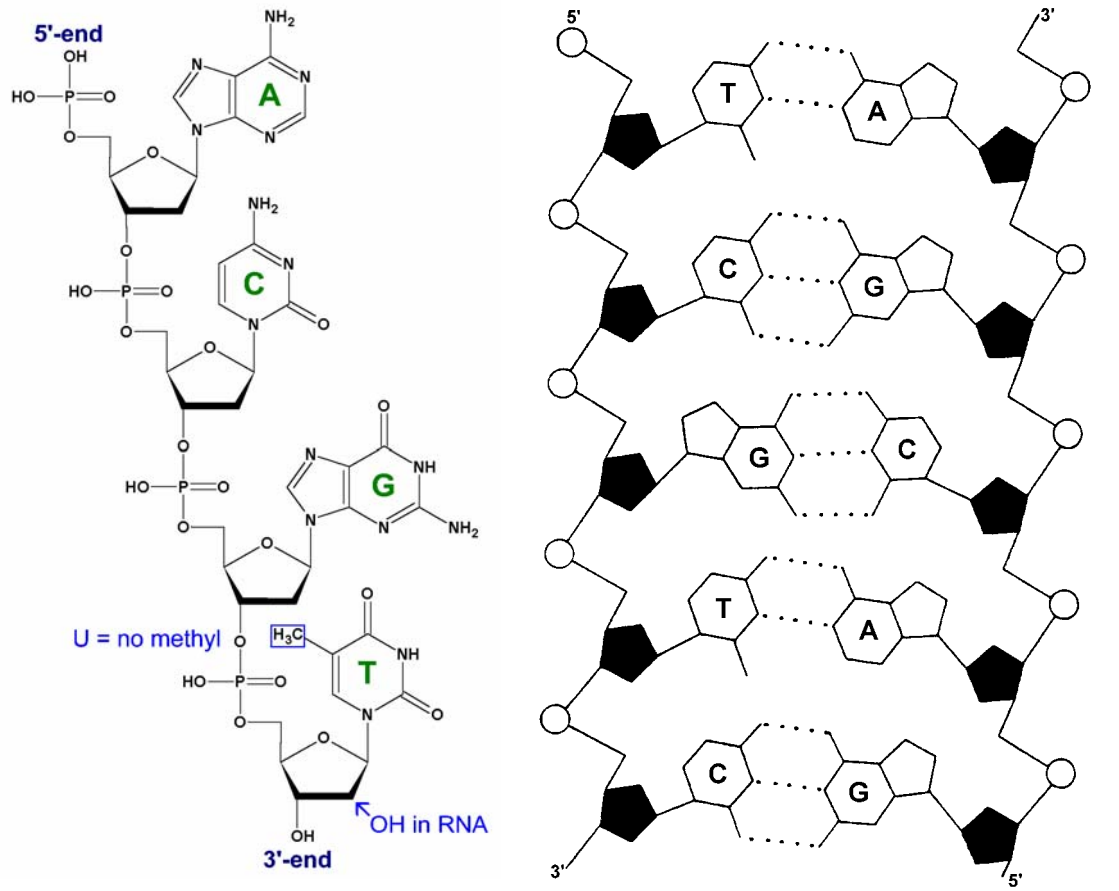
Nucleotide Nomenclature		
Base	Nucleoside	Nucleotide
Adenine	Adenosine	AMP
Guanine	Guanosine	GMP
Cytosine	Cytidine	CMP
Thymine	Thymidine	dTMP
Uracil	Uridine	UMP

Oligonucleotides are formed via a phosphodiester bond joining the 5'-phosphate from one nucleotide with the 3'-hydroxyl on another nucleotide (Figure). These phosphodiester bonds form the 'phosphate backbone' of the oligonucleotide. In addition, oligonucleotides have a polarity with a phosphate group at the 5'-end and a hydroxyl group at the 3'-end. During DNA and RNA synthesis nucleotides are added to the 3'-hydroxyl by polymerases using nucleotide triphosphates (NTPs) as the substrates. The nucleotide is cleaved between the first and second phosphate resulting in the α-phosphate being incorporated into the phosphate backbone.

DNA can also form a double-stranded molecule and the two strands are held together by hydrogen bonds formed between A and T residues and between C and G residues (Figure). The two strands of DNA are oriented in opposite directions in terms of the 5'-to-3' polarity. The restricted base pairing between nucleotides and the opposite polarities of the strands result in the two strands being **complementary**. During DNA replication and RNA transcription one strand will serve as template for the synthesis of the complementary strand. The strands of DNA also twist around each other to form a helix. RNA also forms secondary structures that result from base-pairing between complementary nucleotides within a same strand. Similarly, small

circular DNA molecules can also form higher ordered structures known as super coils. DNA free of proteins, or naked DNA, probably does not exists in cells. Instead, DNA is associated with various DNA-binding proteins which package the DNA into chromosomes.

Nucleic Acid Structure



ISOLATION OF NUCLEIC ACIDS

- Genomic (chromosomal)
- Organellar (satellite)
- Phage/Viral (ds or ss)
- Plasmid (extrachromosomal)
- Complementary (mRNA)

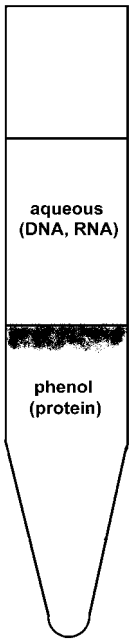
Three major types of techniques, or combinations of them, are employed in the isolation of nucleic acids: differential solubility, absorption methods, or density gradient centrifugation. The choice of method will depend on the type of DNA being isolated (Box) and the application. A major goal of nucleic acid isolation is the removal of proteins. The separation of nucleic acids from proteins is generally easily accomplished due to their different chemical properties. In particular, the highly charged phosphate backbone makes the nucleic acids rather hydrophilic as compared to proteins which are more hydrophobic. Separating the different types of nucleic acids can be more problematic in that they all have similar chemistries. On the other hand, though, this similar chemistry results in a few basic procedures which are common to many nucleic acid isolation protocols. Most nucleic acid isolation protocols involve a cell lysis step, enzymatic treatments, differential solubility (eg., phenol extraction or absorption to a solid support), and precipitation.

- Cell Lysis
- ± Enzyme Treatment
- Phenol Extraction or Adsorption Methods
- EtOH Precipitation

Cell Lysis. Nucleic acids must be solubilized from cells or other biological material. This solubilization is usually carried out under denaturing conditions such as: SDS, alkali, boiling or chaotropic agents. These denaturing conditions efficiently solubilize the nucleic acids and generally do not adversely affect them. In addition, the denaturing conditions promote the removal of proteins during the subsequent steps and inhibit the activity of nucleases which will degrade the nucleic acids.

Enzymatic Treatment. Another approach in the isolation of nucleic acids is to degrade unwanted components. For example, inclusion of proteases (eg., proteinase K) in the lysate will promote the removal of proteins. Proteinase K is still active at 55^{o} in the presence of 0.5% SDS. The elevated temperature and SDS improve solubility and inhibit any DNAse activity that may be present in the lysate. Nucleases can also be used to remove unwanted nucleic acids. For example, many DNA extraction protocols include a RNase treatment step, and visa versa. It is important that the RNase be free of DNase activity. DNase-free RNase is easily prepared by boiling commercial RNase for 10 minutes. The stability of RNase makes the preparation of RNase-free DNase more difficult. RNase-free DNase should be purchased from a reliable vendor or tested before it use.

Phenol Extraction. Phenol is an organic solvent that is used to separate proteins from nucleic acids. Proteins are hydrophobic and partition in the organic phase. Nucleic acids are highly charged and partition in the aqueous phase. The advantages are of phenol extraction are that it is easy to carry out and can be adapted to many applications. It is also easily applied over a wide range of volumes (40 µl to several ml). In particular, phenol extraction is widely used for the isolation of high molecular weight genomic DNA (see below).



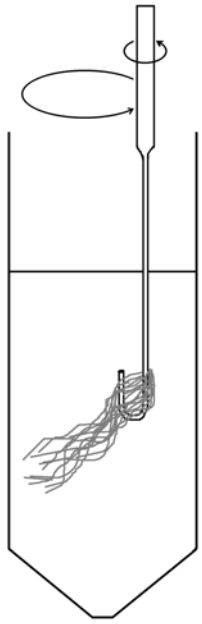
Phenol extraction is accomplished by mixing the sample with an equal volume of phenol which has been previously saturated with a Tris buffer at pH 8 containing EDTA and NaCl. The phenol should be molecular biology grade phenol should and store at -20^{o} until preparing the saturated solution. The saturated solution is stored at 4^{o}. Phenol is easily oxidized, as evidenced by yellowing, and the oxidation products can break DNA. Oxidized phenol should be discarded.

Depending on the application, the two phases are completely mixed by vortexing, or gently mixed (eg., high molecular weight DNA). The phases are separated by centrifugation and the upper aqueous phase, which contains the nucleic acids, is retained. Proteins will often be visible as flocculent material at the top of the phenol phase. The two phases need to be carefully separated in that the nucleic acids and proteins tend to be at the interface. Leaving too much of the aqueous layer behind will lead to undue loss of material and aspirating too close to the interface can include protein. The aqueous phase can be re-extracted with phenol to remove more protein. Phenol is a hazardous waste material that needs to be disposed of properly.

A common variation of phenol extraction is a mixture of phenol:chloroform: isoamyl alcohol (25:24:1). The more organic chloroform removes lipids, denatures more protein and mixes less with the aqueous phase leading to more efficient extraction.

Ethanol Precipitation. Nucleic acids can be precipitated from dilute solutions with ethanol. This precipitation can be a concentration step or a means to change buffers, especially after phenol extraction. Typically either sodium acetate or potassium acetate, pH 5.0-5.5, is added to a final concentration of approximately 0.3 M. The sodium and acidic pH will neutralize the highly charged phosphate backbone and to promote hydrophobic interactions. Two-to-two and a half volumes of ethanol are added and the sample is incubated as -20^{o}. If the nucleic acids are small in size and/or in low concentrations an extended incubation (several hours to overnight) is needed. The precipitated DNA is collected by centrifugation. The pellet is rinsed with 70% ethanol to remove any excess salt, dried and dissolved in the appropriate buffer. A variation is to substitute ammonium acetate if the 'hard' salts are a problem. Another modification is to use an equal volume of isopropanol (instead of 2-2½ volumes of ethanol) which minimizes the increase in sample volume.

Isolation of High Molecular Weight Genomic DNA. High molecular weight chromosomal DNA is usually isolated by multiple rounds of phenol extraction and enzyme treatments as discussed above. Shear forces, which can break long DNA molecules, need to be avoided during all steps and samples should never be vortexed. Therefore, the phenol extraction is carried with gentle rocking for several hours. These precautions against shear forces are not necessary in the isolation of low molecular weight DNA. Another common modification at the ethanol precipitation step is ‘spool out’ the high molecular weight genomic DNA on the end of a sealed Pastuer pipet. The precipitated DNA is wrapped around the end of the pipet is then allowed to partially dry and then dissolved in the appropriate buffer. This minimizes the contamination with RNA and low molecular weight DNA fragments.



Plasmid Minipreps and Adsorption Methods. Historically, phenol extractions were used for the isolation of most forms of nucleic acids. It is now more common to use techniques based upon adsorption chromatography for the isolation of smaller DNA molecules, such as plasmids. Various kits are available for the rapid isolation of small quantities of plasmid DNA. The procedure consists of solubilizing the bacteria in an alkali solution followed by neutralization with sodium acetate. The neutralization results in the precipitation of some of the protein and the genomic DNA which is removed by centrifugation. The soluble material is then mixed with a resin in the presence of chaotropic agents (usually guanidine hydrochloride). The resins are usually either based on silica or diatomaceous earth. Under these conditions DNA binds to the matrix, but proteins and RNA do not. The DNA is eluted in a low salt buffer. These methods are rapid and yield a highly purified plasmid DNA which can generally be used directly in most applications without further processing.

Another common application for an adsorption method is the isolation of DNA fragments following gel electrophoresis. In this case the agarose gel piece containing the DNA is dissolved in NaI, a chaotropic salt, and the DNA adsorbed to silica. The DNA is then eluted with a low salt buffer and sometimes gentle heating.

ISOLATION OF RNA

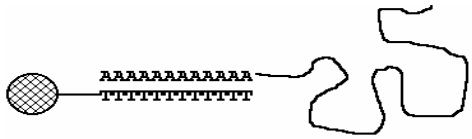
- RNase inhibitors!
- extraction in guanidine salts
- phenol extractions at pH 5-6
- treatment with RNase-free DNase
- precipitation with LiCl
- oligo-dT column

Most RNA isolation protocols also involve phenol extractions and are similar to DNA isolations. However, there are some differences and special considerations (Box). In particular, precautions against RNase activity must be taken. RNase is an extremely stable and active enzyme. Gloves should be worn at all times and sterile plasticware should be used whenever possible to avoid introducing exogenous RNase to the sample. Glassware needs to treated with DEPC-water and autoclaved to inactivate any RNase. Buffers should be prepared from DEPC-water or RNase inhibitors included.

The cell lysis and solubilization of RNA will typically be carried out in guanidine salts (especially guanidine thiocyanate). Guanidine is a strong chaotropic agent and will inhibit RNases. The strong denaturing effect of this salt will also promote better phenol extraction. The phenol extraction is the same as the DNA isolation except that the phenol is usually saturated with a buffer of pH 5-6. The lower pH will result in some DNA partitioning in the organic phase. If DNA contamination is a problem, it is possible to purchase RNase-free DNase. Long-term storage is best as a precipitate in 70% ethanol at -20^{o}.

Precipitation of RNA with LiCl. LiCl has been used to selectively precipitate RNA. Large RNAs (rRNA, mRNA) are insoluble at high ionic strength, whereas small RNAs (tRNA and 5S rRNA) and DNA generally remain soluble. Following either phenol or guanidine extraction, an equal volume of 8 M LiCl is added. The sample is mixed vigorously and incubated at -20^{o}. The precipitate is collected by centrifugation and reprecipitated if necessary.

Affinity chromatography. Most eukaryotic mRNA contains a stretch of A residues at its 3' end which added post-transcriptionally. It is possible to isolate mRNA by affinity chromatography on oligo-dT columns (Figure). The RNA solution is passed over an oligo-dT column under conditions which promote base pairing. Only RNA with a polyA tail binds. The polyA RNA is eluted under conditions (usually low salt and high temperature) which breaks the base pairing.



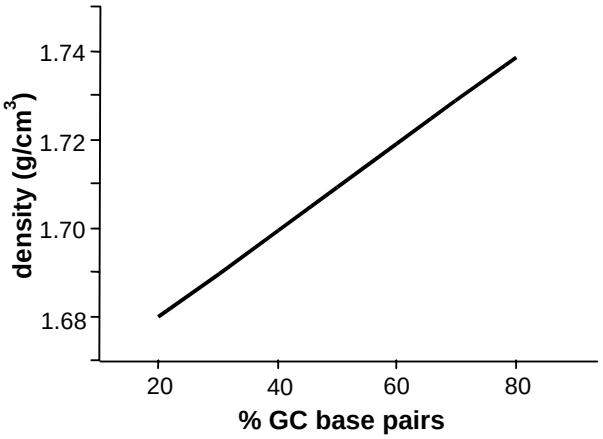
Oligo-dT Column

DENSITY GRADIENT CENTRIFUGATION

Density in CsCl	
DNA	~ 1.7 g/cm^3
Protein	~ 1.3 g/cm^3
RNA	> DNA
ssDNA	> dsDNA

Density gradient centrifugation can also be used in the analysis and isolation of nucleic acids. Double-stranded DNA, single-stranded DNA, RNA and proteins all have different densities (Box) and therefore can be separated by isopycnic (i.e., equilibrium) centrifugation. CsCl is the standard medium for the density gradient centrifugation of nucleic acids and are especially useful for the purification of large amounts of highly purified DNA. The gradients are carried out in the presence of ethidium bromide which fluoresces when bound to DNA. The DNA bands are detected by illumination with ultraviolet light and easily recovered with a syringe and needle by puncturing the wall of the disposable tube and aspirating the DNA. The CsCl can be removed by dialysis or by precipitating the DNA.

The %G:C content affects the density of DNA (Figure). This can result in multiple bands on CsCl gradients if DNA composed different G:C compositions are centrifuged. For example, minor bands, called satellite DNA, are often observed when total DNA from an organism is analyzed by CsCl gradients. These satellite bands are usually due to highly repetitive DNA or organellar DNA. Mitochondrial DNA is noted for its high A:T content. These satellite DNA bands can be purified from the genomic DNA by density gradient centrifugation.



Nucleic acids can also be separated according to size by rate zonal centrifugation on sucrose gradients. However, this method is not widely used since gel electrophoresis is generally a more convenient method for the size fractionation of nucleic acids.

ANALYSIS AND QUANTIFICATION

DNA	A_{260}	1.0 ≈ 50 μg/ml
	A_{260}/A_{280}	1.6 - 1.8
RNA	A_{260}	1.0 ≈ 40 μg/ml
	A_{260}/A_{280}	~2.0

The quality and quantity of isolated nucleic acids can be determined spectrophotometrically (Box). Nucleic acids have an A_{max} of 260 nm and proteins have A_{max} of 280 nm. The A_{260}/A_{280} ratio is therefore indicative of the degree of purity of the nucleic acid. A_{260}/A_{280} ratios of 1.6-1.8 or 1.8-2.0 are usually acceptable for DNA and RNA, respectively. The standard extinction coefficient used for ssDNA, dsDNA and RNA are 0.03 ml/μg, 0.02 ml/μg and 0.025 ml/μg, respectively. Formulas which take into account protein and other contaminants are also available. Indirect spectrophometric assays for DNA quantification are also available, but rarely used. In some instances fluorometry using fluorescent dyes that bind DNA and/or RNA is used to determine nucleic acid concentrations.

APPENDIX 1. COMMON CONVERSIONS

Weight Conversions

1 µg = 10^{-6} g
1 ng = 10^{-9} g
1 pg = 10^{-12} g
1 fg = 10^{-15} g

Spectrophotometric Conversions

1 A_{260} unit of double-stranded DNA = 50 µg/ml
1 A_{260} unit of single-stranded DNA = 33 µg/ml
1 A_{260} unit of single-stranded RNA = 40 µg/ml

DNA Molar Conversions

1 µg of 1,000 bp DNA = 1.52 pmole (3.03 pmoles of ends)
1 pmole of 1,000 bp DNA = 0.66 µg

Protein Molar Conversions

100 pmoles of 100,000 dalton protein = 10 µg
100 pmoles of 50,000 dalton protein = 5 µg
100 pmoles of 10,000 dalton protein = 1 µg

Protein/DNA Conversions

1 kb of DNA = 333 amino acids of coding capacity = 3.7 x 10^4 dalton protein
10,000 dalton protein = 270 bp DNA
50,000 dalton protein = 1.35 kb DNA
100,000 dalton protein = 2.7 kb DNA

Formulas

Picomole Ends per Microgram of Double-Stranded Linear DNA:
$(2x10^6)/(660x\text{Number of Bases})$=pmole ends/ µg

Exact Molecular Weight of an Oligonucleotide:
[(Ax312.2)+(Gx328.2)+(Cx288.2)+(Tx303.2)-61.0] = Molecular Weight of specific oligonucleotide

CHAPTER 17--MODIFICATION OF NUCLEIC ACIDS

All DNA molecules are similar in terms of their biochemical and physical properties. Therefore, unlike proteins, DNA techniques are not highly dependent upon the particular gene being studied. However, genomic DNA is an extremely large molecule. For example, the human genome contains approximately 2 x 10^9 base pairs (bp). The size of a gene for a 50 kDa protein might be as small as 2000 bp (or 2 kb). Therefore, it can be quite difficult to identify and characterize specific genes from an organism.

The study of specific genes involves manipulating nucleic acids. In particular, it is possible to break DNA into smaller fragments, identify fragments of interest and to amplify these fragments so that they can be analyzed. Enzymes are used to carry out these manipulations of DNA and RNA. DNA modifying enzymes include: polymerases, ligases and nucleases. **Polymerases** synthesize nucleic acids in a template mediated fashion. **Ligases** join fragments of DNA. Nucleases cleave the phosphodiester bond between nucleotides. A wide range of specificities are exhibited by nucleases (Box). **Exonucleases** remove nucleotides one at a time from either the 5' or 3' end. In addition, some exonucleases exhibit substrate specificities in terms of preferences for double-stranded (ds) or single-stranded (ss) DNA or RNA. **Endonucleases** cleave the phosphodiester bond in the middle of a oligonucleotide and produce fragments. Such exo- and endonucleases do have some applications primarily directed at removing unwanted types of nucleic acids or removing single stranded overhangs from dsDNA.

Examples of Nucleases:

Bal 31	3'-exonuclease of ds or ssDNA
S1	ssDNA or ssRNA 5'-exonuclease
Mung Bean	ssDNA 5'-exonuclease
DNase I	ss or dsDNA (5' of pyr)
ExoIII	3'-exonuclease of dsDNA
RNase A	ssRNA endonuclease (3' of pyr)
RNase T1	ssRNA endonuclease (3' of G)

RESTRICTION ENDONUCLEASES

In the late 1960's endonucleases that cleave DNA at specific sites were described. These are called **restriction endonucleases**, or restriction enzymes. Restriction enzymes are found in prokaryotes and they function to protect bacteria from phage (viruses of bacteria) infection. For every restriction enzyme, there is a corresponding site-specific DNA modifying enzyme (usually a methylase). Bacteria modify their own DNA by site-specific methylation which will make the DNA resistant to digestion with that particular restriction endonuclease. Foreign DNA from a bacteriophage will not be methylated at the appropriate restriction sites and thus be degraded by the restriction endonuclease within the host bacteria. The name restriction enzyme refers to restriction of some phages to particular bacteria strains.

Restriction Enzymes

Type I	cleavage occurs 400-7000 bp from recognition site
Type II	cleavage occurs adjacent or within recognition site
Type III	cleavage occurs 25-27 bp from recognition site

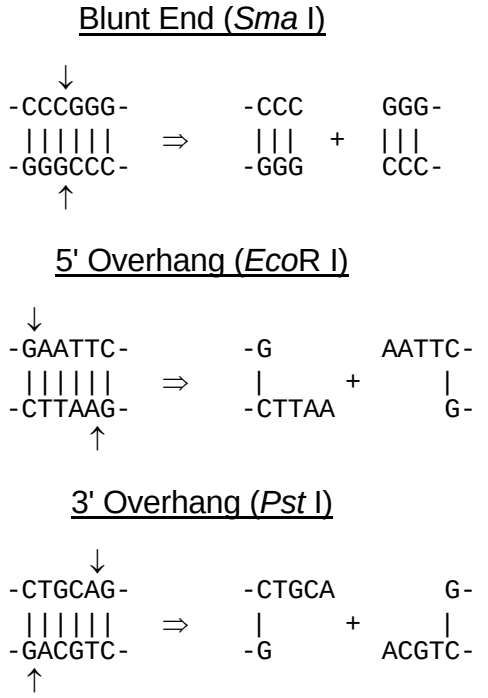
Three distinct classes of restriction endonucleases, designated types I, II and III, have been described (Box). Class II restriction enzymes are the most useful since they cleave DNA at the recognition site. Restriction enzymes are named according to the bacteria species from which they are isolated. The first letter is the first letter of the genus and is

capitalized. The next two letters are the first two letter of the species and are in lower case. Following these three letters (which are italicized) is an optional indicator for the strain of bacteria. The last position is a Roman numeral which designates the number of restriction enzymes discovered in that species and the order of their discovery.

The recognition sequences of type II restriction enzymes are generally 4-8 consecutive nucleotides. Some restriction enzymes do allow for some degeneracy at a particular nucleotide. This degeneracy can be restricted to a purines (R) or pyrimidines (Y), or be completely unrestricted (N). Most recognition sites are **palindromes** in that both strands exhibit the same sequence. The sequence complementarity and opposite orientations of the two strands leads to a dyad symmetry. Hydrolysis of the phosphodiester bond results in the phosphate on the 5'-carbon and the hydroxyl on the 3'-carbon. Both DNA strands are cut between the same two residues. This will result in **blunt ends** if the cleavage site is in the exact center of the recognition site (Figure). Alternatively, **5' overhangs** (or extensions) or **3' overhangs** of varying length will be produced if the cleavage site is not in the center of the recognition site.

Examples of Restriction Sites

Enzyme	Site
*Cla*I	A T↓C G A T T A G C↓T A
*Eco*RI	G↓A A T T C C T T A A↓G
*Fnu*AI	G↓A N T C C T N A↓G
*Hae*III	G G↓C C C C↓G G
*Hin*dII	G T Y↓R A C C A R↓Y T G
*Hin*dIII	A↓A G C T T T T C G A↓A
*Pst*I	C T G C A↓G G↓A C G T C



Isoschizomers are restriction enzymes isolated from different sources that recognize the same sequence, but cleave at different positions resulting in different overhangs. In addition, different recognition sites can produce identical overhangs (called **compatible ends**).

Reaction conditions. Restriction enzymes reproducibly cleave DNA at specific sites. This feature provides a convenient way to break large DNA molecules into smaller fragments for subsequent analysis and manipulation. Restriction digests are carried out by incubating the DNA with the restriction enzyme. Each restriction enzyme has its own optimal reaction conditions (pH, ionic strength, temperature, etc.). Manufacturers usually supply an optimal 10X stock buffer with the enzyme. Appropriate amounts of DNA, the stock buffer, other additives, the enzyme and water are mixed in a single tube. The DNA needs to be quite

pure in that impurities in the DNA preparation (eg., salts, solvents, contaminating proteins, and RNA) can inhibit the digestion. The incubation time will depend on the DNA concentration and the amount of enzyme.

DNA can also be digested with mixtures of restriction enzymes. If the two enzymes have the similar optimal reaction conditions, they can be mixed together and the reaction carried out simultaneously. If not, the digestions are carried out sequentially by diluting the sample in the appropriate buffer or by adding the required components. If necessary, the DNA can be re-isolated after digestion with the first enzyme. The typically procedure is to phenol extract the digestion mixture, precipitate the aqueous phase with ethanol, and dissolve the precipitated DNA in the appropriate buffer for the second digestion.

Conditions Promoting Star Activity

- high glycerol (>5%) concentration
- high enzyme/DNA ratio (100 units/μg)
- low ionic strength (<25 mM)
- high pH (>8)
- organic solvents
- substitution of Mg^{2+}

Some restriction enzymes cleave DNA with less specificity under certain reaction conditions (Box). This is not the same as the degeneracy in the recognition site exhibited by some enzymes. For example, at high enzyme or glycerol concentrations *Eco*RI will cleave at sites in which only 5 of the 6 (GAATTC) nucleotides match and this is referred to as **star activity**. Star activity is generally undesirable since it is random and not reproducible. However, in some applications, digestions are carried under conditions which promote star activity to introduce random cleavages.

Frequency of cutting. Theoretically restriction sites will be random within any genome. The number of times a particular restriction enzyme will cut a genomic DNA can be approximated from the size of the genome and its base composition. For example the probability of finding any 6-base recognition sequence in a genome with 50% GC content is:

(¼)(¼)(¼)(¼)(¼)(¼) = 1/4096

This means that one would expect on average one restriction site every 4000 bp. These restriction sites will not be evenly spaced every 4000 base pairs, but will occur randomly. For example, the genome of *E. coli* (4×10^6 bp) will be digested into approximately 1000 fragments of various sizes. The non-randomness of certain di- and trinucleotides can influence the actual cutting frequencies.

CHAPTER 18--ELECTROPHORESIS OF NUCLEIC ACIDS

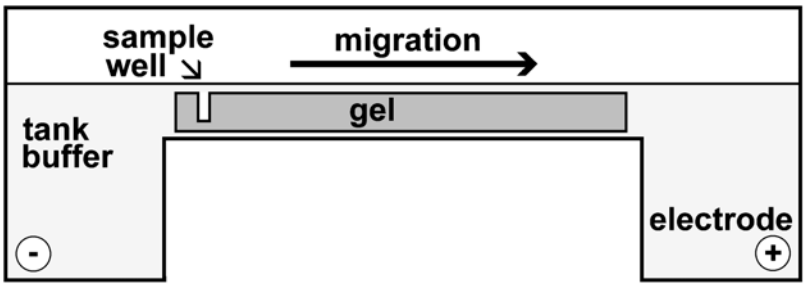
Gel electrophoresis is commonly used to separate and analyze the nucleic acids. The negatively charged phosphate backbone of nucleic acids gives them a uniform charge-to-mass ratio. Therefore migration in gels is inversely proportional to the size of the nucleic acid (as discussed for SDS gel electrophoresis of proteins). The structure of the nucleic acid (eg., linear, circular, double-stranded, or single-stranded) will also affect migration in gels. It is possible to distinguish DNA fragments that differ by as little as 1% and in some applications (eg., DNA sequencing) fragments that differ in size by a single nucleotide can be resolved.

A wide range of sizes (<10 bp to >20 kb) can be analyzed by gel electrophoresis. Agarose is the most common matrix for the electrophoresis of nucleic acids. Nucleic acids ranging in size from 100 bp to 20 kb can be resolved depending on the agarose concentration (Table). Acrylamide is used for smaller nucleic acids and oligonucleotides.

% agarose	range (kb)	% acrylamide	range (bp)
0.7	0.8-20	3.5	100-1000
0.9	0.5-7	5.0	80-500
1.2	0.4-6	8.0	60-400
1.5	0.2-4	12.0	40-200
2.0	0.1-3	20.0	10-100

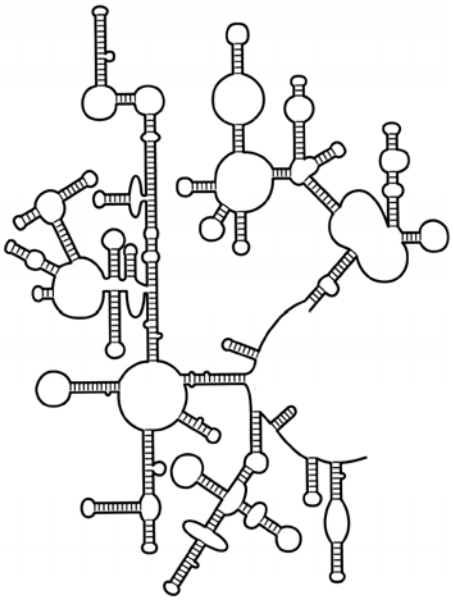
NuSieve® 3:1 agarose (FMC) is specially designed to separate small DNA fragments (30-2000 bp).

Acrylamide gel electrophoresis can be carried out in the same apparatus used for protein electrophoresis (i.e., vertical). Agarose gels are usually poured and electrophoresed in a horizontal apparatus. The horizontal apparatus provides more support for the gel and are easier to prepare. The appropriate amount of agarose is dissolve in the electrophoresis buffer by heating the solution, which is then poured into a mold and allowed to cool. A well-forming comb is inserted before the agarose polymerizes. The gel is submerged in the electrophoresis tank buffer, the samples are loaded, and the appropriate voltage applied. Tracking dyes are used to monitor the electrophoresis.



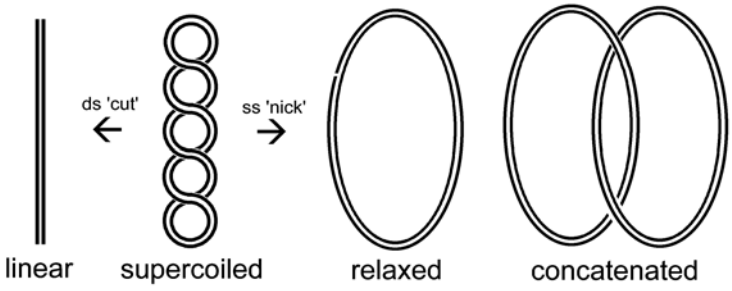
Staining with ethidium bromide (or other fluorescent dyes) is the most common method for detecting nucleic acids after electrophoresis. The ethidium bromide can by co-polymerized with the agarose gel. As the nucleic acids migrate through the gel the ethidium binds and becomes fluorescent. Illumination with UV light will reveal the fluorescent bands. The progress can be monitored during electrophoresis until the desire resolution is achieved. Alternatively the gel can be stained post-electrophoresis by incubating in an ethidium bromide solution and destaining with water. As little as 5-10 ng of DNA can be detected. Specific nucleic acids are detected by transferring the nucleic acids to a membrane and hybridizing with a DNA probe

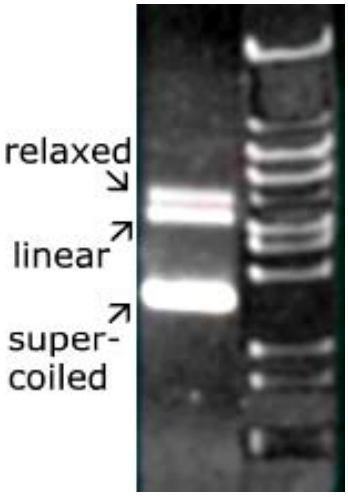
(discussed below).



Shape also effects the migration of nucleic acids in gels. For example, RNA and ssDNA can form secondary structures due to internal base pairing. These secondary structures will lead to a lower mobility. Secondary structures are eliminated by running the gels under denaturing conditions which prevent the formation of base pairs. Chaotropic agents, such as urea, formamide, or formaldehyde, are included in the gel buffers for the electrophoresis of ssDNA and RNA if sizes need to be accurately determined.

DNA can also exist as a circular molecule (eg., plasmids). Furthermore, the circular DNA can exist in multiple conformations. In its natural state, circular DNA is supercoiled due to additional superhelical twists. Breaking a single bond in the phosphate backbone of one strand will result in a relaxed circular DNA molecule. Breaking both strands in the same position (eg., endonuclease) will result in the formation of a linear molecule. Supercoiled circular DNA, relaxed circular DNA and linear DNA of the same molecular weight will migrate at different rates through the gel. The typical order of migration from fastest to slowest is supercoiled DNA, linear DNA, and relaxed circular DNA. However, the exact migration is influenced by the agarose concentration, the applied current, ionic strength and number of superhelical twists. Furthermore, circular DNA can form concatenated forms (dimers, trimers, etc.). Therefore, multiple bands are often observed in highly purified plasmid preparations. Complete digestion with a restriction endonuclease which only cuts once within the plasmid will result in the formation of a single band corresponding to the linear DNA molecule.





Recovery of DNA from gels. DNA fragments are relatively easy to recover from gels (Box). One possibility is to transfer the nucleic acids to membranes and analyze the DNA or RNA by blotting techniques (discussed in Chapter 19). DNA can also be recovered by electroelution. The regions containing the DNA fragment of interest is excised from the gel and put into a dialysis bag with electrophoresis buffer. The dialysis bag is placed into an

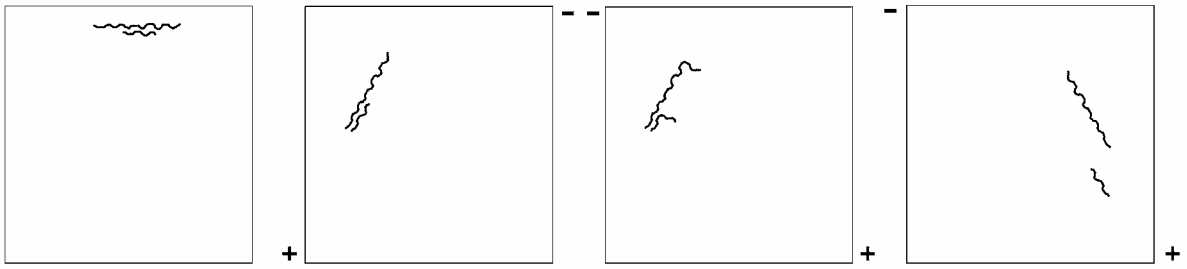
electrophoresis chamber. The DNA fragment will migrate out of the agarose but remain in the dialysis bag. The DNA is then recovered by precipitation.

- transfer to membranes
- electroelution into dialysis bags
- NaI/glassmilk (eg., GeneClean®)
- low-melting temperature agarose

Another method involves dissolving the excised agarose gel piece with NaI. A silica resin which binds DNA is then added to the dissolve gel solution. Following several washes, the DNA is eluted from the resin at low ionic strength (see section on purification by absorption methods). Low-melting temperature (40-50^{o}) agarose is also available. The DNA fragment is excised and the gel is melted. Temperatures below 60^{o} will rarely have an adverse affect on the DNA. The fragment is then recovered from the solution by ethanol precipitation.

PULSE FIELD GEL ELECTROPHORESIS

Conventional gel electrophoresis has an upper limit to sizes that can be resolved. DNA fragments larger than 20 kb tend to "snake" through the gel at the same rate. This phenomenon is known as reptation. Pulsed field gel electrophoresis (PFGE) overcomes the problems of resolution of large DNA fragments by using two sets of electrodes that are at fixed angles to each other (see http://www.nalusda.gov/pgdic/Probe/v2n3/puls.html for review). The current is alternated between these sets of electrodes at defined intervals. Separation is based upon the ability of the smaller DNA molecules to reorient faster than larger DNA molecules. DNA molecules up to 10 Mb can be resolved by PFGE.

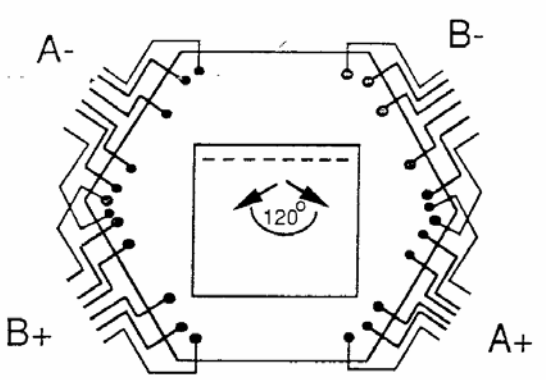


Many lower eukaryotes have chromosomes in 0.5-3 Mb size range. PFGE provides a means to characterize and map genes to the chromosomes of these organisms. In higher eukaryotes PFGE is used for long-range restriction mapping with restriction enzymes that cut infrequently.

After the initial description of PFGE in 1984, several variations of the method were developed (Box). All are based on changing the direction of migration and the differences between the methods are primarily in the positions of the electrodes. The contoured-clamp homogeneous electric field, or CHEF, is probably the more popular method. The electrodes in the CHEF apparatus are arranged in a square or hexagonal contour (Figure) leading to the generation of a homogeneous electric field. This results in the lanes being straight and allows for an easy comparison of different samples. In many of the other variations, the lanes run at an angle and are difficult to compare.

Variations in PFGE

- Orthagonal-Field Alternation Gel Electrophoresis
- Field-Inversion Gel Electrophoresis
- Transverse Alternating Gel Electrophoresis
- Programmable Autonomously Controlled Electrode
- **Contoured-Clamped Homogeneous Electric Field**



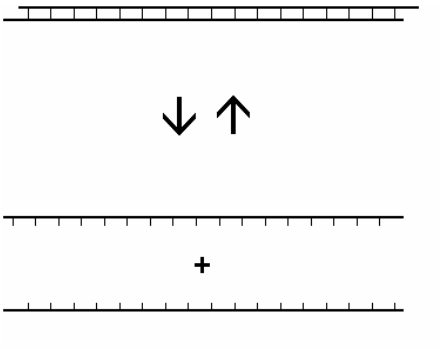
CHEF Electrodes

Factors Affecting Resolution

- uniformity of the two electric fields
- lengths of the electric pulses
- the ratio of the lengths of the pulses
- the angles of the two electric fields
- strengths of the electric fields

Large DNA fragments are easily broken by shear forces. To avoid breakage of chromosome-sized DNA fragments, whole cells are embedded into agarose blocks. The cells are lysed in the agarose block and subjected to electrophoresis. The restriction digests can also carried out in the agarose blocks before electrophoresis. Several electrophoretic variables can be manipulated and these will affect the resolution. The electrophoresis conditions will depend upon the particular apparatus being used and the size range of DNA fragments to by separated.

CHAPTER 19--HYBRIDIZATION AND BLOTTING TECHNIQUES



DNA can be separated into single strands by disrupting the H-bonds which hold the complementary strands together. Complementary DNA strands can reform the double-stranded molecule in vitro. This ability to reform dsDNA molecules, or hybridization, allows for the detection of specific DNA fragments or RNA molecules. Hybridization can be carried out in solution or after target DNA has been transferred to a membrane support. Nucleic acids immobilized on a membrane support can be identified with a labeled DNA probe by blotting techniques.

In 1975 Dr. Southern described a technique to detect specific DNA fragments after gel electrophoresis and the technique became known as Southern blotting (Box). Others modified the technique to detect specific RNA fragments and this method became known as Northern blotting. [When protein blotting was developed it became known as Western blotting.] Nucleic acids can also be bound to membranes without prior electrophoresis and this process is referred to as dot blots.

Southern Blot	≡	DNA on Gel
Northern Blot	≡	RNA on Gel
Dot Blot	≡	No Gel

GENERAL PROCEDURES

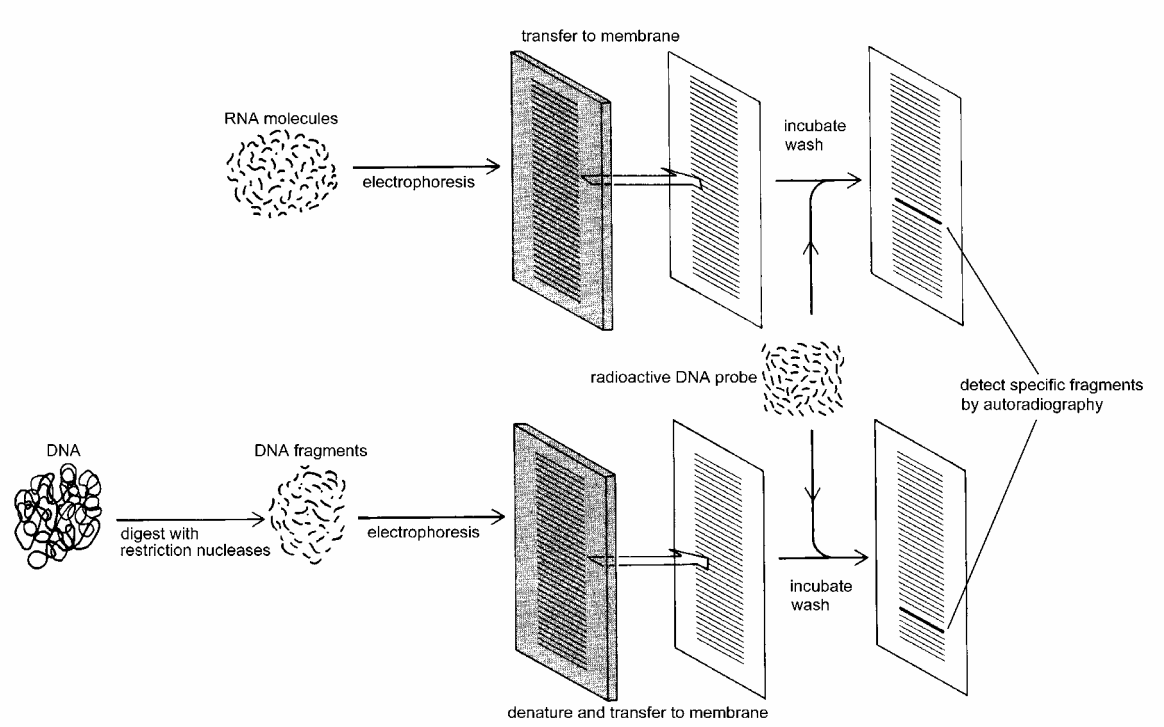
The first steps of nucleic acid blotting are to isolate the nucleic acid of interest and prepare it for analysis. For example, DNA is usually digested with restriction enzyme(s) before electrophoresis. The sample is then subjected to gel electrophoresis on agarose gels. In the case of RNA or ssDNA the gels are usually carried out under denaturing conditions to minimize secondary structures.

Generic Blotting Protocol

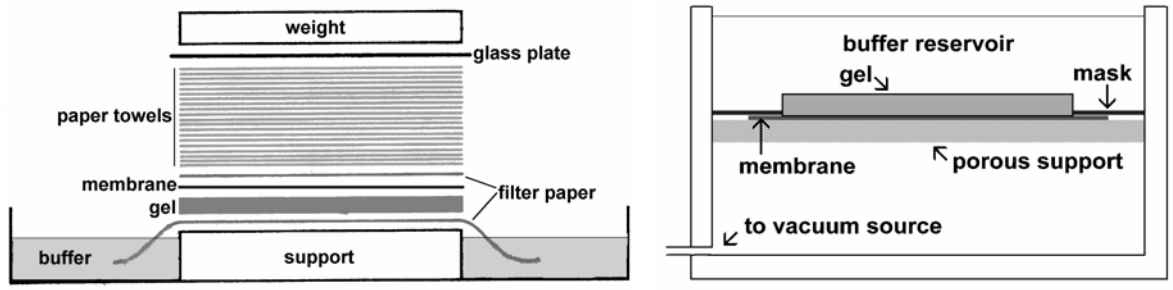
1. Digest DNA or isolate RNA
2. Electrophoresis
3. Depurinate (optional)
4. Denature dsDNA
5. Transfer to membrane
6. Fix nucleic acid
7. Prehybridize
8. Incubate with probe
9. Wash
10. Detect (autoradiography)

An optional depurination step is sometimes carried out following the electrophoresis, especially if very large fragments of DNA are being analyzed. This step is a brief incubation of the gel in a dilute HCl solution followed by an incubation in a neutral buffer. The brief acid treatment will randomly remove some purine bases and lead to random breaks in the DNA. The smaller fragments of DNA are more efficiently transferred from the gel to the membrane.

Double-stranded DNA is denatured (i.e., the complementary strands separated) by treating the gel with a NaOH solution. The gel is neutralized by incubating in the appropriate buffer before transferring the denatured DNA to membranes. It is not necessary to carry out the denaturation step if RNA or ssDNA has been electrophoresed under denaturing conditions.



Transfer of nucleic acids to membrane supports. Four methods for transferring nucleic acids to membrane supports have been described. The original method developed by Southern uses capillary action (Figure). In this method buffer being drawn through the gel carries the nucleic acid. The nucleic acid then binds to the membrane. The method is still widely used but takes several hours and typically is carried out overnight. Electrophoretic transfers are also possible but not widely used. Special devices using either pressure or vacuum have been developed. The vacuum blotting apparatus (Figure) is the more popular and can transfer nucleic acids in approximately 30 minutes.



Several types of membranes are used in nucleic acid blotting techniques. Nitrocellulose has been largely replaced by nylon as the membrane of choice. Nylon is a more durable membrane and binds more nucleic acid. Special nylon membranes, which have been modified to contain positive charges, bind even more DNA.

Fixation of the nucleic acid to the membrane improves the sensitivity by preventing loss of target nucleic acid during the subsequent steps. The membranes can either be 'baked' (heated

to 80^{o} for 2 hr) or cross-linked with UV radiation using a special apparatus. Membranes are then 'prehybridized' to block non-specific binding sites. The prehybridization solution contains non-specific DNA, such as herring sperm or calf thymus DNA. Following the prehybridization, the membrane is incubated with the 'probe' and then washed extensively. The bound probe is then detected by autoradiography if radioactive probes were used, or by ELISA type procedures if non-radiocative probes (see below) were used.

It is also possible to strip the blot of the probe by incubating under conditions which do not allow for hybridization (eg., boiling in low ionic strength). The blot can then be reanalyzed with a different probe. This is a convenient method to compare the expression of two different genes by Northern blotting. However, the stripping and rehybridization can only be carried out a few times.

- temperature
- ionic strength
- chaotropic agents
- probe length
- probe mismatch
- % GC

FACTORS AFFECTING HYBRIDIZATION

Several factors affect the binding of a DNA probe to the target nucleic acid (Box). All of these factors affect the number or stability of the H-bonds formed between complementary strands. Heating DNA destabilizes the H-bonds and increasing the temperature will make it more difficult for the probe to bind to the target DNA. The hybridization of DNA fragments is impeded by electrostatic repulsion due to the negative charge of the phosphate backbone. Cations neutralize this charge-charge repulsion and adding salt will promote the interaction between probe and target. The maximum effect is achieved with 1 M Na^{+}. Chaotropic agents affect H-bond stability and therefore will reduce the hybridization of probe to target DNA. The length and the homology of the probe to the target DNA will determine the total number of H-bonds formed between the target and the probe. However, the total probe length tends to only have a major affect when oligonucleotides are used as probes. DNA with a higher percentage of GC hybridizes will form more H-bonds since GC pairs have 3 H-bonds and AT pairs have 2 H-bonds. All of these factors will play a role in the overall stability of the duplex formed between the target and the probe.

The separation of DNA strands is sometimes called melting and exhibits a **melting temperature**, or T_m. The T_m reflects the overall stability of any particular DNA duplex. The effective T_m for any particular duplex will be determined by the summation of these various factors that affect DNA hybridization (Box). For example, the following formula can be used to approximate the T_m under different hybridization conditions:

$$\textbf{Effective } T_m = 81.5^{o} + 16.6\log[Na^{+}] + 0.41(\%GC) - 0.72(\%formamide) - 1.4(\%mismatch)$$

Consistent with the above discussion, increasing the sodium concentration and percentage of GC base pairs will raise the T_m and increasing the formamide (a chaotropic agent) concentration and decreasing the amount of homology between the probe and target will lower the T_m. Often the %mismatch is not known and it not used in the calculation. In the case of synthetic oligonucleotides the absolute probe length and amount of GC base pairs are the major factors involved in hybridization. Thus, the following formula results in a more accurate estimation of

T_m:

Effective T_m = 2(A + T) + 4(G + C)

where A, T, G and C refer to the absolute numbers of each of the nucleotides.

Stringency vs. T_m	
high	T_m - 15^o
moderate	T_m - 25^o
low	T_m - 35^o

Generally, hybridization is discussed in terms of stringency and not the T_m. **Stringency** refers to the conditions of the hybridization. It is a relative term that is related to the T_m (Box) and reflects the homology between the probe and the target. For example, only a probe with a high degree of homology to the target DNA will hybridize under high stringency conditions, whereas low stingency will allow a less homologous probe to hybridize to the target DNA.

The stringency is controlled by changing the hybridization conditions. For example, increasing the temperature, decreasing the salt concentration, or including formamide all increase the stringency. Likewise, the stringency can be decreased by lowering the temperature, increasing the salt concentration, or decreasing the formamide concentration. In practical terms it is often easier to vary the sodium or formamide concentrations rather than the temperature. A common buffer for hybridization is SSC (standard sodium citrate) which is a citrate buffered sodium solution. A stock solution of 20X SSC (= 3.3 M Na^+) is diluted to achieve different levels of stringency. Typically ranges of 0.1-6X SSC are used with the lower SSC concentrations representing higher stringency. The formamide is typically used to lower the temperature and at the same time maintain a certain level of stringency. For example, hybridizations carried out at 60-65^o in the absence of formamide or at 37-42^o with 50% formamide are about equal in stringency. If incubations at high temperatures are inconvenient then formamide can be included in the buffers and the hybridizations carried out at lower temperatures.

Stringency also applies to the wash steps. It is common to hybridize at low stringency and then to wash at a higher stringency. This insures that the probe will bind to the target DNA and any non-specific hybridization can be removed during the wash steps. The conditions for hybridization (i.e., stringency) need to be determined empirically in conjunction with the above formula as guidelines. A single blot can be sequentially examined under different stringencies. Hybridization and washes are initially carried out at low stringency. The blot is wrapped in plastic and not allowed to completely dry. After exposure to x-ray film, the blot is washed under higher stringency and reexposed to X-ray film. Comparison of the different autoradiographs will allow one to determine how homologous the probe is to the target DNA and the degree of cross-hybridization to other DNA fragments.

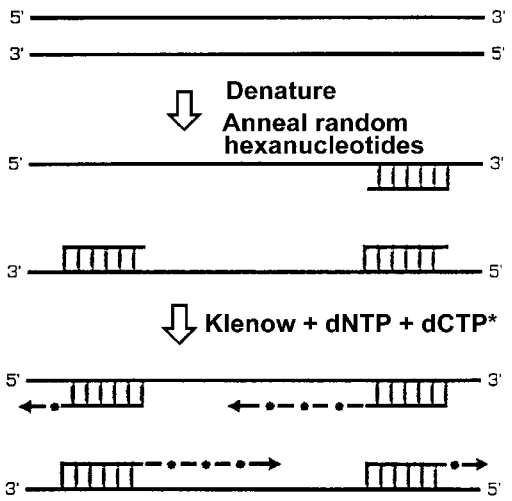
- Nick Translation
- Random Priming
- T4 Nucleotide Kinase
- Terminal Transferase

PREPARATION OF LABELED DNA PROBES

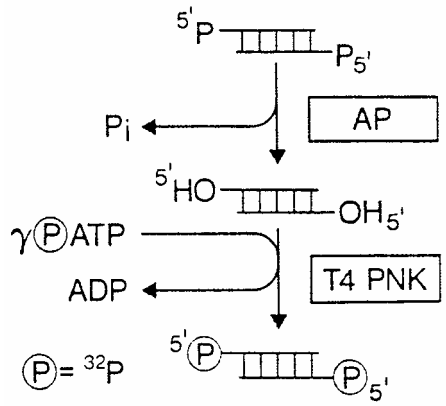
The two major sources of probes are previously cloned genes and synthetic oligonucleotides. In both cases a label needs to be incorporated into the probe DNA. Radioactivity is a common label, but non-radioactive probes are also available. Four methods for incorporating label into DNA probes have been described (Box). Nick translation is an older technique that has been replaced by random

priming. Random priming is the method of choice for labeling cloned DNA fragments. Synthetic oligonucleotides are labeled using T4 nucleotide kinase.

Random Priming.



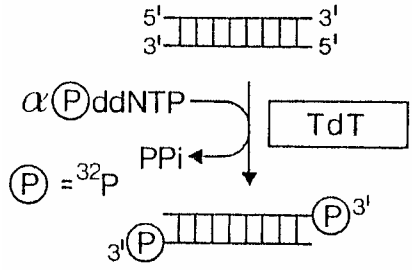
In random priming (Figure) DNA is denatured by heating and mixed with hexamers of random sequence (i.e., random primers). The random primers are usually synthesized and included as part of a kit. They can also be prepared from genomic DNA. A few of primers will be complementary to the probe DNA and the duplex formed between the primer and the probe DNA will serve as an initiation point for the DNA polymerase. The DNA polymerase used is **Klenow**. Klenow is the large subunit of DNA polymerase I in which the 5'→3' exonuclease activity is removed. The four dNTPs including a nucleotide containing a radioactive phosphate in the α-position are also added to the mixture. Therefore, the Klenow will make radioactive copies of the template DNA. The probe DNA is boiled immediately before use in the hybridization assay to convert the dsDNA to ssDNA.



T4 Polynucleotide Kinase.

T4 polynucleotide kinase transfers the γ-PO_4 from ATP to the 5'-hydroxyl of polynucleotides. It is therefore necessary to dephosphorylate the DNA with alkaline phosphatase (AP) before carrying out the phosphorylation. A disadvantage of this technique is that only one radioactive atom is incorporated per DNA strand. However, 5'-terminal phosphorylation is widely used to label oligonucleotide probes that have been prepared synthetically. Synthetic oligonucleotides lack the 5'-phosphate and are too short for random priming. T4 kinase is also used in Maxim and Gilbert DNA sequencing and to phosphorylate (non-radioactive) synthetic linkers.

Terminal Transferase.



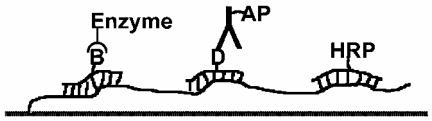
Terminal deoxynucleotide transferase (TdT) adds dNTPs to the 3-OH of either ssDNA or to 3' overhang. In the presence of Co^{2+} TdT will add dNTPs to the 3'-OH of either dsDNA or 5' overhangs. TdT can be used to radiolabel 3' ends if radioactive nucleotides are used. A more common use, however, is to generate homopolymer tails for molecular cloning.

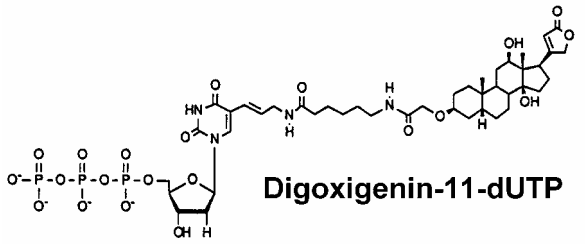
Non-radioactive Probes.

Several procedures have been devised for the detection of hybridization using non-radioactive probes (Box). All are based upon enzyme-linked systems using either alkaline phosphatase (AP) or horse-radish peroxidase (HRP). Biotinylated dNTPs can be incorporated into the probe DNA by random priming. The probe is then be detected with an enzyme-linked streptavidin. Another approach is to incorporate digoxigenin-11-(d)UTP into the DNA probe and then subsequently detected with enzyme-linked antibody against the digoxigenin. A third method is to directly cross-link HRP to the DNA probe.

Labeling/Detection Systems

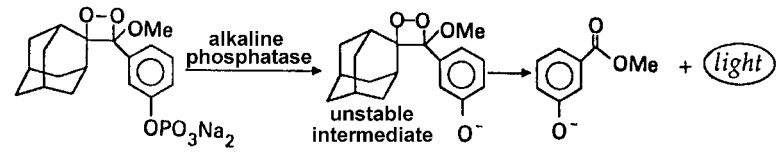
- Biotin/Streptavidin
- Digoxigenin-UTP/Antibody
- Enzyme Crosslinking

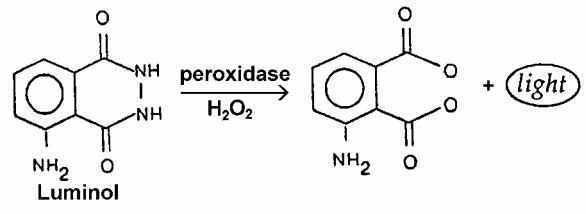




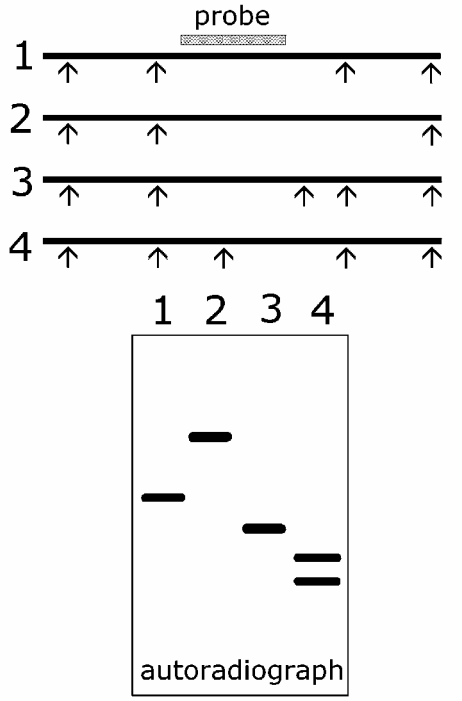
Radioactive probes are generally more sensitive and reliable. However, non-radioactive probes can be adapted to many applications and eliminate some of the problems associated with the use of radioactivity such as waste disposal and safety issues. In addition, the use of non-radioactive probes is are particularly advantages in situation where the same probe is going to be used over a long period of time. The short half-life of ^{32}P (14 days) necessitates that the probe be prepared on a monthly basis, whereas large amounts of a non-radioactive probe can be prepared and stored for long periods of time.

Insoluble substrates, as described for Western blots, or chemiluminescent substrates can be used in association with Northern and Southern blots. Chemiluminescent substrates produce light when cleaved by the appropriate enzyme and this light is detected by autoradiography. Substrates for both alkaline phosphatase (1,2 dioxetane) and peroxidase (luminol) are available (figures). The use of chemiluminescence allows the blot to be striped and reprobed.





RFLP AND RESTRICTION MAPPING

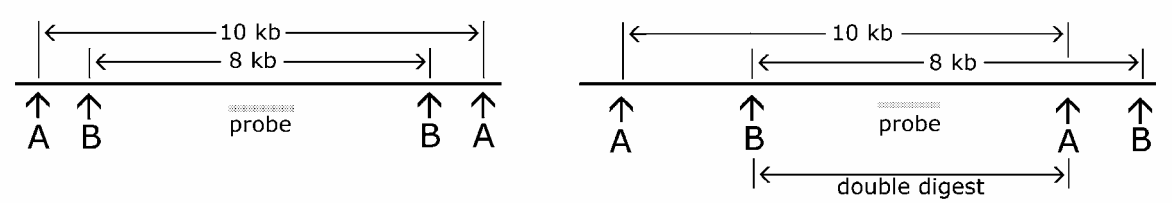


Restriction fragment length polymorphisms (RFLP) refers to the gain of loss of restriction sites associated with a genetic locus. These polymorphisms can be used to distinguish species, strains or even individuals. Single nucleotide mutations can result in the loss or gain of restriction sites, and thus generate different sized DNA fragments associated with that locus. The typical procedure is to digest the target DNA with a restriction enzyme(s), separate the fragments by gel electrophoresis, and to detect the fragments of interest by Southern blotting (Figure). The changes causing the size polymorphisms do not have to be directly associated with the genetic locus being used for a probe.

RFLP applications range from determining specific mutations associated with a particular gene to a genetic 'fingerprinting'. The optimal restriction enzyme(s) and probes will need to be empirically determined for a particular application. A limitation for the application of RFLP is that relatively large amounts of highly purified DNA are needed for many applications. In addition, the procedure is laborious and can take several days to carry out.

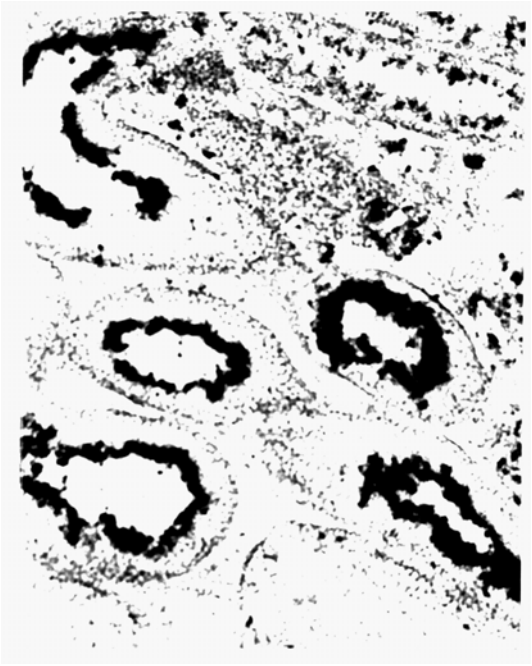
Restriction Mapping. Genetic loci can be defined by restriction maps which show the relative positions of restriction sites. The maps can be generated from sequence data or determined experimentally when sequence data is not available. Restriction maps are generated by digesting DNA with different combinations of enzymes and analyzing the digests by gel electrophoresis or Southern blotting. The sizes of the fragments will indicate the distant between restriction sites. The general procedure is to digest the sample DNA with individual restriction enzymes and combinations of the enzymes. The change in the size of the restriction fragments in double digests is used to determine the relative positions of different restriction sites.

For example, in the situation on the left (see figure below) digestion with enzyme A produces and single 10 kb fragment and digestion with enzyme B produces a single 8 kb fragment. Carrying out a double digest with a mixture of enzymes A and B would result in the same 8 kb fragment as digesting with enzyme B alone. This indicates that both of the B restriction sites are between the A restriction sites, but no information about the distance between site A and site B is obtained. If the situation is as depicted on the right, then a fragment that is < 8 kb is produced. The size of this fragment will indicate the distance between the A and B restriction sites. A more precise map can be generated in both situations if all of the restriction fragments are detectable.



IN SITU HYBRIDIZATION

Cell or tissue specific gene expression can be determined by in situ hybridization. Tissues are fixed on slide and hybridized with cDNA probes. The cells expressing the gene of interest are detected by autoradiography. The cells expressing the gene of interest are identified by the reduced silver grains (dark areas in figure) after development of the photographic emulsion. The probes are usually made with ^{35}S rather than ^{32}P. The probes can also be labeled with fluorochromes and detected by fluorescent microscopy. In situ hybridization is commonly used in developmental biology to determine when and where developmentally regulated genes are expressed.



DNA MICROARRAYS

Most of the hybridization methods are designed to analyze one gene at a time. The DNA microarray technology, also called gene chips, genome chips, biochips, DNA chips, provides a means to simultaneously analyzed thousands of the genes. The array is an orderly arrangement of known DNA samples. Sample spot sizes in macroarrays are generally > 300 microns in diameter and can be imaged with conventional gel and blot scanners. Microarrays contain thousands of spots that are generally less than 200 microns in diameter. Preparation of microarrays requires specialized robotic equipment and analysis of the microarrays requires special imaging equipment.

The DNA microarray is fabricated by spotting known DNA samples, or 'probes', on a solid support such as glass or nylon through the use of high speed robotics. The probes can either be known cDNA sequences (500-5000 bases) or synthetic oligonucleotides (20-80 nucleotides). The array is then exposed to flourescent-labeled cDNA prepared from total mRNA. This free 'target' DNA will bind to the spots containing homologous sequences. The identity and abundance of the complementary sequences can then be determined. It is also possible to mix DNAs prepared with different fluorescent labels. Signals are measured as absolute intensities for a given target, or as ratios of two probes with different fluorescent labels representing two separate treatments to be compared or with one probe as an internal control.

http://www.bsi.vt.edu/ralscher/gridit/intro_ma.htm

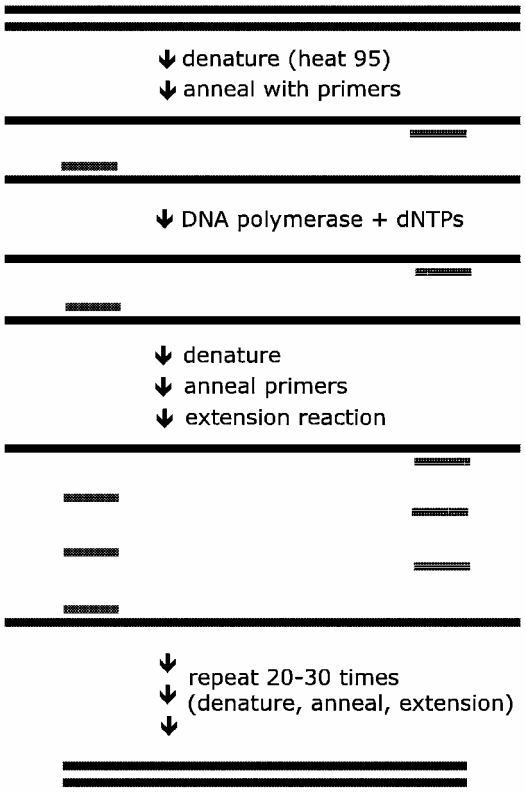
CHAPTER 20--POLYMERASE CHAIN REACTION

PCR Applications
- Diagnosis
- Taxonomy
- Forensics
- Molecular Epidemiology
- Gene Expression
- Sequencing
- Gene Cloning
- Probe Generation
- Site-Directed Mutagenesis

One problem with detecting specific DNA sequences, especially those of unique genes, is that a relatively large amounts of pure DNA are needed. In addition, blotting techniques are laborious and time consuming to carry out. To circumvent these problems it is possible to enzymatically amplify a specific region of DNA using the polymerase chain reaction (PCR). The ability to detect minute amounts of specific DNA sequences has resulted in the broad application of PCR for diagnosis, forensics, molecular epidemiology, etc. In addition, PCR is an integral aspect of many methods, such as gene cloning and sequencing.

PCR MECHANISM

PCR amplifies a specific segment of DNA that lies between two known primer sequences. DNA strands are separated by heating and then annealed with a pair of **primers** which are complementary to the opposing strands (Figure). DNA polymerase recognizes this small region of duplex DNA as a substrate and in the presence of nucleotides will synthesize the complementary strands of both template strands. If this procedure is repeated, the newly synthesized fragments also serve as templates for subsequent rounds resulting a geometric amplification. The end product of PCR is a dsDNA molecule that is defined by the 5'-ends of the primers. In other words, the length of the DNA molecule is determined by the distance between the primers.



Typically, PCR results in a million-fold amplification of the target DNA. Therefore, sequences that only represent a small proportion of the total DNA can be detect after PCR and generally the DNA does not have to be highly purified. Furthermore, the procedure is relatively easy to carry out and does not require expensive equipment or reagents.

One of the major factors in the success of PCR is the use of DNA polymerases isolated from thermophilic organisms such as *Thermus aquaticus*. **Taq polymerase** is thermostable and exhibits an optimal temperature for activity at 72-74^{o}. Other thermostable DNA polymerases are also available. The temperature necessary for the separation of DNA strands will destroy the DNA polymerase activity from other sources. In addition, the optimal temperature for activity

(i.e., 37^o) would result in low stringency hybridization leading to the synthesis of irrelevant fragments. The need to change the temperature during the PCR procedure (i.e., denaturation, primer annealing, polymerization) requires a **thermocycler**. The thermocycler is an instrument that rapidly and accurately changes the temperature of a metal block designed to hold 0.5 ml microcentrifuge tubes.

PCR Procedure
1. Mix DNA, primers, dNTPs and Taq polymerase.
2. Set the thermocycler for desired times.
3. Analyze amplified DNA.

PRACTICAL ASPECTS

The first step of PCR is to combine the template DNA, primers, dNTPs, Mg^{2+} and Taq polymerase in a single tube (Box). Special thin-walled tubes of uniform thickness are used to ensure rapid and equal temperature changes throughout the reaction volume. Typical concentrations are approximately 0.1 pM of the target DNA, 2 nM of primers with a $T_m > 55^o$, 20-200 μM dNTPs (lower concentrations lessen mispriming), and 2 units of enzyme. The optimal Mg^{2+} concentration depends upon the total dNTP concentration which includes: free dNTPs, primers, and template DNA. Since the template DNA is the most variable, it is generally necessary to titrate the optimal Mg^{2+} concentration for different sources of templates. In addition, the EDTA (a typical component of buffers used for nucleic acid isolations) will affect the free Mg^{2+} concentration.

The thermocycler is programmed for the desired times and temperatures of denaturation, annealing and polymerization (Table). The optimal conditions are determined empirically. Most thermocyclers can also be programmed to vary these parameters in different cycles. For example, the denaturation step in the first cycle is sometimes carried out for 2-5 minutes to ensure a more complete melting of the target DNA. Likewise, after the last polymerization cycle it is possible to lower the temperature to 4^o (i.e., storage conditions).

PCR Parameters

STEP	TEMP	TIME	NOTES
Denature	94-96^o	0.5-2 min	longer times ↑ denaturation, but ↓ enzyme and template
Annealing	15-25^o < T_m	0.5-2 min	higher temp. and shorter times ↑ specificity, but ↓ yield
Polymerization	72-75^o	~1 min (<kb)	Taq processivity = 150 nucleotides/sec

Primer Design
- template length
- uniqueness (18-28 bases)
- no primer dimers
- no internal complementarity
- 50% GC ratio
- 3'-GC 'caps'
- HPLC purification

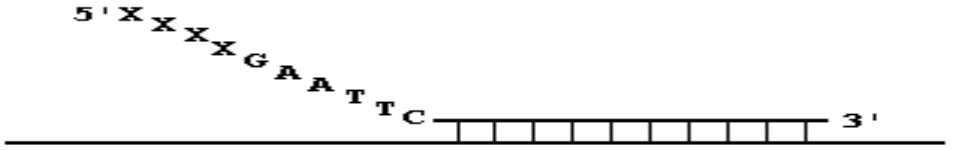
The oligonucleotide primers are the most critical element in terms of successful PCR (Box). Computer programs are used to examine DNA sequences for potential primers. One consideration is distance between the primers. Smaller DNA fragments are amplified more efficiently than longer DNA fragments and it is often particularly difficult to amplify fragments larger than one kb. The primers should also be unique in that they should not hybridize to other sites in the target DNA. Generally primers of 20 nucleotides or greater will provide a satisfactory

level of uniqueness. The primers should not exhibit complementarity to each other or internal complementarity within a primer. Both of these will phenomenon will prevent the primer from annealing to the template DNA. A balance nucleotide composition also tends to improve the function of a primer. Designing a primer so that the last one or two bases on the 3'-end are either a G or a C will result in a stronger hybridization and ensure recognition by the polymerase. Purifying the primers by HPLC will improve their fidelity, but is not absolutely necessary.

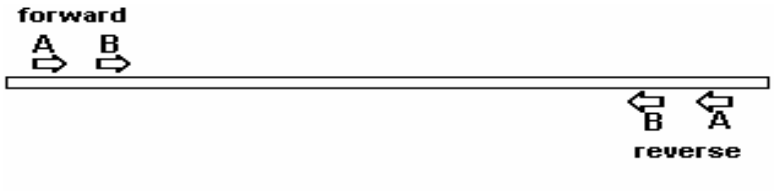
Amplified DNA is analyzed by gel electrophoresis. In most applications the desired result is a single band, or **amplicon**, detectable with ethidium bromide. Primers are usually designed so that a sample can be tested for the presence or absence of the expected band. Size heterogeneity can be determined if exhibited by that particular locus. Southern blotting of the PCR products can also be carried out for either increased sensitivity or specificity.

Site Directed Mutagenesis

PCR can also be used to manipulate DNA. For example, site-directed mutagenesis can be carried out by designing primers with single nucleotide mismatches. Since the primers serve as templates in subsequent rounds of DNA replication the PCR products will contained the introduced nucleotide. Similarly, restriction sites are easily added to the PCR products for subsequent subcloning as illustrated by the following figure:

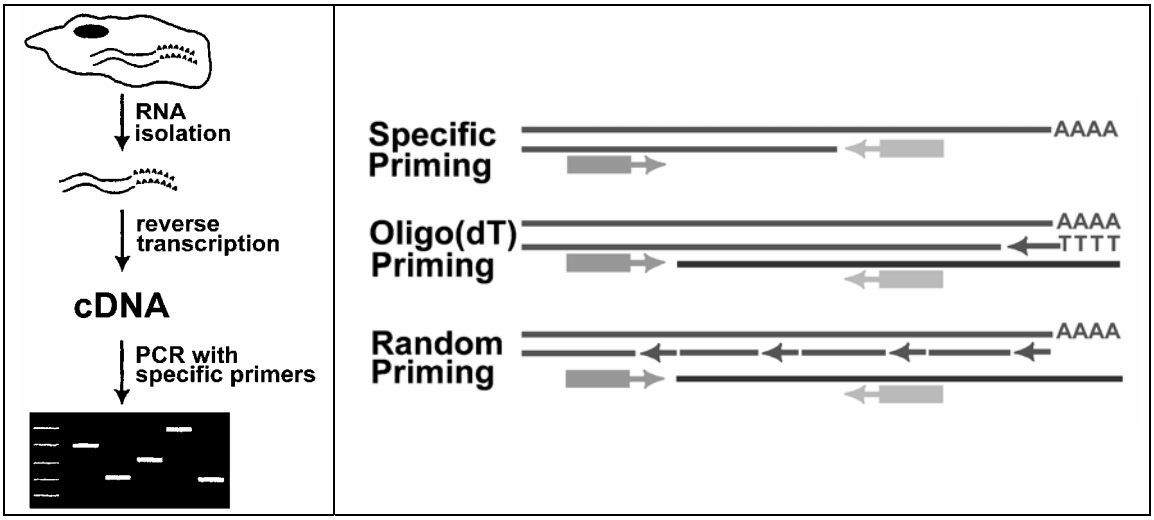


Generally 25-30 cycles will yield the maximum amount of amplification. This is due to loss of enzyme activity associated with the high temperatures and to the accumulation of products that inhibit or interfere with the amplification. Such products include fragments synthesized as a result of mispriming and short fragments formed as a result of the polymerase dissociating from the template before reaching the opposite primer. Subjecting the sample to a second round of amplification can increase the sensitivity. Typically the reaction product from the first cycle is diluted by a 100-1000 fold and the primers, nucleotides and polymerase are re-added. Similarly, a **nested PCR** can be designed in which a second amplification is carried out using a primer pair internal to the first set of primers (Figure).



RNA-PCR

It is also possible to use PCR for the analysis of RNA and gene expression. This method, called RNA-PCR or RT-PCR, detects specific transcripts and exhibits a greater sensitivity than Northern blotting. The first step of RT-PCR is to make a DNA copy of the mRNA (i.e., cDNA). The copy is made using **reverse transcriptase** (RT), an enzyme of retroviruses which exhibits RNA-dependent DNA polymerase activity. There are three basic strategies for the synthesis of cDNA: specific priming, oligo-dT priming, or random priming (Figure). The cDNA is then subjected to PCR using a specific primer pair.



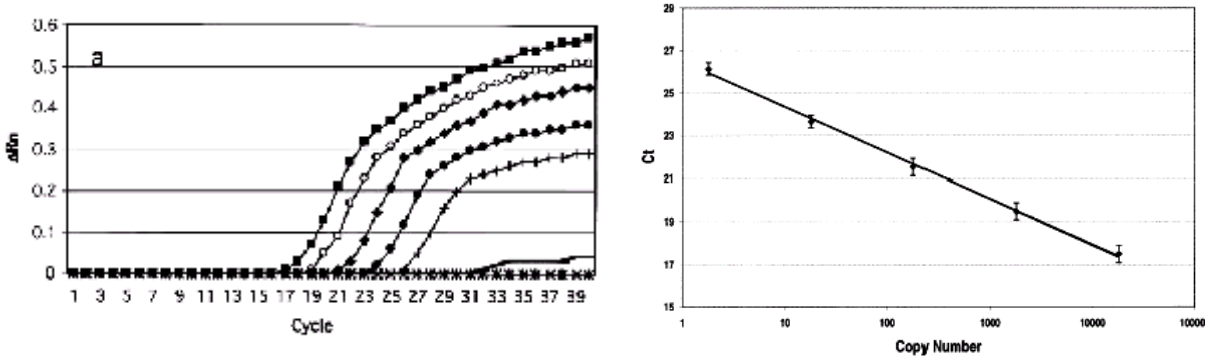
QUANTITATIVE PCR

Information obtained from PCR is primarily qualitative. The major problem is that the analysis is an end-point determination. Since the inhibitory affects accumulate the differences in the initial template concentration will be masked. In addition, quantification of the end-point amplification product usually requires some post-amplification handling which increases the labor and time of the assay.

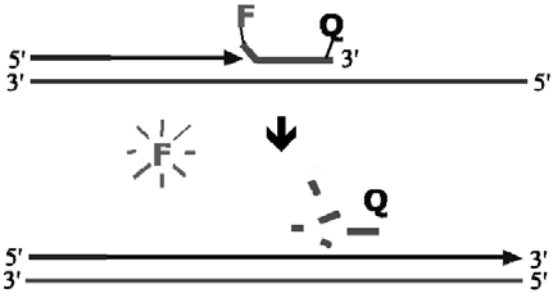
One approach to obtaining quantitative information is to include a competitive reporter with the template DNA. This reporter has the same primers as the target DNA, but produces a different size amplicon. The target DNA is mixed with increasing known amounts of the competitor and analyzed by PCR and gel electrophoresis. Since the reporter and the target DNA are in the same tube, they will be similarly inhibited. Therefore, the amount of the respective amplicons produced will be proportional to the initial template concentrations. Approximately equal amounts of competitor and target amplicons will be produced when their respective template concentrations are equal, and therefore, the amount of target DNA (i.e., number of copies) can be estimated from the amount of reporter known to be in that sample.

Another method for estimating the concentration of target DNA is by 'real time'

PCR. Real time PCR utilizes fluorescent probes for the detection of PCR products and requires a special thermocycle (i.e., 'light cycler') which monitors fluorescence during the reaction. The simplest method to monitor the accumulation of PCR products is to include a dye in the PCR reaction which fluoresces when intercalated into dsDNA. SYBR green is the most widely used dye for this application. The dye binds to the double-stranded PCR products and fluoresces. Fluorescence is measured during the extension stage of each cycle and thus represents the the accumulation of PCR products during the reaction. The cycle in which significant fluorescence is first detectable is inversely proportional to the amount of initial template. A standard curve can be prepared from this threshold cycle (C_t) and the template concentration of known standards. The amount of template present in an unknown sample can then be extrapolated from this standard curve.

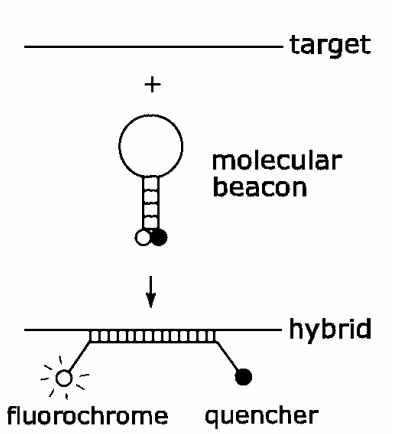


Since these dyes bind to any dsDNA, specificity can be a problem. This is especially true towards the end of the PCR cycles when more non-specific products may be accumulating. Furthermore, it is not possible to quantify multiple PCR products within the same sample (i.e., multiplexing). Several single-stranded DNA detection methodologies have been developed to overcome these limitations. Two examples include hydrolysis probes (aka, F-Q probes or Taqman probes) and molecular beacons. Both of these methods rely upon the use of specific probes to the amplified sequence which have been coupled with fluorochromes and quenchers, thus adding some expense to the these assays. Through the use of fluorochromes with different emission spectra it is possible to develop multiplex assays in which differenent PCR products can be monitored simultaneously in the same sample.



Hydrolysis probes. A probe homologous to a sequence between the two primers is designed. One end of the probe contains a fluorochrome (eg., 5') and the other end contains a quencher (eg., 3'). The energy of the excited fluorochrome is transferred directly to the quencher and therefore this probe does not fluoresce. During annealing the probe also hybridizes to the target DNA. As the polymerase progresses down the template during the extension reaction the endogenous 5'-

exonuclease activity digests the F-Q probe and thereby releases the fluorochrome and quencher. The fluorochrome and quencher are no longer held together in close proximity and therefore the free fluorochrome fluoresces. Free fluorochrome accumulates after each cycle and the fluorescence intenesity is a measure of the amount of specific PCR product. There tends to be a high background associated with the F-Q probes since the fluorochrome and quencher are separated by approximately 20 nucleotides.



Molecular beacons. Probes which form 'hiarpins' have been designed to improve the intramolecular quenching. As in the case of the F-Q probes a probe homologous to a sequence between the primers is designed. Complementary sequences are added to each end as well as the the fluorochrome and quencher. These complementary sequences will anneal and therefore force the fluorochrome and quencher in close proximity and thus increase the efficiency of quenching. Hybridization of the probe to the target DNA results in fluorescence since the fluorochrome and quencher are now held far apart. As in the case of using non-specific dsDNA binding dyes, the fluorescence is only observed during the annealing step. During the extension reaction the probe is displaced and will be recycled.

Special considerations are needed for probe and primer design in both the Taqman and molecular beacon assays. The optimal amplicon length for both assays is 50-100 bp. In both cases the probe needs to bind to the target in a quantitative manner and with the desired specificity. The T_m of the probe-target hybrid should by 8-10^oC above the annealing temperature of the PCR reaction. This is usually obtained with probes of 20-26 nucleotides. The exact length will depend on GC content and the amount of tolerance for mismatches. The stem region for the molecular beacon probes should also exhibit a T_m of 8-10^oC above the annealing temperature so that unhybridized probe is kept in the non-fluorescent conformation during the annealing step. Stems of six nucleotides, of which five are G:C pairs are usually the optimum.

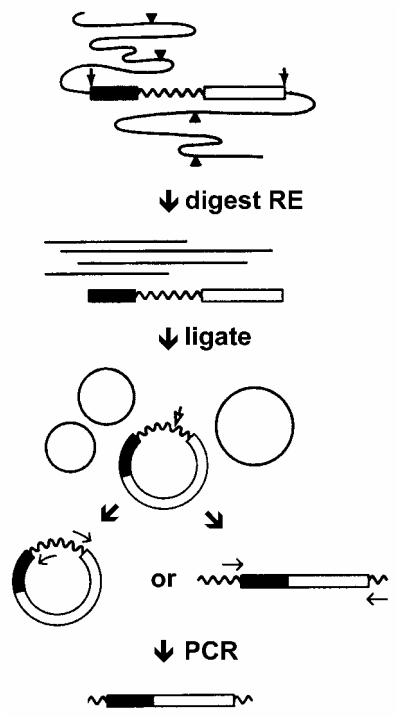
- inverse PCR
- anchor PCR
- RAPD
- AFLP

LIMITATIONS

One limitation of PCR is that only DNA between two known primer sequences is amplified, thus making it difficult to analyze unknown sequences. Modifications of the PCR method which overcome this limitation have been described (Box).

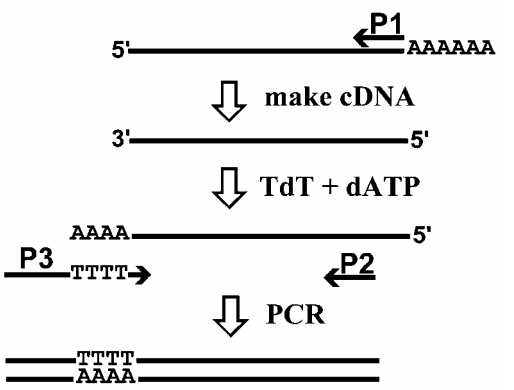
Inverse PCR (also known as chromosome crawling) is a method to amplify the sequences flanking a known sequence. Genomic DNA is digested with a restriction enzyme which produces appropriately sized fragments (eg., 3-5 kb) containing the target sequence. The fragments are circularized with DNA ligase and amplified with primers that are directed away from each other. This results in the amplification of the sequences that directly flank the known

sequence. The boundary between the two flanking sequences is demarcated by the restriction site used to disgest the genomic DNA.



Known sequences, or 'anchors', can be also added to unknown sequences and used as the source of PCR primers. This anchor PCR is often used to amplify the unknown 5'-end of a cDNA (i.e., mRNA). (Sometimes called RACE for rapid amplification of cDNA ends.) In this method a cDNA copy of the mRNA is made using a primer based on known sequence (P1). A homopolymer tail is added to the 3'-end of the cDNA with terminal deoxynucleotide transferase. A primer containing sequence complementary to the homopolymer tail (P3) is used in conjuction with a primer from the known sequence (either P1 or P2) to amplify the target DNA.

Random portions of the genome can also be amplified as a means to identify polymorphisms without prior knowledge of the particular polymorphism. For example, random amplified polymorphic DNA (RAPD), also called arbitrarily primer PCR (AP-PCR), uses short (usually 10 nucleotides) random sequence primers to amplify random portions of the genome. These short primers will anneal to a number of locations within a genome and in loci with inverted repeats the appropriate distance apart an amplicon will be produced. The fingerprint of DNA fragments can then be compared between species or strains and analyzed for polymorphisms. The RAPD method has several advantages as an assay for polymorphism. However, it does tend to exhibit poor reproducibility.



Advantages of RAPD

- requires small amounts of DNA (15-25 ng)
- no prior sequence information needed
- easily produced and screened
- relatively easy to convert into more reproducible PCR assay

Another method based on the amplification of random portions of the genome is amplified fragment length polymorphisms (AFLP). The advantage of this method over RAPD analysis is that the polymorphisms are based on restriction sites and not primer binding. Therefore, AFLP analysis tends to be more reproducible than RAPD analysis. The first step in AFLP analysis is to digest genomic DNA with two restriction enzymes. Typically one of the enzymes has a 6-base recognition sequence and the other a 4-base recognition sequence. The DNA is then ligated with adaptors corresponding to the two restriction sites and PCR using primers based on the adaptor sequences are carried out in the

presence of either radioactive nucleotides or fluorochrome-labeled nucleotides. The PCR products are analyzed on sequencing gels for polymorphic fragments. Both AFLP and RAPD are used to identify strain or species specific markers similar to RFLP analysis. However, much less DNA is need and these methods are less laborious and time consuming than traditional RFLP analysis.

PRECAUTIONS

The ability of PCR to amplify minute amounts of DNA is not only a strength, but also a potential weakness. Under controlled situations amplification of a single copy of DNA has been demonstrated. This ability poses a serious problem in terms of contamination of buffers and samples with unwanted DNA. Precautions against spurious contamination need to be taken (Box). It is also advisable to include a control sample (no target DNA added) in every PCR run.

Precautions

- avoid contamination
- aliquot reagents
- add target DNA last
- no target DNA control
- prepare '+' control elsewhere

CHAPTER 21--RECOMBINANT DNA

Molecular cloning, or recombinant DNA, is the process by which a single gene, or segment of DNA, is isolated and amplified. This is accomplished by recombining DNA fragments in vitro and then propagating the DNA fragments of interest. A source of nucleic acids and the ability to manipulate them are needed to carry out molecular cloning (Box).

Cloning Needs
- foreign DNA
- vector + host
- means to cleave DNA
- ability to join DNA fragments
- introduction of rDNA into host
- screening mechanism

The source and type of nucleic acid will depend on the gene being cloned and the species of interest. The **vector**, when introduced into the appropriate host, is a replicating piece of DNA. The foreign DNA is recombined with a vector, which is then replicated by the host. Available cloning vectors include plasmids, phage/viruses, or some combination of the two. Phage, short for bacteriophage, are viruses of bacteria.

In order to carry out recombination between the vector and the foreign DNA, it is necessary to specifically and reproducibly cleave DNA. Restriction enzymes cleave DNA in a site specific manner and thus provide this requirement. DNA ligase is an enzyme that covalently joins two pieces of DNA. Therefore, restriction enzymes and DNA ligase allow for the recombination of foreign DNA with vector DNA.

The DNA of interest is amplified by introducing the recombinant DNA (rDNA) into a compatible host. The vector is replicated by the host. The introduction of vector in the host depends upon the nature of the vector and host. For example, some viral vectors can be introduced by infection. Introduction of naked viral DNA or plasmid DNA is referred to as **transfection** and **transformation**, respectively. Finally a means to detect the molecular clone of interest, or a screening mechanism, is needed.

PREPARATION OF FOREIGN DNA

The two major sources of foreign DNA for molecular cloning are genomic DNA (**gDNA**) and complementary (or copy) DNA (**cDNA**). Complementary DNA is prepared by making a copy of mRNA. One consideration in deciding whether to use gDNA or cDNA is the copy number and the level of expression of the gene of interest. Single copy genes in large genomes may require extensive screening. A gene expressed at high levels, though, will require less screening to isolate a cDNA clone. Conversely genes expressed at low levels will be difficult to detect in cDNA libraries. In addition, some genes may only be expressed at certain times or in certain tissues that are inaccessible or unavailable, thus necessitating gDNA libraries.

One also needs to consider how the recombinant DNA library will be screened and the questions that will be asked about the cloned gene. For example, many eukaryotic genes contain 'introns' which will result in an interruption of the open reading frame. Therefore, cDNA is usually a better starting material if an expression library is going to be screened or if the cloned gene is going to be used for the production of recombinant protein. Similarly, comparing gDNA

and cDNA will allow for a precise mapping of the introns and splice sites.

Fragmenting gDNA. The extreme length of gDNA necessitates that it be fragmented into smaller pieces before molecular cloning (Box). The size of the DNA fragments should take into account the vector capacity and the size of the desired clone. Restriction enzymes reproducibly cleave DNA at specific sites which are relatively easy to ligate into restriction sites of vectors. If some randomness is needed (eg., expression libraries and overlapping clones), it may be possible to carry out the digestions under 'star' conditions or to use frequently cutting enzymes under conditions where the target DNA is only partially digested.

Fragmenting gDNA
- restriction enzymes
- random nucleases
- mechanical shearing

Preliminary Southern blots are often carried out when a DNA probe is available. The gDNA is digested with several different enzymes and analyzed for appropriate sized fragments. The appropriateness of the size depends on the size of the genetic locus and the vector being utilized. After deciding which restriction enzyme(s) is the most appropriate, a sufficient quantity of DNA is digested and electrophoresed. The region of the gel corresponding to the desired size range is excised and the DNA isolated. This size-fractionated DNA is then used to prepare a library.

It is also possible to use a non-specific nuclease under controlled conditions to cleave gDNA into appropriately sized pieces. For example, DNAse I in the presence of Mn^{2+} will cleave dsDNA at approximately the same site on both strands. Pilot experiments are carried out to determine the optimal DNase/DNA ratios for the production of the desired sized fragments. An application of this method is in 'shotgun' sequencing strategies which result in the production of several overlapping clones. Mung bean nuclease in the presence of formamide will also randomly cleave dsDNA.

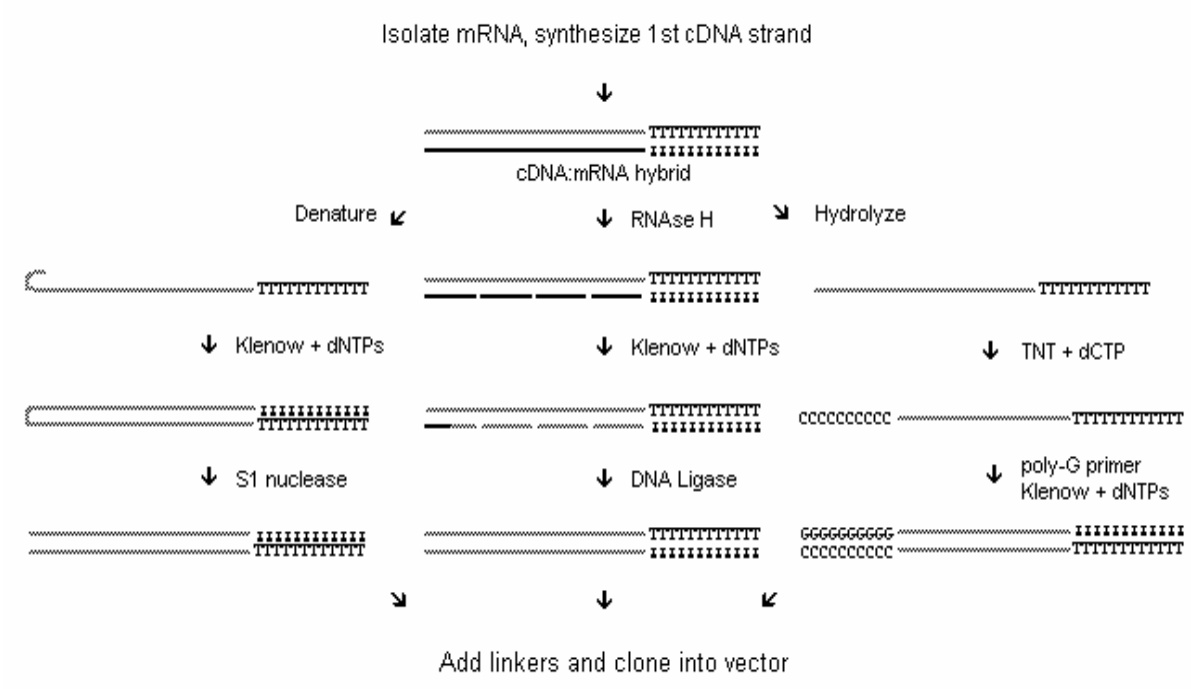
DNA can also be broken by mechanical shear forces. Passage of a DNA solution through a small gauge syringe needle or sonication will result in the DNA molecule being broken in random positions. Following fragmentation by either method the DNA can be size fractionation by either rate zonal centrifugation on sucrose gradients or by gel electrophoresis. Restriction sites can be added to the ends of the gDNA fragments to increase the efficiency of cloning (see below) as compared to blunt-end ligations.

PCR fragments can also be cloned. Amplifying DNA fragments of interest before cloning will eliminate, or at least reduce, the screening of recombinant DNA libraries. However, the cloning of PCR fragments is less efficient than anticipated due to the non-template dependent addition of a single nucleotide to the 3'-end of PCR products by many thermostable DNA polymerases. The most frequently added nucleotide is an adenosine residue. Vectors containing a single 3'-T overhang (i.e., TA cloning) greatly enhances the efficiency of cloning. In addition, it is possible to generate a restriction site during the amplification process (see PCR).

Preparing cDNA Libraries. The preparation of cDNA from mRNA is a multistep process (Box). The first step in preparing cDNA is to isolate mRNA. mRNA tends to be less stable than DNA and requires more precautions than gDNA. In addition, an enrich source of mRNA of interest can be used. Enrichment for a particular mRNA could include: choice of species, tissue, developmental stage, induction of expression (eg., hormones), etc. It may also be possible to enrich for the mRNA of interest by size fractionation, immunological purification of polysomes, or substractive libraries. The integrity of mRNA should be evaluated by Northern blotting and/or in vitro translation before proceeding with the preparation of the library.

Preparation of cDNA
1. Isolate mRNA
2. Synthesize DNA-RNA hybrid
3. Synthesize 2nd DNA strand
4. Add termini

After obtaining the desired mRNA a DNA-RNA hybrid is synthesized with reverse transcriptase (RT). Reverse transcriptase is a RNA-dependent DNA polymerase isolated from retroviruses. In the presence of a mRNA template and primer, RT will copy the RNA template resulting in the formation of a DNA-RNA hybrid. An oligo-dT primer of 12-18 bases in length is used to initiate synthesis on poly-A messenger RNA. Sequences corresponding to restriction sites can also be included in the oligo-dT primer. For example, a common primer-linker for the synthesis of cDNA is 5'-(GA_{10})ACTAGTCTCGAG$(T)_{18}$-3' which includes *Spe*I and *Xho*I sites. Random priming can also be used to increase the chances of obtaining the 5'-end of long mRNA or if the mRNA of interest lacks a poly-A tail.



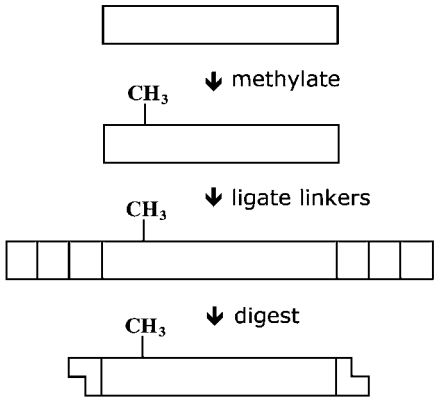
Three Methods for the Synthesis of cDNA 2nd Strands

The RNA strand is replaced with a second DNA strand following the synthesis of the RNA-DNA hybrid. Three methods for synthesizing the 2nd DNA are: 1) self-priming, 2)

replacement synthesis, and 3) primed synthesis (Figure). In self-priming the RNA-DNA hybrid is denatured by heating or the RNA is hydrolyzed with NaOH. The 3'-termini of single-stranded cDNA forms a hairpin loop that can function as a primer for DNA polymerase. Following the completion of second strand synthesis, the ssDNA loop is then removed with S1 nuclease. The self-priming is a poorly controlled reaction and treatment with S1 nuclease sometimes leads to loss of the 5'-end of the message.

Replacement synthesis uses the DNA-RNA hybrid as a template for a 'nick translation' reaction. The DNA-RNA hybrid is treat with RNAse H which produces nicks in the RNA strand. The nicked RNA then functions as primers for DNA polymerase. In primed synthesis, a homopolymeric tail (or other known sequence is added to the 3'-end using terminal transferase (TdT). A primer complementary to the homopolymer tail is then used to synthesize the second strand. This method may result in a higher efficiency of recombinant clones containing the complete 5'-end of the message.

The synthesis of cDNA results in blunt-end DNA without convenient restriction sites for cloning into vectors. Linkers are added to the termini of the cDNA molecules which will provided the necessary restriction sites (Figure). Restriction sites within the insert DNA are protected by methylation prior to the addition of the linkers. The overhangs are then generated by digestion with the appropriate restriction enzyme. The use of primer-linkers allow for different restriction sites on the 5'-end and the 3'-end (i.e., poly-A tail) and directional cloning. Another strategy to create sticky ends is to add complementary homopolymer tails with TdT to both the insert DNA and the vector.



Useful Plasmid Features
- relaxed replication
- selectable markers
- streamlined
- unique restriction sites
- recombinant identification

PLASMID VECTORS

Recombinant DNA vectors function as carriers of the foreign DNA. Plasmids are extra-chromosomal genetic elements found in a variety of species and are widely used in molecular cloning. They are autonomously replicating circular DNA molecules that range in size from 1-200 kb. Their replication can be under either **stringent** control (low copy number) or **relaxed** (high copy number). Stringent control means that the plasmid is only replicated when the genome is replicated, whereas relaxed plasmids are replicated continuously. Many plasmids confer antibiotic resistance and are naturally transmitted between bacteria by conjugation.

Naturally occurring plasmids have been modified to make them more useful as cloning vectors (Box). To be useful as cloning vectors, plasmids need to be relatively small, replicate in relaxed fashion, have selectable markers (eg., antibiotic resistance) and contain unique restriction sites. The most widely used plasmids are derivatives of either pBR322 or pUC (where the lower case 'p' means plasmid). Typically modifications of plasmids include the

removal of unnecessary portions, or streamlining, and the introduction of several useful restriction sites in one location. These restriction sites, call the **multiple cloning site** (MCS) or

Example of a multiple cloning site

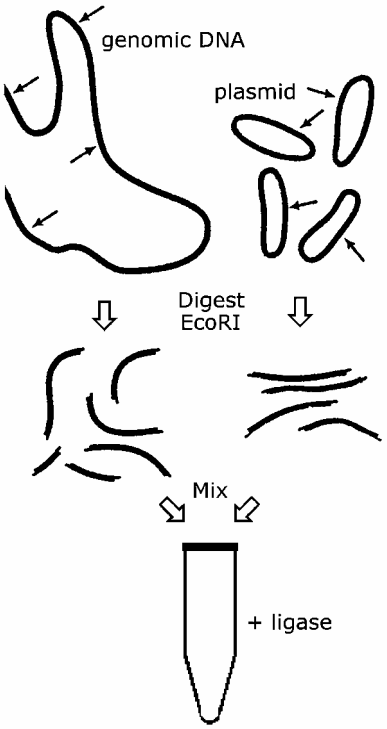
```
|SacI|  |ScII|          |XbaI||SpeI||BamH||SmaI||PstI||EcRI||EcRV||HIII||ClaI|   |SalI||XhoI|         |KpnI|
GAGCTCCACCGCGGTGGCGGCCGCTCTAGAACTAGTGGATCCCCCGGGCTGCAGGAATTCGATATCAAGCTTATCGATACCGTCGACCTCGAGGGGGGGCCCGGTACC
CTCGAGGTGGCGCCACCGCCGGCGAGATCTTGATCACCTAGGGGGCCCGACGTCCTTAAGCTATAGTTCGAATAGCTATGGCAGCTGGAGCTCCCCCCCGGGCCATGG
```

polylinker, facilitate DNA cloning. This MCS is usually located within a gene which allows for easy identification of recombinants.

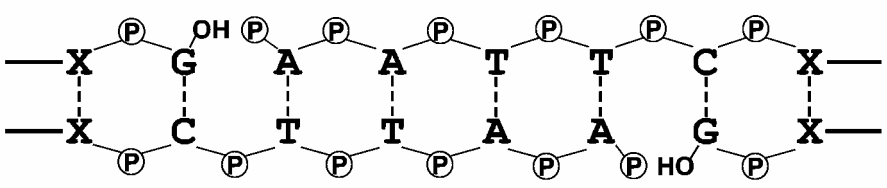
Generic Recombinant DNA Protocol
1. Prepare foreign DNA.
2. Prepare vector.
3. Combine foreign DNA with vector.
4. Introduce recombined vector into host.
5. Screen for rDNA of interest.

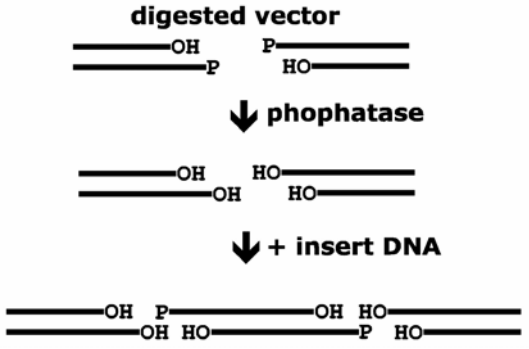
Recombining target DNA with vector.

The first step in cloning involves preparing the foreign or target DNA and simultaneously preparing the plasmid vector (Box). The exact method for preparing the foreign DNA will depend on the source of the nucleic acid and the nature of the gene being cloned. For example, genomic DNA can prepared by digesting with appropriate restriction enzyme(s). The digested DNA can also be size-fractionated on agarose gels and the desired size range of DNA fragments excised from the gel.



The vector is prepared by digesting the plasmid with the restriction enzyme(s) that are compatible with the foreign DNA. Recombinant DNA molecules are generated by simply mixing the prepared vector DNA with the prepared foreign DNA in the presence of **DNA ligase** (Figure). The complementary overhangs, or 'sticky-ends', produced by the restriction enzyme(s) allow the foreign DNA and vector DNA to transiently anneal. DNA ligase catalyzes the formation of a phosphodiester bond between a 5'-phosphate and a 3'-hydroxyl (Figure) and will result in the covalent joining of DNA fragments. Typically, equal molar amounts of vector and foreign DNA are mixed during the ligation reaction to minimize the formation of recombinants with tandem copies of the insert. The molar ratio of vector to foreign DNA can be increased (eg., 1.5-2) to increase the amount of target DNA incorporated into plasmids. Ligation reactions are usually carried out at lower temperatures (8-15^{o}) for extended time periods (2-24 hours) to promote the annealing of the short (2-4 base) overhangs.





During DNA ligations the vector can ligate with itself in cases were the vector has identical overhangs on both ends (including blunt-ends). This intramolecular formation of circles is kinetically favored and many of the vectors will not contain insert DNA (i.e., be non-recombinants). The formation of non-recombinants is minimized by treating the vector with phosphatase prior to the ligation step. This pretreatment favors the formation of recombinants since DNA ligase requires a 5'-phosphate, which can now only be supplied by the foreign DNA insert (Figure).

Commonly used phosphatases are calf intestinal alkaline phosphatase (CIAP) and shrimp alkaline phosphatase (SAP). SAP is inactivated by moderately high temperatures whereas CIAP is more stable and requires higher temperatures or proteinase K treatment for complete inactivation. The phosphatase activity needs to be completely inactivate before mixing the vector with the foreign DNA so that the 5'-phosphates are not removed from the insert DNA. The efficiencies of removing the 5'-phosphate from overhangs (i.e., extensions), recessed ends, or blunt ends are different and the reaction conditions need to be adjusted accordingly. The completeness of the phophatase step is easily checked by carrying out a ligation and transformation without insert DNA and comparing the number of colonies formed to the number of colonies obtained with an equal amount of plasmid which was not treated with phophatase.

Ligation of DNA Fragments into Plasmid Vectors

TERMINI	CLONING REQUIREMENTS	COMMENTS
Identical Overhangs	Phosphatase treatment of linear plasmid improves efficiency.	Restriction sites at junctions preserved. Both orientations of insert DNA possible. Tandem copies of insert possible.
Blunt-end	High concentrations of DNA and ligase needed. Phosphatase treatment.	Restriction sites at junctions often eliminated. Tandem copies of insert DNA possible. Both orientations possible.
Different Overhangs	Purification of double-cut plasmid increases efficiency.	Restriction sites at junctions preserved. Background of non-recombinants is low. One possible orientation of insert. Tandem copies unlikely.

The vector and foreign DNA can also be prepared with restriction enzymes that produce blunt ends. The advantage of blunt ends is that all blunt ends are compatible. However, the restriction sites will be lost or changed when blunt ends from different restriction enzymes are ligated together. In addition, the ligation efficiency is much lower because there are no cohesive termini to temporarily hold the two DNA fragments together. Therefore, higher DNA concentrations (both vector and foreign) and DNA ligase concentrations are needed for blunt-end ligations.

Plasmids with different protruding termini can also be prepared. Advantages of different termini are a low background of non-recombinants, tandem copies of the insert DNA are less likely, and the insert DNA can only be in one orientation.

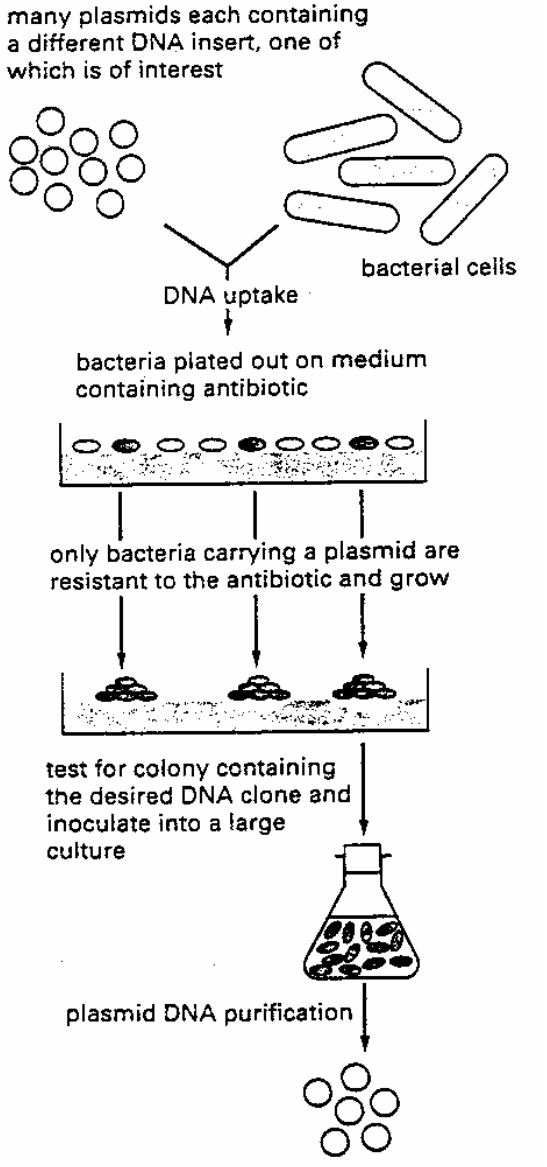
Transformation and Screening.

Transformation Efficiency

	5000	bp/plasmid
X	660	daltons/bp
=	$3x10^{6}$	g/mole plasmid
X	10^{6}	μg/g
=	$3x10^{12}$	μg/mole plasmid
1/X	$3x10^{-13}$	moles/μg DNA
X	$6x10^{23}$	plasmids/mole
=	$2x10^{11}$	plasmids/μg DNA

Heat Shock
(10^{5}-10^{9} colonies/μg DNA)

Electroporation
(10^{9}-10^{10} colonies/μg DNA)



Following ligation, the rDNA is introduced into the host. Bacteria, and in particular *E. coli*, are typically used as host cells. The two most common methods used to **transform** bacteria with plasmids are 'heat-shock' and 'electroporation'. In both cases the bacteria must be made 'competent' (i.e., ready to receive the DNA). In the heat shock method the bacteria are pretreated with $CaCl_2$. The competent cells are incubated with the recombined plasmid to allow DNA to bind to the bacteria surface and then briefly incubated at 40-42^{o}. The heating increases the fluidity of the membrane and

supposedly opens up temporary 'pores' in the bacterial membrane that allows the DNA to enter.

The efficiency of heat shock is typically 10^5-10^9 colonies per μg DNA, which is relatively low. For example, a 1 μg of a 5 kb plasmid will contain approximately 2 x 10^{11} copies. This means that even at an efficiency of 10^9 colonies per μg of DNA only 0.5% of the plasmids were taken up by bacteria during transformation. Electroporation exhibits efficiencies of 10^9-10^{10} colonies per μg DNA and can be used for the transformation of prokaryotic cells. In electroporation, the competent cells are exposed to pulses of high voltage which will produce transient pores in membranes and permit DNA uptake. With both electroporation and heat shock the efficiency of transformation decreases with increasing plasmid size.

Sources of Probes
- previously cloned genes
- synthetic oligonucleotides
- mixed oligonucleotides
- antibodies

After transformation with the recombinant plasmids the bacteria are plated on antibiotic containing media (Figure). The antibiotic will kill the bacteria which do not contain plasmid. Therefore, all of the colonies will represent bacteria with plasmid. However, the bacteria will have different plasmids and it will be necessary to identify the clone of interest by screening the rDNA library. A common method for screening rDNA libraries is to use a 'DNA probe' which hybridizes to the recombinant clone of interest. Sources of DNA probes include previously cloned genes or synthetic oligonucleotides that exhibit sequence homology to the gene of interest (Box). For example, a previously cloned gene from another species that is homologous to the gene of interest can be used as a DNA probe. It is also possible to make a synthetic oligonucleotide probe from highly conserved regions of known genes. Screening expression libraries with antibodies raised against the protein of interest is also possible (see section on expression).

In situations where a homologous gene has not been cloned or not much is known about the gene of interest it may be possible to develop a probe based upon a partial protein sequence (referred to as **reverse genetics**). A mixture of synthetic oligonucleotides, based upon all the possible codons for the polypeptide sequence, is synthesized and used to screen the recombinant DNA library. If possible, the amino acid sequence should be chosen to minimize the redundancy of the genetic code (see Table). For example, methionines and tryptophans, since they only have one possible codon, and amino acids with only two possible codons are especially useful. Conversely, amino acids with six possible codons should be avoided.

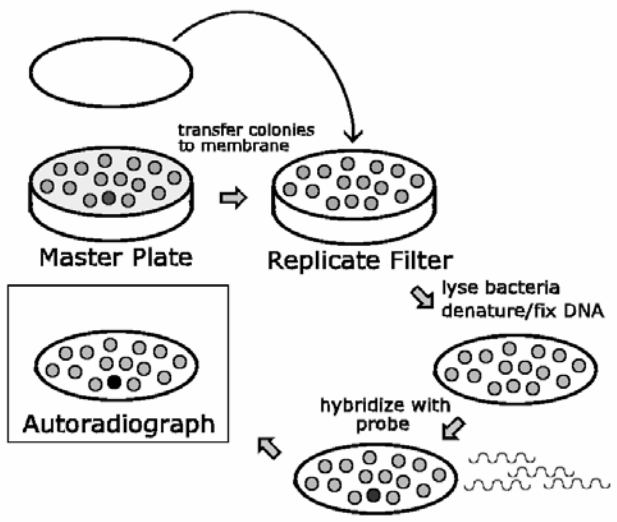
Codons for Hypothetical Heptapeptide

Met	Trp	Glu	Leu	Ile	Ala	Gly
ATG (100)	TGG (100)	GAA (87)	CTT (11)	ATT (40)	GCT (42)	GGT (47)
		GAG (13)	CTA (7)	ATA (53)	GCA (43)	GGA (46)
			CTG (2)	ATC (7)	GCG (4)	GGG (3)
			CTC (2)		GCC (11)	GGC (3)
			TTA (64)			
			TTG (13)			

Corresponding degenerate oligonucleotide for the heptapeptide:

A-T-G-T-G-G-G-A-R-Y-T-N-A-T-H-G-C-N-G-G-N
1x1x1x1x1x1x1x2x2x1x4x1x1x3x1x1x4x1x1x4 = 768

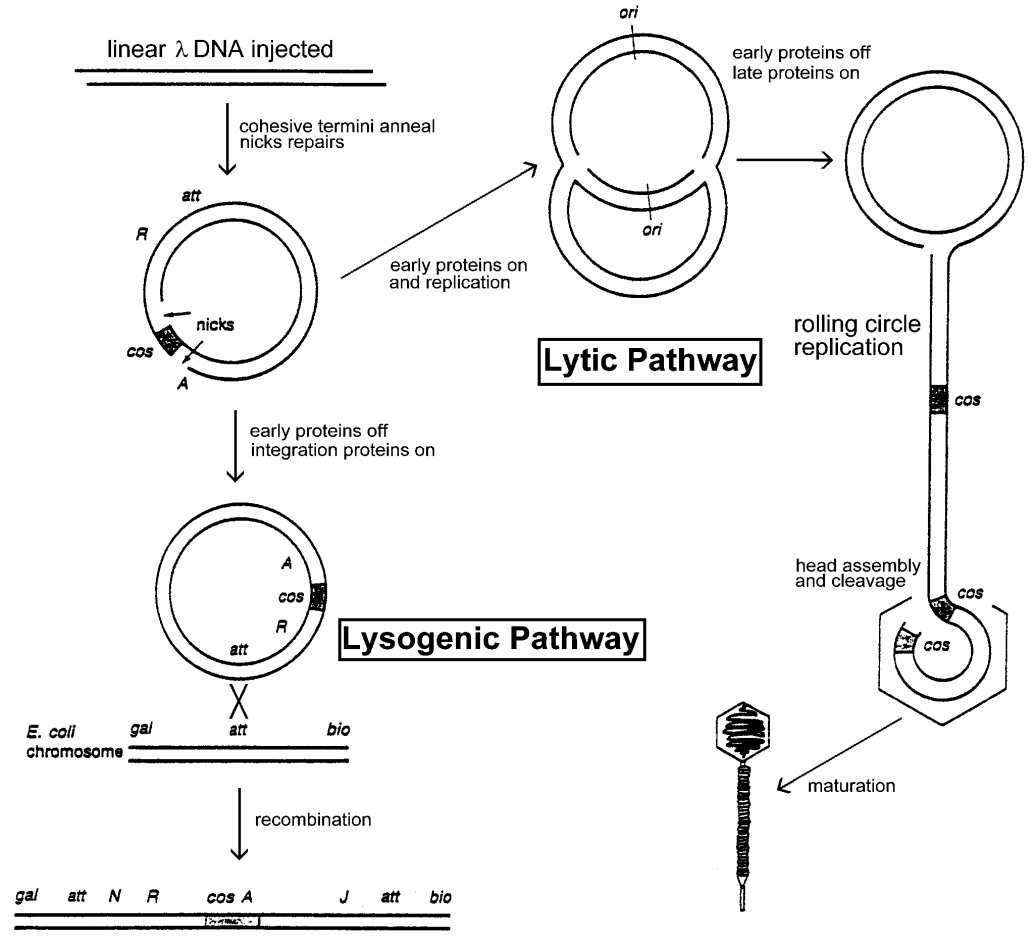
It may be possible to decrease the number of degenerate oligonucleotides by adjusting for the codon usage in the species of interest. Many organisms exhibit a codon-usage bias in that not all codons are used at equal frequencies (see http://www.kazusa.or.jp/codon/ for codon usage tables). For example, the numbers in parentheses are percentages for each the codons in *Plasmodium falciparum*. Since *Plasmodium* species have an AT-rich genome (approximately 80%) the third codon position is very biased towards A and T and it is possible to make less degenerate probe which still may hybridize to the target sequence.



A 'colony lift' is used to detect bacteria containing plasmids of interest (Figure). In this procedure, a DNA probe is used to screen the rDNA library utilizing blotting techniques as discussed previously. The transformed bacteria are grown on agar plates in the presence of antibiotic. When bacterial colonies become visible a sterile nylon membrane is laid on top of the plate. Orientation marks are made on both the membrane and plate so that the pertinent colonies can be identified later. The membrane is then carefully remove and laid onto another nutrient agar plate with the colony side up. In other words, this membrane is a replica of the original master plate. Both the original plate and the membrane are returned to the incubator until the colonies are the desired size. The bacteria are lysed and the plasmid DNA is released directly onto the membrane by laying the membrane (colony side up) on filter paper saturated with buffers containing SDS. The released DNA will bind to the membrane and it is then denature by moving the membrane to a filter paper saturated with NaOH. Following neutralization, the DNA is fixed to the membrane by baking or UV cross-linking, and then incubated with a radioactive DNA probe as described for Southern blotting. The probe will hybridize to plasmids containing homologous insert sequences and result in a spot on the autoradiograph corresponding to the position of the colony containing the recombinant plasmid. The positive colonies are identified on the original plate by aligning the autoradiograph with the master plate. The bacteria can be streaked on an agar plate and retested for the plasmid of interest to insure that the original bacteria colony was derived from a single bacteria. Bacteria from a positive colony are expanded in liquid culture media containing antibiotic and the recombined plasmids (i.e., cloned DNA) isolated for further characterized.

BACTERIOPHAGE λ

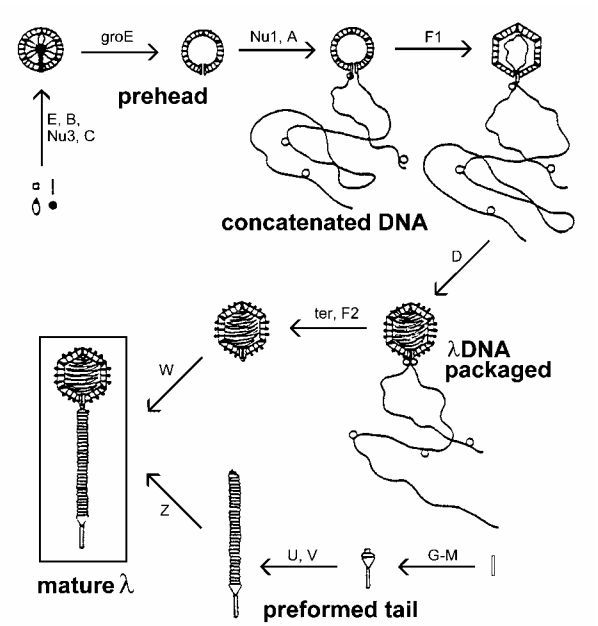
Viruses can also used as cloning vectors. The ability of a virus to replicate within a host cell is exploited as a means to amplify DNA of interest. Bacteriophages are viruses that infect bacteria and are used in various recombinant DNA methods. Phage λ has proven to be an exceptionally powerful vector for molecular cloning.



Phage λ Life Cycle. 1) Phage attaches to bacteria. 2) DNA is injected and circularizes vis-a-vis cos ends. 3) DNA is transcribed and replicated*. 4) Phage assembly. 5) Cell lysis (plaque formation). *Alternatively λ can intergrate into host genome and undergo a lysogeny.

Phage λ infects *E. coli* by injecting its DNA into the bacteria (Figure). The phage DNA is initially linear but forms a circle as a result of cohesive, or **cos**, ends. Phage proteins are transcribed from the phage DNA and the genome is replicated. DNA replication proceeds by a 'rolling circle' mechanism that generates long linear molecules composed of consecutive λ genomes. Infectious particles are then assembled from the phage proteins and DNA. The proteins making up the phage heads self-assemble into preheads. Other phage proteins bind the linear phage DNA near the cos site, wind the DNA into the prehead, and cleaves the DNA at the cos sites. A fully assembled tail structure is then attached to the prehead resulting in a mature and infectious phage particle. The phage then induces host cell lysis releasing infectious phage

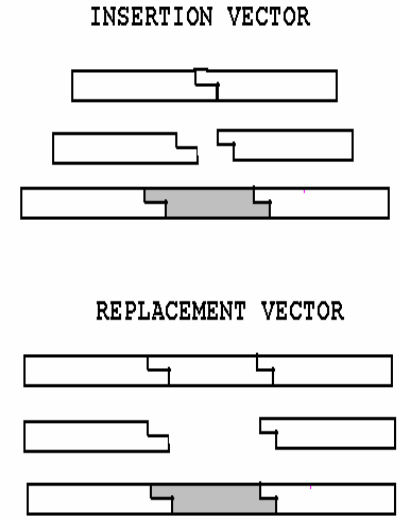
particles which are able to infect more *E. coli*.



The infection and multiplication of phage λ is known as the lytic cycle. Alternatively, λ can also undergo a lysogenic cycle in which λ incorporates into the host genome at a specific site and is replicated with the host. At a later time, the phage genome can excise from host genome and undergo lytic cycle producing infectious phage particles.

Lambda as a Cloning Vector.

Two features of λ make it a powerful cloning vector: 1) foreign DNA can be inserted into the λ genome, and 2) infectious phage particles can be assembled *in vitro*. Phage λ has been extensively modified to make it a more efficient cloning vector and many different λ vectors have been developed. Each particular λ vector has an optimal host and sometimes different applications require different host bacteria. It is important to choose the right host strain for each particular λ phage and the application being carried out.



The two basic types of λ vectors are **insertion** and **replacement** vectors (Figure). The λ genome is approximately 50 kb. Phage assembly can occur if the genome is 78-105% of the normal size. In other words infectious phage particles can be assembled if the λ genome is 40-52 kb. Insertion vectors have a small non-essential central region of the λ genome removed. This removed portion of the genome allows for the insertion of up to 10 kb of foreign DNA (depending upon the particular λ vector). In replacement vectors a 13 kb fragment, called the stuffer fragment, is excised with restriction enzymes and discarded. Only genes necessary for the lysogenic cycle are on the stuffer fragment and the lytic cycle is not affected. The stuffer fragment is replaced with a foreign DNA insert of 11-21 kb. The insertion vectors are easier to work with but accommodate less insert DNA than replacement vectors.

The overall strategy of cloning into lambda is similar to that of cloning into other vectors (Box). The λ DNA is isolated and digested with the appropriate restriction enzyme(s). In the case of replacement vectors the left and right arms will be purified away from the stuffer fragment. The vector can also be dephosphorylated to minimize vector self-ligation (as

discussed for plasmids) when identical overhangs. Generally vectors which have already been digested with restriction enzymes and dephosphorylated (i.e., ready for cloning) are obtained commercially.

Cloning into λ

1. Prepare vector DNA.
2. Prepare foreign DNA.
3. Ligate vector and foreign DNA.
4. *In vitro* packaging.
5. Titer and amplify(?)
6. Infect host *E. coli*.
7. Screen plaques.
8. Plaque purification.
9. Subclone fragment

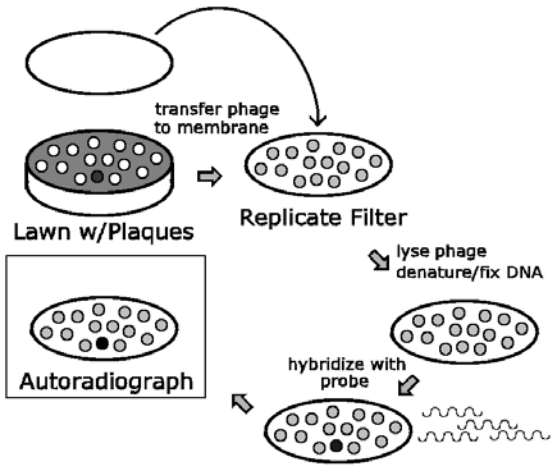
The vector and foreign DNA (digested with the appropriate restriction enzyme) are mixed and ligated. Generally, equal molar ratios of vector and foreign DNA work best. Ligation results in the foreign DNA being combined with the vector. In addition, the cos ends of phage λ are ligated together forming very long DNA molecules of sequential recombinant phage. The recombined concatenated phage DNA is then mixed with an extract containing all of the proteins needed for phage assembly. This *in vitro* packaging extract is prepared from *E. coli* which have been infected with a defective λ and usually obtained from a commercial source. One of the mutations in the defective lambda reduces the packaging of the endogenous phage DNA, but does not affect the synthesis of the phage coat proteins. Therefore, the recombined phage DNA is preferentially packaged into phage particles when mixed with extracts prepared from *E. coli* infected with this lambda mutant. Nucleases on the phage head recognize and cleave the DNA at the cos sites resulting in the in vitro assembly of an infectious phage particle. The phage particles will contain different insert DNA fragments and this mixture of recombinant phage is called a library.

The next step is to titer the recombinant library and determine the plaque forming units (pfu) per ml. **Plaques** are clear zones on a bacteria lawn which are the result of localized bacteria lysis. The bacteria lawn is formed by mixing the phage with *E. coli* in a low concentration of agar or agarose. This mixture is then poured onto an agar nutrient plate and allowed to solidify. As the bacteria grow the plate will become opaque. Infected bacteria will result in a clear plaque due to the destruction of neighboring bacteria following many cycles of lysis and reinfection.

It is sometimes desirable to amplify the library. Amplifying will increase the total volume of the library allowing for multiple screenings with different probes. Amplification is accomplished by subjecting a portion of the library to several rounds of replication before screening the library. Libraries can be amplified by either a plate lysate method or a liquid culture method. Both methods consist of infecting the host bacteria with the recombinant phage and allowing the phage to replicate until the bacteria are completely lysed. In both methods the ratio of phage to bacteria, called the multiplicity of infection (m.o.i.), has a large impact on the final yield of phage (i.e., titer). In the plate lysate method a bacterial lawn is prepared and buffer is then added to the plate following the complete lysis of the host bacteria to recover the phage particles. One drawback of amplification is that the relative frequencies of the different recombinants may change since slowly replicating recombinants will be under represented in the amplified library. λ libraries are stable for years if stored at 4^{o}C in the presence of chloroform.

Screening λ libraries

The library is plated out on the appropriate strain of *E. coli* and screened. Recombinant clones of interest are identified with DNA probes in a plaque lift. When plaques appear, the plates are overlaid with a nylon membrane. The phage in the plaques stick to the membranes. In general, plaque lifts are similar to colony lifts (Figure) except that it is not necessary to continue growth after transfer to the membranes. Phage particles are lysed directly on the membrane and the recombinant phage DNA is denatured and fixed to the membrane. The membranes are then hybridized with a specific DNA probe. The recombinant phage with the insert DNA of interest is then identified by aligning the autoradiograph with the original plate.



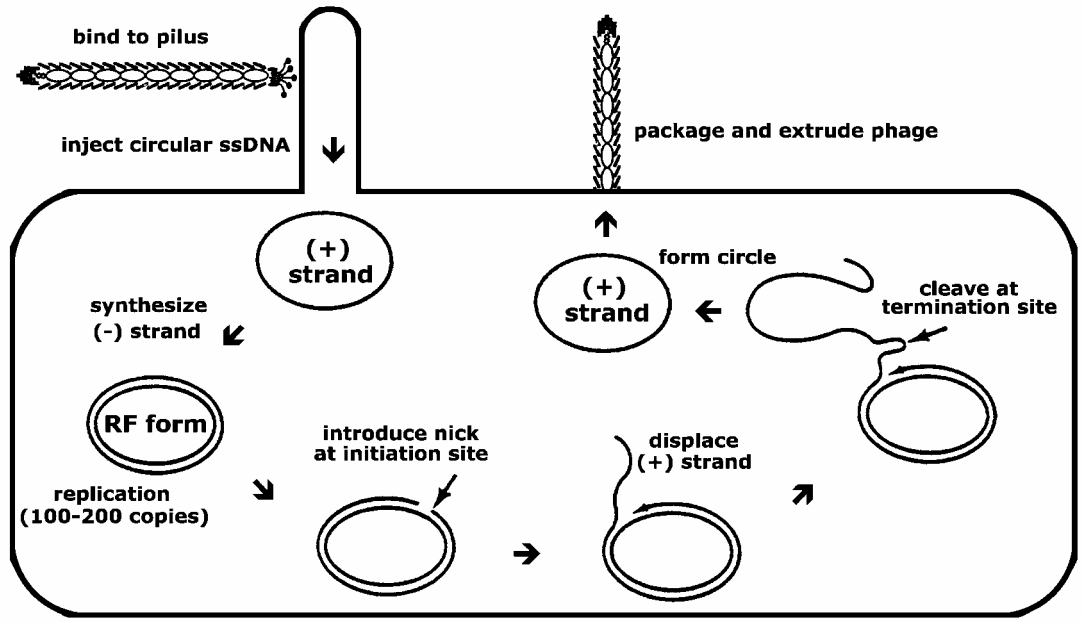
The recombinant phage of interest are recovered by punching out the plaque of interest from the agar plate with a Pastuer pipet. The phage are eluted from the agar and can be expanded by reinfecting *E. coli*. A 'plaque purification' can be carried out by replating and rescreening the recombinant phage. It is recommended to repeat this plaque purification until all of the plaques are positive for the insert DNA of interest. The cloned phage are then expanded and the DNA isolated.

The high efficiency of λ infection and the ease at which relatively large numbers of plaques can be screened make λ a good vector for cloning genes. However, isolating DNA from λ is somewhat cumbersome compared to isolating plasmid DNA. Therefore, the recombinant DNA insert is usually subcloned into a plasmid for further characterization. This is easily accomplished by digesting the phage DNA with the appropriate restriction enzyme(s) and ligating with plasmid previously digested with the same enzyme(s). It is usually not necessary to purify the insert fragment since the phage arms are not efficiently incorporated into the plasmid or transfected into the host bacteria.

FILAMENTOUS BACTERIOPHAGE

Applications of ssDNA phage
- DNA sequencing
- site-direct mutagenesis
- phage display
- in vivo excision

Single-stranded DNA filamentous phage exhibit features that can be exploited for molecular biology research. For example, M13 has been widely used for the generation of ssDNA for DNA sequencing and site-directed mutagenesis. More recently, though, sequencing of dsDNA has replaced ssDNA and many site-directed mutagenesis applications now utilize PCR. Filamentous phage are also used for 'display' of ligands or epitopes (known as phage display).

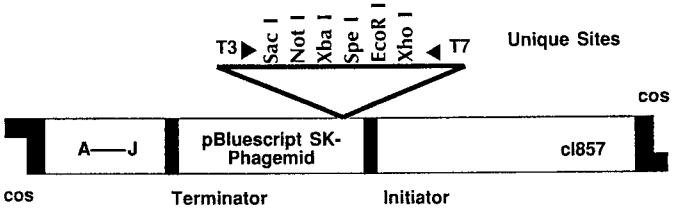


Life cycle. Filamentous phage attach to the F-pili of *E. coli* and inject single-stranded circular DNA into the bacteria. The ssDNA is converted to a dsDNA form known as replicative form (RF). The replicative form behaves as a plasmid within the bacteria and undergoes multiple rounds of replication. It is this replicative form that is isolated and manipulated as a cloning vector. Foreign DNA can be incorporated into cloning sites and bacteria transformed by standard procedures. Structural genes are transcribed from the replicative form and the ssDNA genome is produced from a special replication origin (usually called the f1 origin). The product of gene II forms a nick in the initiation site of the f1 replication origin. Host DNA polymerases replicate the phage DNA starting at the nick and displace one strand will progressing along the template. When the polymerase completes the circle, the gene II product cleaves the displaced strand at the termination site within the replication origin and the ssDNA is circularized.

Cloning in M13
- isolate RF form and manipulate as plasmid
- transfect *E. coli*
- differential centrifugation
  - bacteria pellet = RF form
  - phage supernatant = ssDNA

The ssDNA is packaged at the membrane of the bacteria and the mature phage is extruded without lysing the cell. Phage particles containing ssDNA are isolated from the culture supernatant medium by differential centrifugation and the ssDNA is extracted from the phage particles. The phage can also be plated on a bacterial lawn to produce plaques. However, these plaques are 'fuzzy' since they are not formed as a result of cell lysis, but formed as a result of diminished host cell growth.

Phagemids. Many plasmids have been modified to include a f1 origin of replication. These 'phagemids' allow for the generation of ssDNA without some of the difficulties of working with M13. In particular, incorporation of foreign DNA into M13 will often severely affect phage replication. Foreign DNA is incorporated into these phagemids and propagated as conventional plasmids. Single-stranded DNA is rescued by super-infecting the transformed cells with a helper phage (eg., R408, M13KO7). The helper phage supplies enzymes (i.e., gene II) that are necessary for synthesis of circular ssDNA molecules and the genes for the phage coat proteins. The helper phage, however, has a defect in its own f1 origin so that the f1 origin of the recombinant phagemid is the preferred substrate. The ssDNA is then packaged into the phage particles and secreted into the culture media with some contamination by the helper phage due to its slow replication. The exact yield of the ssDNA is dependent on the nature of the insert DNA.



λZAP is another example of a vector exploiting the f1origin. In this case, a phagemid (pBluescript) is contained within the λDNA such that initiation and termination regions of the f1 origin are at opposite ends (Figure). Co-infection with a helper phage leads to the displacement of a ssDNA molecule from the initiation site followed by cleavage and circularization at the termination site. The ssDNA is then packed into phage particles and extruded from the bacteria. The plasmid containing the foreign DNA is obtained by infecting F' strains of *E. coli* with these phage particles isolated from the culture supernatant. This process, called 'in vivo excision', eliminates the subcloning step.

Phage display. Filamentous phage have also been used for expression vectors. Peptide or protein sequences are expressed as fusion proteins with a coat protein of the bacteriophage, resulting in the display of the fused protein on the surface of the phage. Fusions to the N-termini of gp3 and gp8 have minimal effects of phage infectivity. Two common applications for this phage display are peptide libraries and single-chain antibodies. Screening in both applications is carried out by 'biopanning'.

Peptide libraries can be used for epitope mapping, analysis of protein-protein interactions, and the isolation of inhibitors or other ligands. The libraries are prepared by cloning oligonucleotide sequences, corresponding to random peptide sequences, into the N-terminal portion of either gp3 or gp8. The gp8 site can only accommodate peptides of approximately 6 residues.

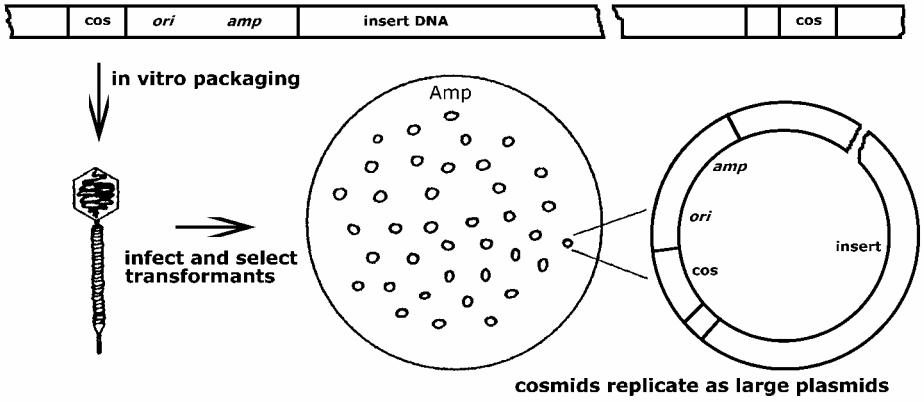
Single chain antibodies, or scFv, are produced by combining the heavy chain variable region with the light chain variable region via a linker. This artificial antibody protein can fold such that the light and heavy chain domains come together to form a functional binding site. Combinatorial antibody libraries are generated from naive B-cells and cloned into the gp3 site.

In both cases the desired recombinant phage are isolated by a biopanning process. The target protein (eg., antibody for peptide libraries or antigen for scFv libraries) are absorbed to a solid support. This target is incubated with the phage display library and non-bound phage are wash away. The bound phage are eluted and amplified, thereby enriching for the phage which express peptides or proteins that bind the target. This enrichment is typically 100-1000 fold, thus 3-4 rounds of biopanning are typically carried out. (In general, the maximum complexity of the library ranges from 10^8-10^9.) The individual phage in the final biopanning are then analyzed for the peptide sequences (eg., epitope) recognized by target protein, or the scFv are isolated and used as a reagent for subsequent analyses.

COSMIDS AND YACS

One problem with cloning vectors is their limited size as compared to the genome size of many organisms. For example, the human genome contains more than 10^9 bp. Typical λ vectors will accommodate only 10-20 kb of DNA. Similarly plasmids, although in theory can accommodate an unlimited amount of DNA are rarely functional above 10 kb. Cosmids and YACS are two additional cloning vectors that increase the size of insert DNA that can be incorporated.

Cosmids are plasmids that contain the cos ends from λ, an *E. coli* origin of replication and a selectable marker (eg., ampicillin resistance gene). Foreign DNA with a size range of 35-45 kb is ligated with the cosmid resulting in long concatemers of DNA. The DNA is then packaged into λ heads and used to infect *E. coli*. Even in the absence of λ genes the DNA will be packaged into phage heads if the cos sites are appropriately spaced. Once inside of the bacteria the injected DNA will form circles and replicate as a large plasmid under antibiotic pressure.



Yeast artificial chromosomes (YACS) are another means to clone large fragments of DNA. YACS are linear pieces of DNA that replicate as chromosomes within yeast cells. The vector contains elements that allow for its replication in both bacteria and yeast (Box). The centromere, autonomously replicating sequences (ARS), and telomere sequences are particularly important for its function as a linear chromosome. The bacterial elements allow for the production of the plasmid in *E. coli*. Foreign DNA (100 kb-2 mb) is then incorporated so that it is a linear piece of DNA and is then introduced into yeast cells.

YAC Vector
- *E. coli* origin
- antibiotic resistance
- yeast centromere and ARS
- ciliate telomere
- yeast selectable marker

QUALITY CONTROL

The success of molecular cloning often depends upon the ability to screen large numbers of potential recombinants. Generally the screening step is the most labor intensive of the entire process. Therefore, it is often desirable to evaluate the quality of a DNA library before screening. For example, in the case of cDNA libraries the average insert size is determined by selecting a few plaques or colonies at random and determining the size of the insert DNA. This information gives some indication of the amount of full length cDNA present in the library. Another measurement of library quality is the percentage of non-recombinants. Many vectors have been modified so that recombinants and non-recombinants can be easily distinguished (see below).

In addition, the ability to distinguished recombinants from non-recombinants eliminates the need to screen in many subcloning procedures. For example, cloning PCR fragments or other highly enriched DNA fragments often does not require screening if recombinants are easily identified.

Distinguishing recombinant from non-recombinant clones. One of the first methods for distinguishing recombinants was through the loss of antibiotic resistance in pBR322. pBR322 contains both an ampicillin resistance gene and a tetracycline resistance gene. The foreign DNA is cloned into either the ampicillin resistance gene or the tetracycline resistance gene. Replicate plates are made on either ampicillin or tetracycline. Clones that are resistance to both antibiotics are non-recombinants, whereas recombinants will only be resistant to one of the antibiotics.

One of the simplest examples of distinguishing non-recombinants is illustrated by λgt10. Non-recombinant λgt10 in the proper hosts (eg., HflA) undergo lysogeny and do not produce plaques. Therefore, all plaques are recombinant and non-recombinants are excluded from the screening process.

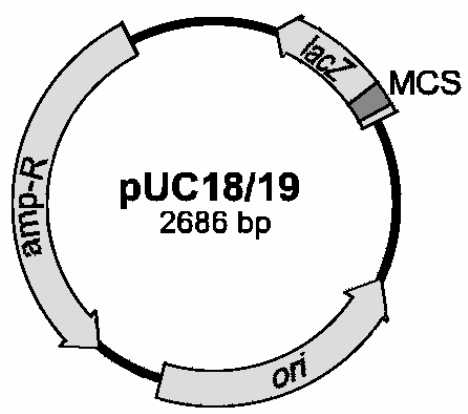
The most common method used to distinguish non-recombinants from recombinants is through the interruption of the β-galactosidase gene. Cloning of foreign DNA into this gene will result in the loss of its activity. The bacteria or phage are grown on nutrients plates which contain a β-galactosidase substrate which turns blue in the presence of enzyme. Colonies or

plaques from non-recombinants will turn blue in the presence of 5-bromo-4-chloro-3-indoly-β-D-galactoside (X-gal), whereas recombinants will remain colorless. This process is often referred to as blue/white color selection. Some vectors (eg., λgt11) utilize the entire β-galactosidase gene whereas others (eg., pUC series) only utilize a portion of the β-galactosidase gene (lacZ in Figure). The latter is called **α-complementation** (Box).

Identifying Recombinants	
λgt10	non-recombinants undergo lysogeny (=no plaques)
pBR322	loss of antibiotic resistance in recombinant
λgt11	interruption of β-galactosidase gene in recombinants
pUC__	loss of α-complementation in recombinants (Box)

Blue/White Color Screening (α-complementation)

- Host expresses C-terminal portion of β-galactosidase gene on F-factor.
- Vector expresses N-terminal portion of gene interrupted by MCS (Figure).
- If both are present, then β-gal is active and substrate (X-gal) is cleaved to produce blue color.
- If vector contains insert, then β-gal is inactive and colonies (or plaques) remain colorless.



CHAPTER 22--EXPRESSION OF RECOMBINANT PROTEINS

One potential of recombinant DNA is the production of proteins. Many systems have been designed for the expression of recombinant proteins. Although *E. coli* are easy to grow in large quantities there are some problems associated with the expression of eukaryotic proteins in prokaryotes. Therefore, expression systems in eukaryotic systems have also been devised. In addition, it is also possible to use expression vectors for the cloning of genes.

Over Production of Recombinant Proteins

- strong promoter
- regulatable promoter
- gene dosage
- localization signals
- protease defective hosts
- fusion proteins

General considerations in the over production of recombinant proteins (Box) include such features as promoters and the nature of the host for expression. Ideally the recombinant protein should be cloned in conjunction with a strong promoter. Promoters are elements found on the 5'-end of genes and control their expression. A strong promoter results in a high level of gene expression. In addition to being strong, promoters should be regulatable so that transcription can be turned off and on. Recombinant proteins are sometimes toxic for the host cell and this toxicity can be minimized by controlling the expression. The use of high copy number plasmids also increases the production of recombinant proteins by increasing the gene dosage.

Other features are included in many expression vectors. For example, it is often convenient to have recombinant proteins that are exported from the host cell. This can be accomplished by engineering localization signals within the recombinant protein so that they are directed to a particular cellular compartment. In addition, the recombinant protein can also be expressed as a fusion protein with another protein. This will sometimes stabilize the recombinant protein and/or assist in the purification and characterization of the recombinant protein. Proteolysis of recombinant proteins is often a major problem that is partially alleviated by cloning into protease deficient hosts.

EXPRESSION IN *E. COLI*.

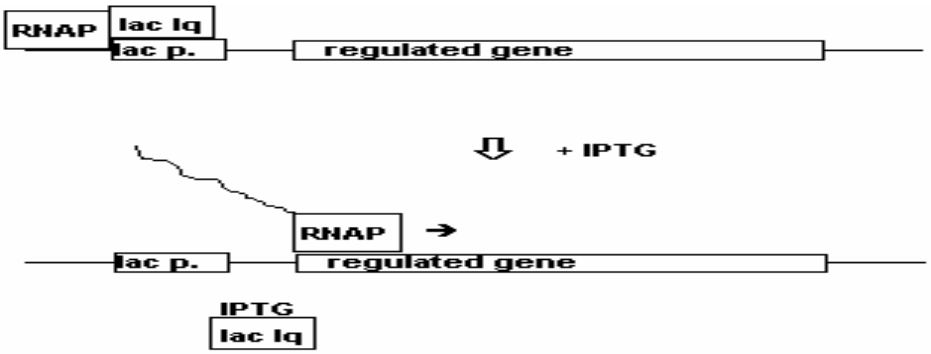
Common Bacterial Promoters

Promoter	Repressor	Induction
λP_L	cI857 (ts)	$32^o \rightarrow 42^o$
lac	$lacI^q$	IPTG
tac	$lacI^q$	IPTG
T7	-	T7 gene 1

The general procedure for expression of cloned genes in *E. coli* involves the insertion of the coding region of interest into a vector, usually a plasmid, so the region is efficiently transcribed and translated. Since eukaryotic genes do not contain the proper signals for transcription initiation, ribosome recognition, translation initiation, and translation termination, these signals need to be supplied by the vector.

Common promoters used in bacterial expression vectors are: λP_L, lac, tac and T7 (Table). Of these, the lac and tac promoters are the most widely used. The control elements that regulate expression from these promoters are supplied by either the host or the vector. The λP_L promoter is controlled by a mutant cI repressor protein that is non-functional at 42^o. At

temperatures less that 42^{o} the repressor protein (cI) binds to the promoter and prevents expression. To induce expression the temperature is raised to that the repressor becomes non-functional and expression is now permitted. One problem with this system is that heat-shock proteins may also be induced. The T7 promoter is from the T7 phage and is only transcribed by the T7 RNA polymerase (T7 gene 1). The host cell must also contain the T7 gene 1 in order to express from this promoter. To regulate expression from this promoter it is necessary to be able to regulate the expression of the T7 gene 1.



Promoters and Repressors

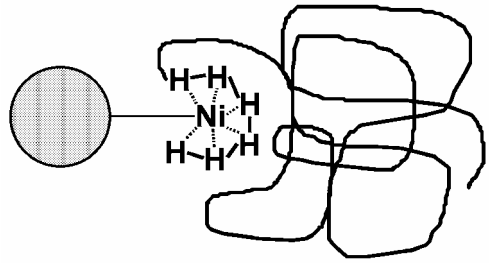
The **lac** and **tac** promoters are controlled by the lac repressor (Figure). The lac repressor binds to the lac promoter and prevents RNA polymerase from transcribing the gene. Host strains with lacI^{q} gene express the lac repressor at 10-fold higher concentrations than the normal lacI gene. IPTG (isopropyl-1-thio-β-D-galactopyanoside, an analog of lactose) binds to the repressor and prevents its interaction with the lac promoter and allowing RNA polymerase to transcribe the regulated gene. The tac promoter is a fusion of trp and lac promoters and is also regulated by IPTG.

Fusion Protein	Affinity Matrix
Glutathione-S-Transferase	glutathione-agarose
Thioredoxin	phenylarsine oxide-agarose
Maltose Binding Protein	amylose-agarose
Six Histidine Residues	Ni-agarose

The production of recombinant **fusion proteins** often stabilizes the expression of foreign proteins in *E. coli*. Several plasmids that express recombinant proteins as fusion proteins have been developed (Table). In addition to stabilizing the recombinant protein, the fusion partner is often exploited for affinity purification or for the analysis of the recombinant protein (see Appendix). However, for many applications the fusion partner may interfere with the activity of the recombinant protein. Some expression vectors include a protease site between the fusion partner and the multiple cloning site (MCS). The engineered protease site allows for the removal of the fusion partner after its purification.

Other expression vectors will only include a small fusion partner, such as epitope tagging and His_6 vectors. In both of these examples, only a few additional amino acids are added at either end of the recombinant protein. In general, these few residues will not interfere with the normal activity of the recombinant protein. Affinity columns prepared from mAbs

against the epitope can be used for the purification of the fusion protein in the case of epitope tagging. Similarly the six consecutive histidine residues added by His_6 vectors bind tightly to metals and are easily purified on metal-chelating columns.



Metal Chelate Chromatography

Recombinant proteins expressed at high levels will sometimes form insoluble aggregates known a **inclusion bodies**. In some applications it is possible to take advantage of this phenomenon. For example, the inclusion bodies can be isolated by differential centrifugation and solubilized under denaturing conditions (eg., urea). It is sometimes possible to renature the protein and regain activity. In addition, fusions with *E. coli* thioredoxin can circumvent inclusion body formation.

PREPARATION OF EXPRESSION VECTORS.

The 'insert' DNA must be subcloned into the expression vector so that it is in frame with the fusion protein (or a start ATG). Most expression vectors will have a MCS with several different restriction sites. In addition, many expression vectors are designed so that variants with all three reading frames are available. Therefore, it is generally simple to choose restriction enzymes that will result in a continuous open reading frame (ORF) between the fusion protein and the foreign protein. It is also possible to shift the reading frame by cutting with restriction enzymes, filling in with Klenow, and religating (see Appendix). Digesting with an enzyme producing a 4-base overhang will result in a -1 frameshift and digesting with an enzyme producing a 2-base overhang will produce a +1 shift. In addition, stop codons can be produced by this method with certain restriction enzymes (eg., *Hin*dIII, *Spe*I).

SCREENING λ LIBRARIES WITH ANTIBODIES.

It is also possible to utilize the expression vectors as a means to screen for the gene of interest. Recombinant DNA clones expressing proteins of interest can be detected with antibodies, ligands or other protein activity. λgt11 is a common vector for cloning genes using antibody probes. Foreign DNA is cloned into a lac-promoter controlled β-galactosidase gene resulting in the expression of fusion proteins. Recombinant phage expressing the protein of interest are detected with antibodies (Box).

Screening λgt11 expression libraries
1. Plate phage on Y1090 lawn.
2. Grow at 42^o until plaques appear.
3. Overlay with IPTG saturated membrane.
4. Continue incubation 1-3 hr at 37^o.
5. Analyze plaque lift by Western blot.

A recombinant λgt11 library is plated on a lawn of *E. coli* Y1090 and incubated at 42^o until plaques are visible. An IPTG saturated nitrocellulose membrane is then laid on the plate and the incubation is continued for 1-3 hr at 37^o. This will induce the expression of β-galactosidase fusion proteins within infected bacteria. Upon lysis the recombinant proteins are released with other bacterial proteins into the plaque and adsorbed onto the membrane. The

membrane (i.e., plaque lift) is then analyzed by Western blotting using an antibody against the protein of interest. This will result in a spot on membrane which corresponds to a plaque formed by the recombinant λgt11 expressing the protein of interest. Such plaques are identified by aligning the membrane with the original plate. Rabbit antisera are often problematic due to a strong background reactivity with *E. coli* proteins. It is sometimes possible to pre-absorb the anti-sera with *E. coli* proteins to reduce the background.

It is also possible to use a cloned recombinant λgt11 phage for the production of recombinant protein. However, this is generally more difficult than subcloning the fragment into a plasmid and therefore not widely used. The procedure involves the production of lysogens. The lysogens are induced to produce high levels of phage from which the fusion protein is expressed. The fusion protein is then isolated from the infected bacteria.

EXPRESSION IN EUKARYOTES

Shuttle Vectors
- prokaryotic replication origin
- *E. coli* selectable marker
- eukaryotic replication origin
- eukaryotic promoters/enhancers
- polyadenylation signals
- eukaryotic selectable marker

Although expression of recombinant proteins in *E. coli* is usually fairly straight forward, it is often desirable or necessary to express cloned genes in eukaryotes. Eukaryotic expression systems are often needed to insure correct folding and disulfide-bond formation, post-translational modifications, and processing. Yeast, such as *Saccharomyces cerevisiae* and *Pichia* species, are useful hosts for the expression of recombinant proteins since they can be grown and manipulated like a bacteria. *Pichia* expression systems often allow for high level expression of recombinant proteins using a strong alcohol oxidase promoter which is induced by methanol. **Shuttle vectors** contain both a bacterial origin of replication and the origin of replication of interest (Box). This allows for manipulations of the plasmids to be carried out in *E. coli* before transforming yeast or other eukaryotes.

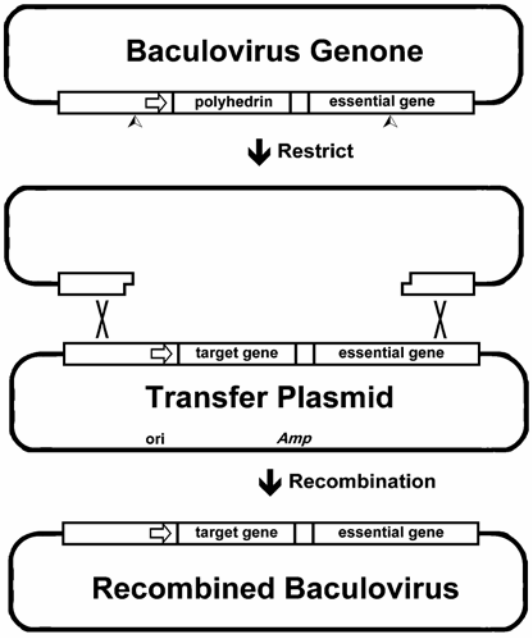
- calcium phosphate
- DEAE-dextran
- electroporation
- liposomes
- protoplast fusion
- ballistics (gene gun)
- microinjection

The two major strategies for expression of recombinant sequences in mammalian cells are 1) stable or transient expression of transfected DNA and 2) the use of viral expression vectors. Mammalian and other eukaryotic cells are able to take up DNA. The introduction of DNA to cells is often cell type dependent and problematic, though. Several different methods for introducing DNA to cells are available (box). One of the early methods for transfecting eukaryotic cells was to incubate a calcium phosphate precipitate of the DNA with the target cells. This method is relatively inefficient. Using DEAE-dextran instead of calcium phosphate is a little more efficient. Electroporation is the application of a short pulse of high voltage. This method is effective for many cell types once the optimal conditions are determined. DNA can also be incoporated into liposomes prepared from cationic detergents. These liposomes (i.e., micelles) will fuse with the plasma membrane and deliver the DNA. Similarly, protoplasts can be prepared by digesting the bacterial cell wall and fusing these protoplasts with the target cells. Finally, DNA can be directly introduced into cells by ballistics (i.e, gene gun) or microinjection. The gene gun shoots small colloidal gold particles covered with DNA into cells with doing permanent membrane

damage. In many cases the foreign DNA is quickly loss from the host cell and the expression is only transient.

Several different eukaryotic viruses can also be used as cloning vectors for the expression of recombinant proteins. Lytic viruses are good for transient expression whereas episomal viruses are better for constitutive expression. Retroviruses can become incorporated into the host cell genome and possibly lead to a stable transformation.

Baculoviruses are used for the production of recombinant eukaryotic proteins. *Autographica californica* is a nuclear polyhedrosis virus (AcNPV) that infects insects. Sf9 cells, derived from *Spodoptera frugiperda*, are readily grown in vitro and can be infected with baculovirus. Extremely high levels of recombinant proteins can be expressed and many eukaryotic post-translational modifications are correctly made in the host insect cells. This high level of expression is driven by a strong promoter for the polyhedrin gene. The polyhedrin protein is expressed late in infection as the virus is killing the host cell and is needed for the dissemination of the virus in nature. However, the polyhedrin gene is unnecessary for viral growth in tissue culture and can be replaced with a gene of interest.



Baculovirus expression vectors consist of the viral genome (≈130 kb) and a transfer plasmid vector which contains the regions flanking the polyhedrin gene. The target gene is inserted into the plasmid vector and host cells are co-transfected with the parental baculovirus and the engineered plasmid. Recombination between the viral genome and the plasmid results in replacement of the polyhedrin gene with the target gene. The efficiency of identifying recombinants is greatly enhance by digesting the baculovirus DNA with a restriction enzyme which disrupts an essential gene and using a transfer vector which contains this essential gene (Figure). Recombinants between the transfer vector and the virus will have a selective advantage and produce plaques. The recombinant virus can be isolated from plaques and then amplified to produce large amounts of protein.

APPENDIX 1. CODES AND CODONS

A
C
G
T or U (DNA or RNA)

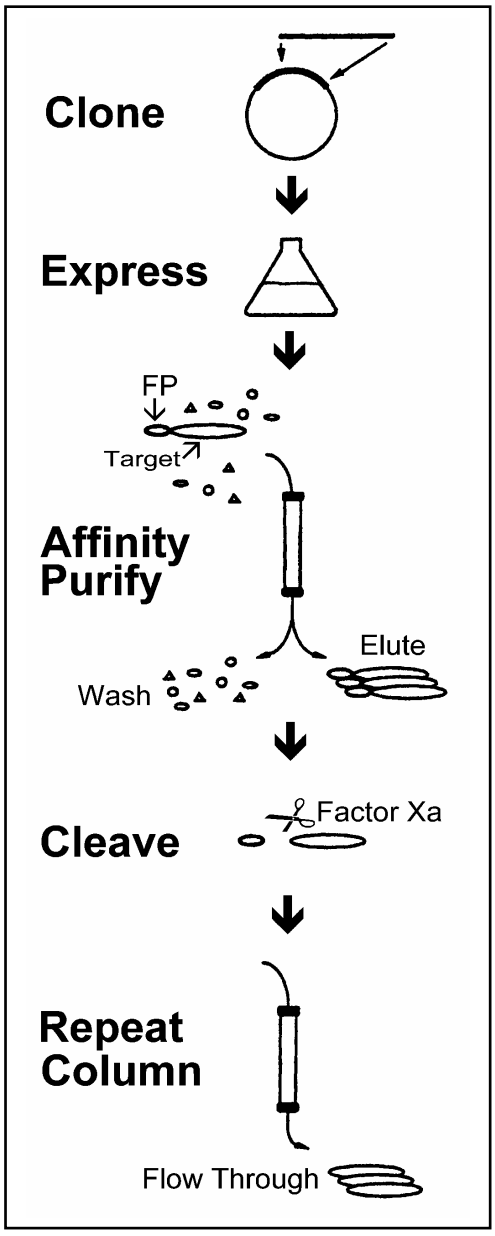
M	A or C (methyl)
R	A or G (purine)
W	A or T (weak)
S	G or C (strong)
Y	C or T (pyrimidine)
K	G or T (keto)
V	A or C or G (not T = V)
H	A or C or T (not G = H)
D	A or G or T (not C = D)
B	C or G or T (not A = B)
N	A or C or G or T (any nucleotide)

A	GCN
C	TGY
D	GAY
E	GAR
F	TTY
G	GGN
H	CAY
I	ATH
K	AAR
L	CTN
M	ATG
N	AAY
P	CCN
Q	CAR
R	CGN or AGR
S	TCN or AGY
T	ACN
V	GTN
W	TGG
Y	TAY
*	TAR or TGA

Second Codon Position

First Codon Position	U		C		A		G		Third Codon Position
U	UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys	**U**
	UUC	Phe	UCC	Ser	UAC	Tyr	UGC	Cys	**C**
	UUA	Leu	UCA	Ser	UAA	Stop	UGA	Stop	**A**
	UUG	Leu	UCG	Ser	UAG	Stop	UGG	Trp	**G**
C	CUU	Leu	CCU	Pro	CAU	His	CGU	Arg	**U**
	CUC	Leu	CCC	Pro	CAC	His	CGC	Arg	**C**
	CUA	Leu	CCA	Pro	CAA	Gln	CGA	Arg	**A**
	CUG	Leu	CCG	Pro	CAG	Gln	CGG	Arg	**G**
A	AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser	**U**
	AUC	Ile	ACC	Thr	AAC	Asn	AGC	Ser	**C**
	AUA	Ile	ACA	Thr	AAA	Lys	AGA	Arg	**A**
	AUG	Met	ACG	Thr	AAG	Lys	AGG	Arg	**G**
G	GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly	**U**
	GUC	Val	GCC	Ala	GAC	Asp	GGC	Gly	**C**
	GUA	Val	GCA	Ala	GAA	Glu	GGA	Gly	**A**
	GUG	Val	GCG	Ala	GAG	Glu	GGG	Gly	**G**

APPENDIX 2. FUSION PROTEINS



The gene of interest is cloned into the expression vector leading to the creation of a gene fusion between the target gene and the fusion partner (FP).

Transformed *E. coli* are grown in large-scale liquid cultures and after reaching the appropriate density are induced to express the recombinant protein.

A bacterial lysate is passed over an affinity column made from a ligand binding the fusion partner. The recombinant fusion protein is retained by the column and other *E. coli* proteins are removed. The fusion protein is eluted with free ligand or by other means.

Many expression vectors have a specific protease site (eg., factor Xa) engineered between the fusion partner and the protein of interest. Treatment with the specific protease may result in separation of the two proteins.

The fusion partner can be separated from the the protein of interest by repeating the affinity chromatography since the fusion partner will be retained by the column.

APPENDIX 3. GENERATING FRAMESHIFTS

Generation of a -1 Frameshift

```
 | SalI |    | ClaI || HIII |
GGT CGA CGG TAT CGA TAA GCT TGA
CCA GCT GCC ATA GCT ATT CGA ACT
Gly Arg Arg Tyr Arg *** Ala ***
          ⇓ Sal I
GG        T CGA CGG TAT CGA TAA GCT TGA
CCA GCT         GCC ATA GCT ATT CGA ACT
          ⇓ + Klenow + dNTPs
GGT CGA   T CGA CGG TAT CGA TAA GCT TGA
CCA GCT   A GCT GCC ATA GCT ATT CGA ACT
          ⇓ + DNA ligase
GGT CGA TCG ACG GTA TCG ATA AGC TTG A
CCA GCT AGC TGC CAT AGC TAT TCG AAC T
Gly Arg Ser Thr Val Ser Ile Ser Leu
```

Generation of a +1 Frameshift

```
 | SalI |    | ClaI || HIII |
GGT CGA CGG TAT CGA TAA GCT TGA
CCA GCT GCC ATA GCT ATT CGA ACT
Gly Arg Arg Tyr Arg *** Ala ***
          ⇓ Cla I
GGT CGA CGG TAT       CGA TAA GCT TGA
CCA GCT GCC ATA GC      T ATT CGA ACT
          ⇓ + Klenow + dNTPs
GGT CGA CGG TAT CG    CGA TAA GCT TGA
CCA GCT GCC ATA GC    GCT ATT CGA ACT
          ⇓ + DNA ligase
GGT CGA CGG TAT CGC GAT AAG CTT GA
CCA GCT GCC ATA GCG CTA TTC GAA CT
Gly Arg Arg Tyr Arg Asp Lys Leu
```

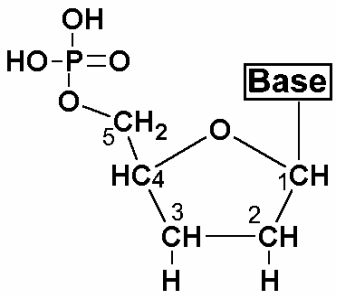
Generation of a Stop Codon

```
              | HIII |
AAT TCG ATA TCA AGC TTA TCG
TTA AGC TAT AGT TCG AAT AGC
Ile Ser Ile Ser Ser Leu Ser
          ⇓ Hind III
AAT TCG ATA TCA        AGC TTA TCG
TTA AGC TAT AGT TCG A       AT AGC
          ⇓ + Klenow + dNTPs
AAT TCG ATA TCA AGC T    AGC TTA TCG
TTA AGC TAT AGT TCG A    TCG AAT AGC
          ⇓ + DNA ligase
AAT TCG ATA TCA AGC TAG CTT ATC GAT
TTA AGC TAT AGT TCG ATC GAA TAG CTA
Ile Ser Ile Ser Ser ***
```

CHAPTER 23--DNA SEQUENCING

Simple and reliable DNA sequencing methods have allowed for a relatively rapid accumulation of gene sequence data. Two methods for sequencing DNA are the dideoxy chain termination method developed by Sanger and the chemical cleavage method developed by Maxim and Gilbert. The dideoxy procedure is the less laborious and the most widespread. The chemical cleavage procedure is primarily used to sequence synthetic oligonucleotides.

DIDEOXY CHAIN TERMINATION



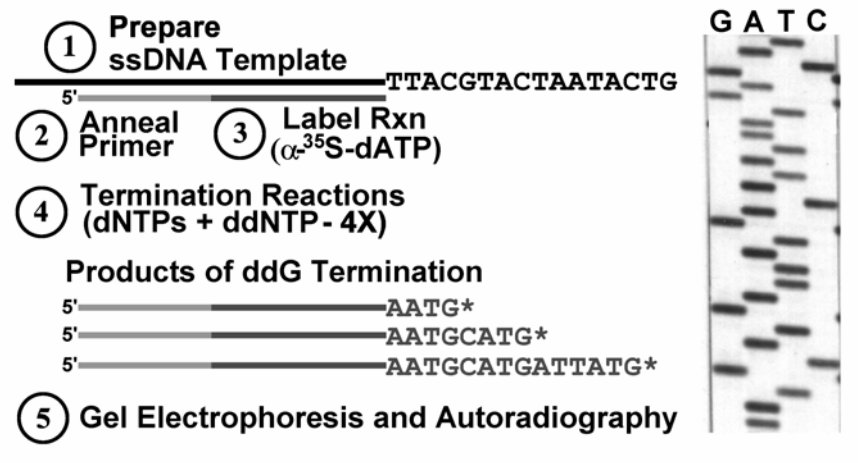
The dideoxy chain termination method is based upon the use of 2',3'-dideoxynucleotides (figure) as chain terminators. These ddNTPs will be randomly incorporated into the growing poly-nucleotide chain if DNA synthesis is carried out in a mixture of dNTPs and ddNTPs. Dideoxynucleotides lack the 3'-OH and, therefore, additional nucleotides cannot be added during template mediated DNA synthesis. Since the incorporation of the ddNTPs is random the results of the DNA synthesis will be a mixture of newly synthesized DNA fragments of various lengths with a dideoxy-nucleotide at the 3'-end. This mixture is then analyzed by gel electrophoresis under conditions which will resolve DNA fragments that differ in length by single nucleotides. The sequence is then determined from the dideoxy-nucleotide which is located at the 3'-end of progressively longer oligonucleotides.

Dideoxy chain termination requires ssDNA as a template. The ssDNA can either be isolated from M13, phagemids, etc., or be prepared by denaturing dsDNA with NaOH. The sequencing is more efficient and less template DNA is needed if ssDNA is used. However, it is more laborious to prepare and isolate ssDNA from filamentous phage, and therefore, plasmid DNA denatured with NaOH is the most widely used template. Following denaturation in NaOH, the ssDNA is precipitated with ethanol and redissolved in an appropriate buffer.

Original Method. The original dideoxy sequencing method is based on radiolabeling the product strand. The first step is to incubate the template ssDNA with a primer under conditions which promote hybridization. Primers are oligonucleotides complementary to the DNA to be sequenced. Common sequencing vectors have well characterized primer sequences that flank the MCS. The annealed primer/DNA is mixed with DNA polymerase, α-^{35}S-dATP, dGTP, dCTP and dTTP and subject to a short **labeling reaction**. The nucleotide concentrations used in this labeling reaction are relatively low so that DNA synthesis proceeds slowly. This results in the synthesis of a short segment (15-30 bases) of radioactive DNA.

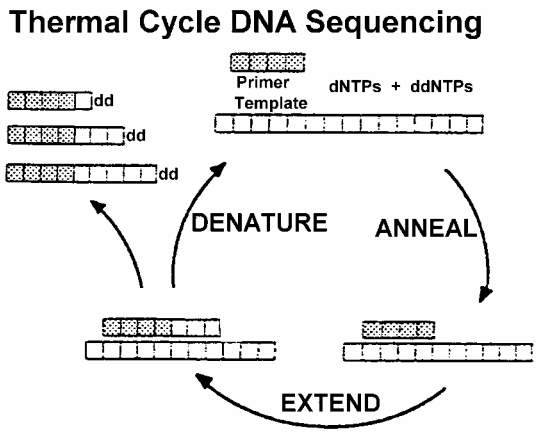
The labeled DNA is then divided into four tubes containing higher concentrations of dNTPs and either ddATP, ddGTP, ddCTP or ddTTP. This step is called the **termination reaction**. The increase in the dNTP concentration result in an increased rate of DNA synthesis. The ddNTPs are at relatively low concentrations and are incorporated at random in the growing DNA molecule. When a ddNTP is incorporated into the growing DNA strand no additional nucleotides can be added resulting in a chain termination. Since ddNTP incorporation is random, the result will be a mixture of nucleotides terminated at all possible positions. The ratio

of ddNTP/dNTP is crucial for the success of DNA sequencing. If the ddNTP is too high then the DNA synthesis will terminate prematurely and only sequence close to the primer will be determined. If the ddNTP concentration is too low then insufficient oligonucleotide chains of appropriate lengths will be produced.



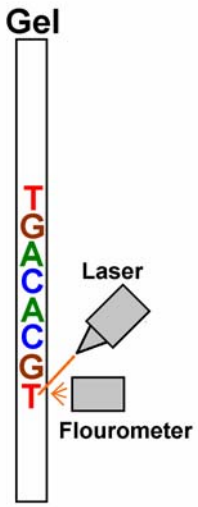
The newly synthesized DNA is denatured by heating and subjected to gel electrophoresis under conditions which separate DNA molecules that differ in size by only a single base. The standard conditions are polyacrylamide gels (6-8%) containing urea. The urea will prevent H-bonding and minimizes the formation of dsDNA and secondary structures. The four termination reactions are electrophoresed side-by-side and DNA strands are detected by autoradiography. The smaller oligonucleotides representing termination events that occurred near the primer will be at the bottom of the gel. Starting at the bottom of the gel the DNA sequence can be read directly from the autoradiograph by noting the lane (i.e., ddNTP termination reaction) of each progressively larger band.

Semi-automated Sequencing. Pouring, running and analyzing sequencing gels are laborious procedures. Semi-automated DNA sequencers based on fluorescent dyes are now the preferred method of sequencing DNA. In addition, to reducing the amount of work involved in DNA sequencing, automated sequencing yields significantly more bases of readable sequence. The basis of automated sequencing is similar to that of manual sequencing except that product strand is labeled with fluorescent dyes during the extension reaction and then detected during electrophoresis. Therefore, no preliminary labeling reaction is needed.

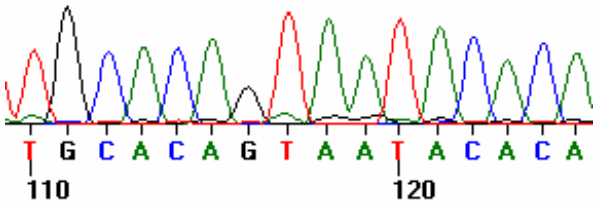


Sequencing reactions are carried out with a thermocycler in a single tube containing the template DNA, a primer, a heat stable DNA polymerase (eg., Taq), all four dNTPs and all four ddNTPs. The four ddNTPs that are conjugated with different fluorochromes (i.e.,

different emission spectra). Use of the thermocycler does not result in a geometric amplification as observed during PCR since a single primer is being used and the product strands are not used in subsequent cycles. The template strand is reused in subsequent cycles though. As in manual DNA sequencing the ddNTPs are randomly incorporated into the product strands during the extension reaction and no additional nucleotides can be added to that particular DNA product. The result is a mixture of DNA molecules complementary to the template strand but differing in length by a single nucleotide and containing one of the four possible dideoxy nucleotides at their 3'-ends.



The products of this DNA synthesis (i.e., sequencing reaction) are subjected to gel electrophoresis in a special apparatus that continuously analyzes the DNA molecules for fluorescence as they pass by an excitation laser. As in manual sequencing the gel can resolve DNA molecules that differ by a single nucleotide. These progressively longer DNA molecules pass through this fluorometer one at a time and the fluorescence intensity for each of the four possible emission spectra will be determined. The wavelength of the emission spectra will indicate which of the dideoxynucleotides was incorporated into the terminal position of that DNA molecule. Therefore, the DNA sequence can be directly determined by the instrument during electrophoresis. The instrument provides the sequence of the DNA as well as a graphical printout of the fluorescence intensities for each of the four chromophores plotted against nucleotide number. The graphical printout can help resolve any ambiguities in the sequence.



Compressions and stops. Occasionally it is not possible to determine the DNA sequence because of compressions and stops. Compressions are caused by secondary DNA structure in the 'product' strand. DNA strands containing secondary structures that are stable in urea will migrate faster during electrophoresis than unfolded DNA molecules. This will result in bands that are close together and exhibit irregular spacing. Stops, or bands in all four lanes, are the result of false termination during the sequencing reaction. These stops are due to secondary structures in the 'template' strand. When the polymerase encounters a secondary structure in the template it will dissociate without incorporating a nucleotide. The secondary structures are often the result of dyad symmetry and or G-G base pairing that occurs in GC rich DNA.

Alleviating Compressions

- sequence 'other' strand
- formamide in gel
- nucleotide analogs
  - dITP (inosine)
  - 7-deaza-dGTP
- Taq polymerase, 72°
- formamide or DMSO in rxn
- terminal transferase chase

Several different ways to alleviated or minimize compressions and stops have been described (Box). The simplest and usually most reliable method is to sequence the other strand. The factors contributing to the secondary structure are typically not found on the

complementary strand. One strategy to eliminate compressions is to include formamide in the gel in an attempt to reduce the amount of base pairing. Another strategy is to lower the stability of the secondary structures in the product strand by substituting ITP or 7-deaza-dGTP for GTP. These nucleotides will prevent some of the secondary structure due to G-G base pairing. Stops may be eliminated by minimizing the secondary structure in the template strand either before or during the sequencing reactions. For example, using the Taq polymerase (heat stable) at 72^{o} might eliminate the secondary structure in the template. It is also possible to minimize template secondary structure by including chaotropic agents such as formamide or DMSO (up to 10%) in the sequencing reactions. An alternate strategy in cases template secondary structure persists, such as especially GC-rich DNA, the product strands can be treated with terminal transferase. DNA strands that are the result of false stops will still have a 3'-OH. The terminal transferase will add a large random number of nucleotides to the 3'-OH and thus eliminate these false stops from the analysis since these fragments will now migrate near the top of the gel and not interfere with readable sequence. Furthermore, the random number of nucleotides added to the 3'-end will diminish the effects on the analysis.

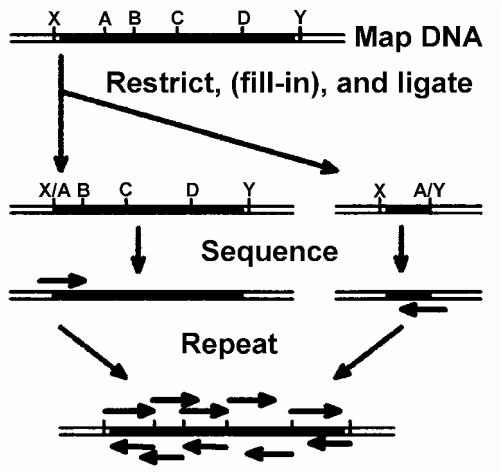
EXTENDED SEQUENCING STRATEGIES

The number of nucleotides that can be determined from a single sequencing reaction is limited. Several strategies have been described for the sequencing of long DNA fragments (Table). The choice of method(s) will depend largely and how much sequencing needs to be done and the resources available.

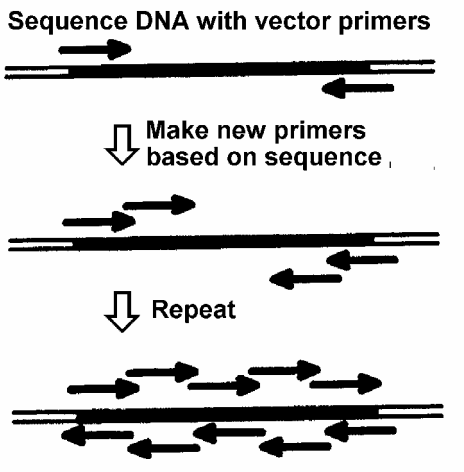
Extended Sequencing Strategies

Method	Advantages	Disadvantages
Restriction Digestion and Subcloning	• preparation simple • contig assembly easy	• relies on convenient restriction sites
Primer Walking	• preparation simple	• slow and expensive
Nested Deletions	• contig assembly easy	• laborious to generate
Shotgun	• quick and easily automated	• gaps and inefficient

The most straightforward and simplest method is to take advantage of restriction sites within the target DNA fragment. The cloned DNA is cut with a restriction enzyme within the MCS and a restriction enzyme within the insert fragment (Figure). If the restriction sites are not compatible, the recessed ends are filled-in with Klenow and dNTPs. The plasmid is then re-circularize by treating with DNA ligase, and thereby creating a deletion mutant that can be sequenced using primers flanking the MCS. Restriction fragments can also be subcloned and sequenced. One limitation of this method is the potential lack of conveniently located restriction sites.

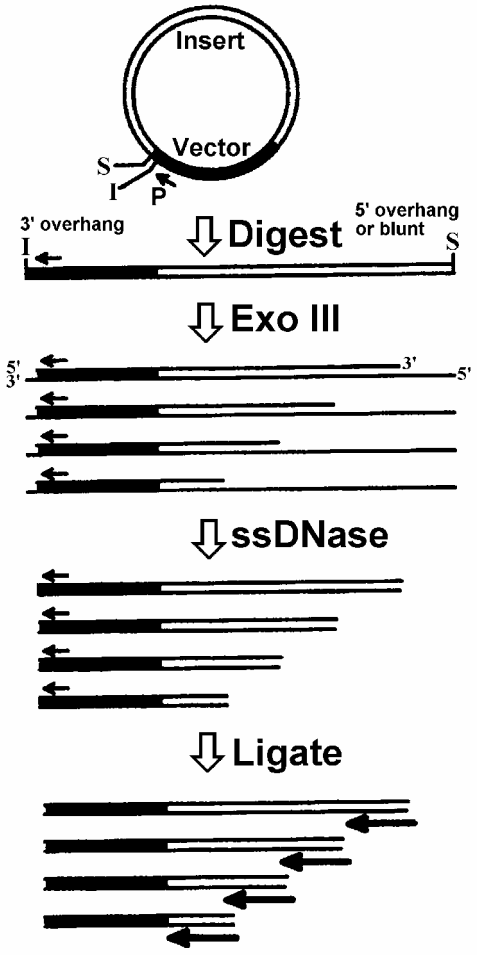


Sequence Restriction Fragments



Primer Walking

Another strategy for extended sequencing, often called 'prime and run' or 'look and leap', is based upon synthesizing primers based upon known sequence (Figure). In this method one sequences a region of DNA and then designs an oligonucleotide primer corresponding to most distal tract of reliable sequence. This new primer is then used to sequence more of the insert DNA and the procedure is continued until the complete sequence is obtained. The same general rules for the design of PCR primers also apply to the design of sequencing primers and several computer programs that predict potential primer sequences are available. Generally primers should be 40-60% GC composition and be at least 18 nucleotides in length. In addition, primers should not fold into hairpin loops or other secondary structures.



Prime and run strategies tend to be slow and expensive in that one needs to continuously synthesize primers and only a small stretch of DNA is sequenced before the next primer can be synthesized. In addition, it may not be possible to identify optimal sequencing primers resulting in primers that do not work well in the sequencing reaction. Therefore, this method is best for filling in small gaps in sequence information or confirming questionable stretches of DNA sequence.

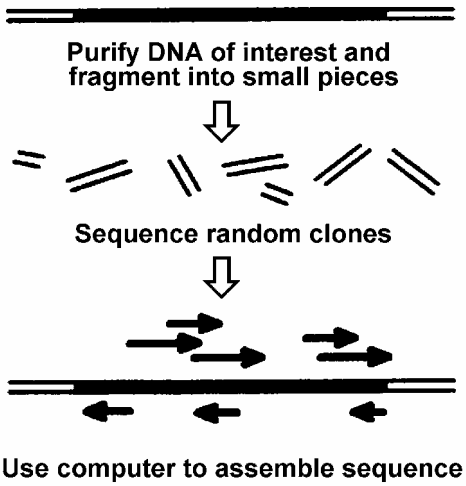
The generation of a series of nested unidirectional deletion mutants overcomes some of the limitations in prime and run strategies and does not rely on the presence of convenient restriction sites. The plasmid containing the insert to be

sequenced is cut with two different restriction enzymes within the MCS (Figure). The site closest to insert DNA should leave either a blunt end or 5'-overhang (S). The other restriction enzyme should leave a 3'-overhang (I). Many of the common sequencing vectors have poly-linkers in which the outermost restriction sites leave 3'-overhangs.

After the plasmid is linearized, it is digested with Exonuclease III which removes nucleotides in the 5' to 3' direction. However, ExoIII cannot digest dsDNA which has a 3'-overhang. Therefore, treatment of the linearized plasmid with ExoIII will only result in an unidirectional digestion toward the insert DNA fragment. The vector, including the primer site, will be protected because of the 3'-overhang. The ratio of DNA to ExoIII needs to be carefully controlled so that the digestion proceeds at a known rate. Aliquots are remove at defined intervals after the start of the ExoIII digestion and the reaction is stopped. Each subsequent aliquot will contain progressively shorter DNA fragments.

The samples are then treated with S1 (or mung bean) nuclease to remove the single-stranded protruding 3'-ends. The DNA is then treated with Klenow to repair any 5'-overhangs generated by the S1 treatment, insuring that the ends are blunt. Following ligation and transformation, colonies are picked and analyzed for the size of the insert fragment remaining. Ideally this method will yield a series of clones with inserts that are progressively shorter by an appropriate length. The nested deletion clones are then sequenced from the primer of the plasmid and the entire sequence is obtained from the overlapping sequences. It is somewhat laborious to generate the series of unidirectional mutants, but once they are obtained, the sequencing proceeds rather rapidly. The nested deletions may also be useful for other applications.

Another method for carrying out extending sequencing is the 'shotgun method'. In this method a series of random overlapping clones is generated and sequenced. Digestion of DNA with DNaseI in the presence of Mn^{2+} will randomly cut dsDNA. The resulting fragments are then size-fractionate, subcloned, and sequenced. The sequences are then analyzed by computer programs that will align the series of overlapping clones and generate the entire sequence. The DNAse treatment is very sensitive to DNA/DNAse ratios and often fails to generate appropriate sized fragments. In addition, to insure that there are no gaps in the sequence due to lost fragments, more sequencing than necessary is carried out. Because of these problems, this method is not widely used in small sequencing projects. However, this method is amenable to automation and is used in large genome projects.



MAXIM AND GILBERT

Chemical Cleavage
1. End label DNA.
2. Cleavage reactions.
3. Gel electrophoresis.

The Maxim and Gilbert method is based on the principal that DNA can be specifically cleaved at certain residues. Initially the Maxim-Gilbert method was more reproducible and accessible than the dideoxy method. However, the dideoxy procedure is now superior and chemical cleavage is primarily used to check the sequence of synthetic oligonucleotides. In addition, the chemical cleavage method is sometimes used to sequence short stretches of GC-rich DNA, which is predisposed to compressions and stops in the dideoxy method.

The first step in Maxim/Gilbert sequencing is to end label the DNA (Box). The most straight forward method is to use T4 polynucleotide kinase and γ-^{32}P-ATP, resulting in a single ^{32}P at the 5' end (see section on labeling DNA probes). Double stranded DNA will be labeled at both ends and therefore the two stands must be separated. Sometimes it is possible to separated the two single strands by electrophoresis. Generally, though, such preparations will result in preparations in which the complementary strands contaminate each other. The labeled DNA can also be digested with a restriction enzyme and the desired fragment isolated which would result in only one of the strands containing a radiolabel. It is also possible to label recessed 3'-ends by filling in with Klenow and α-^{32}P-dNTPs. This strategy, through the choice of restriction enzymes and radioactive nucleotide, may result in the asymmetric labeling of the two strands.

Base	Chemical
G	dimethyl sulfate
R (A+G)	piperidine formate
Y (C+T)	hydrazine
C	hydrazine + NaCl
A>C	NaOH at 90^{o}

The labeled ssDNA is placed in four or five separate tubes containing different chemicals that specifically cleave at certain residues (Table). Some protocols will omit the A>C reaction. The reaction conditions must be carefully controlled so that the cleavages are random (i.e., partial). The samples are then analyzed by gel electrophoresis under conditions which allow the resolution of fragments differing by a single nucleotide, as described for dideoxy sequencing. Since the chemical cleavage reactions are not absolutely base specific, it is more difficult to determine the sequence than in the dideoxy-chain termination method.

CHAPTER 24--SEQUENCE ANALYSIS AND BIOINFORMATICS

The field of bioinformatics represents a convergence of explosive growth in both biotechnology and information technology. Traditionally bioinformatics has been synonomous with the management and analysis of DNA and protein sequences. The term now is more broadly used to include epidemiology (especially genetic epidemiology) and evolutionary studies. One aspect of computational biology research is the development of new algorithms and statistics. The other aspect of bioinformatics is the use of these tools for the analysis and interpretation of sequence data.

Common Sequence Analyses

- compiling and editing sequences
- manipulating sequences
  - complement, reverse, or both
  - translating
- structural analysis
  - searching for coding regions
  - physical properties
- comparisons and homology searches
  - aligning sequences
  - sites and motifs
  - gene identification

DNA sequencing produces large amounts of data that would be extremely tedious to analyze without computers. Several commercially available software packages will carry out routine sequence analyses (Box). Many sequence analysis programs are also available on the internet (see Appendix). Some of these programs can be downloaded and used on a free standing computer. In other cases the sequence is submitted via the website or by email and the results of the analysis are provided.

A DNA sequencing project will require the use of a computer to assist in compiling and editing the sequence data. Overlapping clones are found and assembled into contiguous sequences. Computers are also used to generate the complementary sequence, reverse the 5'→3' orientation, or provide amino acid translations of nucleic acid sequences. It is also possible to calculate various physical, biochemical or structural properties of both nucleic acids and proteins. For example, algorithms to predict the secondary structure of either proteins or RNA are available. Specific sites or motifs within a gene, such as restriction sites, promoter elements or other DNA signals, as well as various protein motifs (eg., glycosylation sites, signal sequences, etc.) can be identified with the computer.

Alignment of related sequences and homology searches are common analyses performed on both nucleic acid and protein sequences. For example, it is relatively easy to obtain sequence data, but it is quite difficult to predict protein structure and function. Protein structure and function can be approximated from similarities to known sequences. Such similarities are identified by aligning DNA or protein sequences.

SEQUENCE ALIGNMENT

Alignments provide a powerful way to compare sequences for either evolutionary relatedness or structural/functional relatedness. Sequences can be compared by either **global or local alignments**. Global alignment forces complete alignment of the input sequences, whereas local alignments will only align their most similar segments. The choice of method will depend on whether the sequences are presumed to be related over their entire lengths or to only share

isolated regions of homology. The algorithms to carry out global and local alignments are similar, but the statistics associated with the output are different.

Homologs are structures or objects that share a common evolutionary origin. Objects with similar structure or function, but no common ancestor are **analogs**. Homologs can be classified as **orthologs** or **paralogs**. Orthologs are homologous genes from different species that arose from a common ancestor gene. Paralogs are homologous genes that are the result of gene duplication within an evolutionary lineage (eg., species) and may have different or similar functions.

Pair-wise sequence alignments use a **scoring matrix** to calculate the best alignment between two sequences. A scoring matrix contains information on how to score each match in an alignment and penalties for mismatches. For example in DNA alignments, matches are assigned a value of 0.9 and mismatches a value of -0.1. Proteins are more complicated and the scoring matrix must also take into account similarity between residues and their relative abundance. For example, tyrosine and tryptophan residues are uncommon and weighted heavier that alignments between common residues such as alanine and serine. Similarly, alignments between glutamate and aspartate are scored positively because of their chemical similarity, whereas matches between dissimilar residues are penalized. Furthermore, amino acid similarity can be defined as similar on either a chemical basis or a structural basis (Table). No single scoring matrix can be universally used for proteins. BLOSUM and PAM are two well known scoring matrices.

```
GCGCCTC
|||  ||
GCGGGTC
```

(5 x 0.9) +
(2 x -0.1) = 4.3

Amino Acid Similarities

Chemical	Physical
A, G	C, S
D, E	D, L, N
F, Y	E, Q
K, R	F, H, W, Y
I, L, M, V	I, T, V
Q, N	K, M, R
S, T	

The **gap penalty** is another parameter used in sequence alignments. Mathematically, the penalty for opening a gap of length k is defined as:

$$W_k = a + bk$$

where a is the gap opening penalty and b is the gap extension penalty. Both the opening penalty and the extension penalty can be varied and it may be useful to run alignments with different values unless it is already known the types of matches being searched for. For example, having a large open and a large extend penalty is good for analyzing closely related proteins. A large open penalty and small extend penalty may be good for situations where the distance between domains is not crucial. A small open penalty and large extend penalty (ie, many small inserts) is useful for distantly related homologous proteins.

Aligning sequences involves submitting the two sequences, choosing the scoring matrix and prescribing the gap penalties. The alignment giving the best score for a particular scoring matrix and the prescribed gap penalties is returned. However, this is not necessarily the most biological significant alignment. Quite often finding the optimal alignment

involves carrying out several alignments using different values for these parameters. In addition, the final alignment often involves using human insight.

In contrast to pair-wise alignments, which always return the best possible mathematical match, multiple alignments are a first approximation. The human eye coupled with biological insight is much better at spotting patterns in multiple alignments that the currently available computer algorithms. Alignments are a non-trivial computational task and it is best not to treat them like a black box programs

SEARCHING DATABASES

Comparing a sequence against a database to discover similarities is one of the most frequently used and powerful tools in bioinformatics. In essence, searching a database is the same as aligning two sequences. However, to perform a pair-wise alignment of the query sequence with every sequence in the database would be too time consuming. Therefore programs use heuristic strategies to speed up the analysis are usually used. FASTA and BLAST are the two most commonly used programs for searching databases. Both programs use rapid exact match procedures to first identify sequences which are most likely to be related. These sequences are then subjected to further analyses by alignment programs. BLAST (basic local alignment search tool) uses an approach based on matching short sequence fragments and then finds the best local alignments between the query sequence and the database sequences. Several BLAST programs are available for different types of sequence matching (see Table).

Databases

- primary (original biological data) vs. secondary (value added)
- three 1^{o} DNA databases
  - GenBank
  - EMBL
  - DDBJ
- subdivisions
- annotated

BLAST PROGRAMS

PROGRAM	QUERY	DB	COMMENTS
BLASTP	protein	protein	compares amino acid query against protein sequences
BLASTN	DNA	DNA	compares nucleotide query against DNA sequences
BLASTX	DNA	protein	compares 6X translations of nucleotide query against protein sequences
TBLASTN	protein	DNA	compares protein query against 6X translations of DNA sequences
TBLASTX	DNA	DNA	compares 6X translations of nucleotide query against 6X translations of DNA sequences

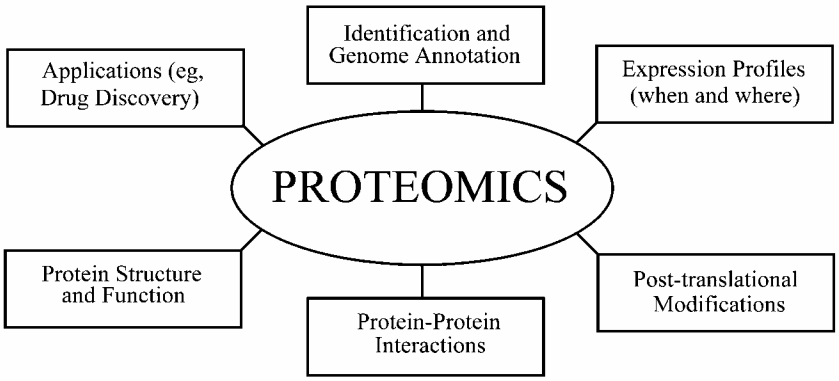
BLAST searches of the most current version of GenBank can be performed at the National Center for Biotechnology Information (NCBI) by cutting and pasting the query sequence onto their web page (http://www.ncbi.nlm.nih.gov/ BLAST/). The entire database or

subdivisions of the database can be searched. It is also desirable to 'filter' low complexity sequences such as long runs of repetitive sequence (eg., tracts of poly-alanine, etc.) and most BLAST programs filter by default. Such regions of sequence can give spuriously high scores against unrelated proteins. Similarly, iterative searches are prone to contamination by regions corresponding to coiled-coils or transmembrane domains. These characteristics might match the initial search and the program then emphasizes these inappropriate characteristics. The user can also choose cut-off values (i.e., E-values), the number of matches to report, and the number of alignments to show.

The entire database(s) is searched and sequences with similarity to the query sequence are identified and ranked accordingly (see Appendix 2 for sample output) to the alignment score derived from the pair-wise alignment. In addition to the this 'raw or bit score', the output includes a 'statistical score' and the alignments. The statistical score, or expectation value (E-value), provides a measure of the expected number of sequences in the database that would achieve a given alignment score by chance. The E-values are more useful in terms of judging the significance of a match than the alignment score. E-values for proteins and DNA should be less than 0.1 and 0.0001, respectively, to be considered significant. Examining the alignments will also help in inferring whether the hit is significant. For example, protein global alignments which show 25% or greater sequence identity over a stretch of at least 80 amino acids should exhibit the same basic structure. The identified sequences can be also retrieved and subjected to further analyses. Database searches may assist in identifying an unknown gene or provide clues about its function.

PROTEOMICS

The availability of the complete genomic sequence from several organisms represents a major advancement in the understanding of biology. However, the gene sequences provide a limited amount of information about the proteins that are encoded by the genes. In addition, phenotype will be determined by the proteins and not the genes (i.e., genotype). Efforts are now being directed towards the characterization of the proteome, or the complete set of proteins found in a cell, tissue or organism. Proteomics encompasses many types of activities and analyses with a goal of having a complete understanding of protein function and to apply this knowledge (Figure).



One objective of proteomics is to identify the complete complement of proteins. Genomics provides a first step in this process since all of the genes within a genome can potentially be identified. However, the algorithms used to predict the exon/intron structure of a gene, and to the lesser extend the beginning and end of a gene, are not always accurate. In addition, some genes can undergo alternate splicing and thus produce multiple proteins from the same gene. Therefore, many of the predicted genes will need to be confirmed by additional analysis.

The proteome is a dynamic entity in that genes are expressed at specific times and places and environmental conditions will also influence gene expression. Furthermore, many proteins are found in specific subcellular compartments or modified post-translationally (eg., phosphorylation, glycosylation, acylation, proteolysis, etc.). The subcellular location of a protein as well as the post-translational modifications will have an impact on cellular phenotype. Protein function is also dependent on protein-protein interactions and resolving these networks of protein-protein interactions is also central to understanding protein function and cellular phenotype. Knowing the structure of proteins also contributes to understanding protein function. Protein structure can be determined by X-ray crystallography and biophysical techniques. Some elements of protein structure can also be inferred from the protein sequence and comparing these sequences to related proteins in which the structures have been determined. The information about a protein's structure and function can then be used in applications such as drug development.

Gene identification and expression. Some information on proteins is already available due to previous research and protein characterization and this information is easily incorporated into the genome databases. Similarly information on the identity and expression of proteins will be continuously generated as particular proteins are studied. However, this approach is selective in that proteins and genes which are associated with interesting phenomena will be preferentially studied and it will take substantial time to individually characterize all of the genes with unknown function.

Expressed sequence tags (ESTs) and DNA microarrays provide a more rapid means in regards to characterizing genes and their expression. ESTs represent a collection mRNA sequences (i.e., cDNA sequences) maintained in DNA sequence databases which have been cloned from a defined source and partially sequenced. The cDNA clones have been chosen at random from a cDNA library without selection and represent the mRNA profile of the source (i.e., organism, tissue, cell, etc.) and thus provide information about gene expression. These ESTs can also be used to assist in the identification of expression start sites and positions of introns, even in the cases of unknown genes, by aligning the genomic sequence with the EST sequence. The original clones from which the EST sequences were obtained are maintain in respositories and can be obtained and sequenced in their entirities. Micro-array technology (see section on DNA Microarrays in Hybridization Chapter) is being used to identify the expression of mRNA on a large scale.

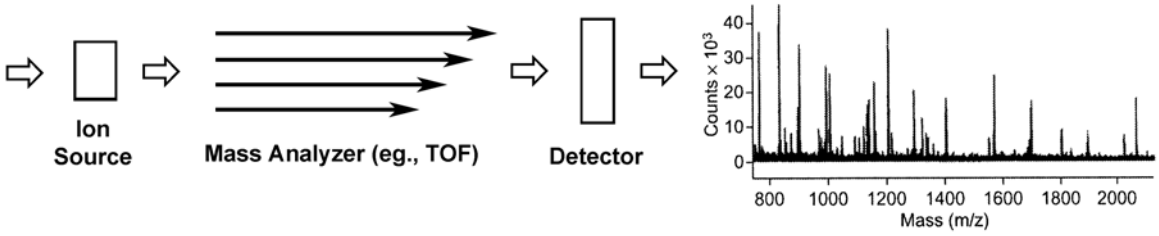
Protein identification. However, in many cases the genes of specific proteins have not been identified and the presence of mRNA does not always correlate with protein expression. The evaluation of proteins on a large scale, or by high throughput methods, is

not as advanced as the generation and analysis of DNA sequence. Much of protein analsysis still involves the analysis of one protein at a time and requires the separation, identification and characterization of proteins resolved from complex mixtures.

Gel electrophoresis of proteins is the predominant technique used in the separation and isolation of proteins (see Chapter on Protein Electrophoresis). One-dimensional SDS gel electrophoresis is capable of resolving many complex mixtures of proteins into individual polypeptide chains. Two-dimensional gels (IEF + SDS-PAGE) can be used if additional resolution is needed. Proteins are then be transferred to a membrane and subjected to N-terminal microsequencing (see Microsequening section in Protein Purification Overview chapter). In cases where the N-termini are blocked it may be possible to break the protein into smaller peptides with site-specific proteases or chemicals and determine an internal sequence. However, substantially more protein is needed to determine these internal sequences.

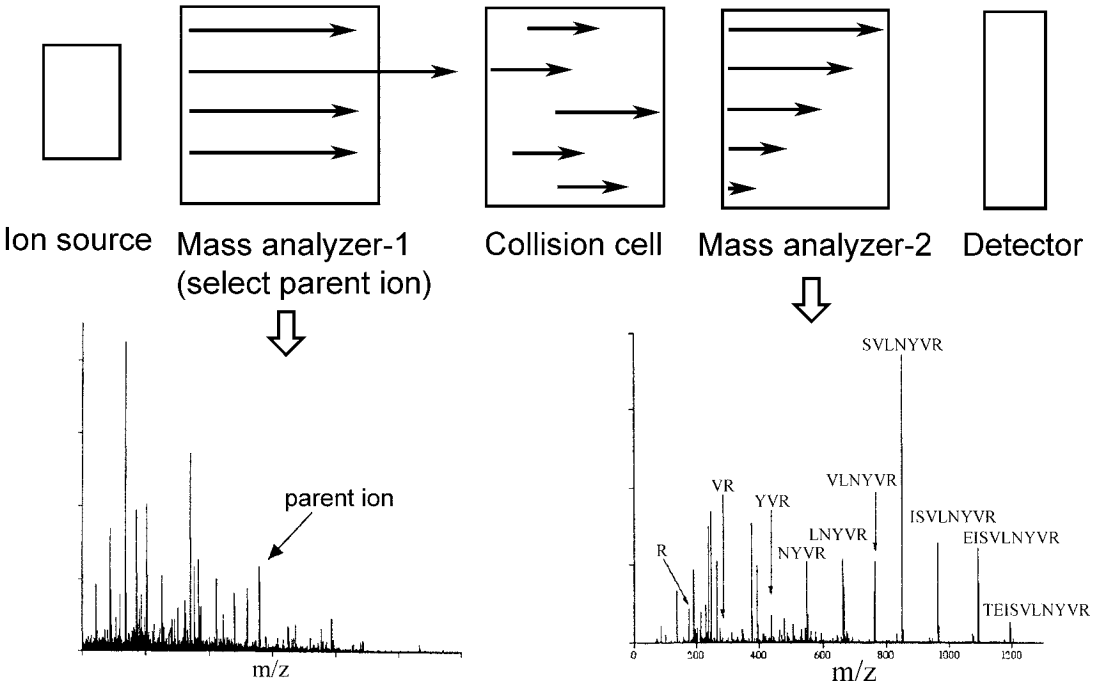
Problems with blocked N-termini and low abundance may be alleviated by a mixed peptide approach. In this method proteins are subjected to gel electrophoresis, transferred to a membrane and the band of interest excised as in a typical microsequencing protocol. The isolated protein is then cleaved into peptides with CNBr (cleaves at Met) or skatole (cleaves at Trp). On average 3-5 peptides will be generated depending on the frequency of these these amino acids. The membrane is then subjected to 6-12 automated Edman cycles. Every cycle will result in several amino acid residues. This mixed sequence data are then analyzed by a computer algorithm which will sort and match the data against protein or DNA databases to identify candidate protein(s). (Additional information on these programs is available at http://fasta.biochem.virginia.edu/)

Mass spectrometry. An increase in the sensitivity of protein identification can be achieved with a mass spectrometer. Mass spectrometers consist of three basic components: an ionization source, one or more mass analyzers, and an ion detector. Various types of instruments utilizing different types of ionization sources and mass analyzers exist. Biological samples are ionized through the addition or loss of protons. Following ionization the masses of the components are determined. For example, in a time-of-flight mass analyzer the time it takes to traverse the length of the flight tube is a function of mass. A spectrum of the components plotted according to mass is recorded.



The mass spectrometer can be used to determine accurate peptide masses or amino acid sequences of peptides. This information is then be used to identify proteins by searching databases. For example, a protein is digested with trypsin, or another site specific

protease, and the resulting peptides analyzed by mass spectrometry. This acquired MS spectrum can then be compared with predicted MS spectrums of all the proteins in a database. The advantage of this method is that the analysis can be carried out quickly and can be completely automated. However, there is some ambiguity in protein identification due to peptide mass redundancy (i.e., same amino acid composition but different sequence will be the same mass). In addition, factors which affect the accurate determination of peptide mass (eg., post-translational modifications) will limit the abililty to successfully identify proteins in databases.



Tandem mass spectrometry (MS/MS) can be used to deduce the amino acid sequence of a peptide. In this approach ions with particular masses (i.e., m/z values) are separated from the other ions and then fragmented in a collision chamber. The collision chamber is filled with an inert gas (usually argon or nitrogen) and collisions with the ions and the gas atoms result in covalent bonds being broken. The masses of the resulting fragments are then determined by a second mass analyzer. The mass spectrum of the fragments is diagnostic of the molecular structure of the parent ion. For example, peptides tend to fragment primarily along the peptide backbone at the amide bond which creates a ladder of ions that is indicative of the amino acid sequence. The sequence is used to search databases. Multiple peptides from the same protein can be analyzed in this fashion and thus increasing the confidence of the protein identification. The disadvantage of MS/MS is that the process is not easy to completely automated. Although computer programs are available, interpretation and analysis of the mass spectrum requires some human guidance.

High throughput analyses. The procedures described above rely upon identification of proteins by gel electrophoresis followed by their subsequent identification using either one- or two-dimensional gel electrophoresis. The ability of two-dimensional gels to resolve

thousands of polypeptides on a single gel makes it a powerful technique in proteomics. However, 2-D gels are somewhat limited to the identification of relatively abundant proteins in the mid-to-low molecular weight ranges. Furthermore, membrane associated proteins and cytoskeletal proteins are difficult to solubilize and hydrophobic proteins do not always resolve well in the isoelectric focusing dimension. In addition, two-dimensional gel electrophoresis is a labor intensive and time consuming process and the protein 'spots' need to be further characterized by mass spectrometry or Edman degradation individually.

To overcome these limitations high throughput methods to analyze complex mixtures of proteins are being developed. One approach is to combine a high throughput microcaillary liquid chromatography with tandem mass spectrotometry. In this approach complex mixtures of proteins are treated with site-specific endoproteases. The resulting peptides are subjected to liquid chromatography (HPLC) without a prior gel electrophoresis step and then directly introduced into the tandem MS. Elimination of the gel electrophoresis step provides provides a greater amount of flexibility in the proteolytic digestion and also increases the overall sensitivity. The first MS analysis will determine an accurate mass of the peptides which is indicative of the amino acid composition and then the fragmentation of those peptides can be used to predict the amino acid sequence. Databases are then searched for potential matches.

This LC/MS/MS approach can be used to generate protein expression maps that are analogous to EST databases. Crude proteins from a particular tissue can be fractionated into cytosolic, membrane/organellar, cytoskeletal, etc. fractions and then analyzed for the expressed proteins in the mixture. This type of analysis will work best in organisms in which the genome sequence or large banks of EST sequences are available. This approach can also be used in the identification of the components of less complex protein mixtures such as large multi-protein complexes. Similarly protein-protein interactions can be discerned by affinity purifying all of the proteins that interact with the target protein and then analyzing the mixture. Methods to purify binding proteins include: co-immunoprecipitation with antibodies against the target protein, co-precipitation with an affinity tagged (eg., His_6) recombinant target protein, or affinity chromatography with the target protein.

Additional Reading on Proteomics

Yates, J.R. 2000. Mass spectrometry from genomics to proteomics. Tr. Genet. 16:5-8.

Graves, P.R. and Haystead, T.A.J.. 2002. Molecular biologist's guide to protomics. Microbiol. Mol. Biol. Rev. 66:39-63

APPENDIX 1. WEBSITES

List of links to molecular biology sites

http://www.expasy.ch/alinks.html

Sequence retrievial system (capable of searching numerous databases and medline simutaneously)

http://www.ncbi.nlm.nih.gov/Entrez/

Database Searches (submit a query sequence and search for homologies in the databases)

http://www.ncbi.nlm.nih.gov/BLAST/
http://www2.ebi.ac.uk/fasta3/

Translation (translate nucleotide sequence into protein sequence)

http://www.expasy.ch/tools/dna.html

Pairwise Alignments

http://genome.eerie.fr/fasta/align-query.html

Multiple Alignments

http://www.ibc.wustl.edu/service/clustal.html
http://dot.imgen.bcm.tmc.edu:9331/multi-align/Options/clustalw.html

Protein Analysis (numerous programs for analyzing proteins)

http://www.expasy.ch
http://www.ebi.ac.uk/Tools/index.html

2^{o} Protein Structure

http://www.embl-heidelberg.de/predictprotein/predictprotein.html

APPENDIX 2. RESULTS OF BLAST SEARCH

Probability of match appearing by chance

```
Query= Pbpp58b (423 letters)

Database: nr (493,611 sequences; 154,780,071 total letters)

                                                                    Score     E
Sequences producing significant alignments:                         (bits)  Value

sp|Q08168|HRP_PLABE 58 KD PHOSPHOPROTEIN (HEAT SHOCK-RELATED PRO...   334  1e-90
gb|AAC37300.1| (L21710) 58 kDa phosphoprotein [Plasmodium berghei]    329  3e-89
pir||T10455 heat shock related protein - Plasmodium berghei >gi|...   250  2e-65
sp|P50503|HIP_RAT HSC70-INTERACTING PROTEIN >gi|4379408|emb|CAA5...   106  5e-22
sp|P50502|HIP_HUMAN HSC70-INTERACTING PROTEIN (PROGESTERONE RECE...    87  3e-16
gb|AAF45894.1| (AE003429) CG2947 gene product [Drosophila melano...    87  4e-16
pir||T24865 hypothetical protein T12D8.8 - Caenorhabditis elegan...    86  5e-16
pir||T04562 hypothetical protein T12H17.60 - Arabidopsis thalian...    81  2e-14
.
.
.
.
.
gb|AAC60555.2| (S59774) STI1 stress-inducible protein homolog [S...    48  2e-04
gb|AAD33401.1|AF129086_1 (AF129086) carboxy terminus of Hsp70-in...    48  2e-04
pir||A56534 P58 protein - bovine >gi|468012|gb|AAA17795.1| (U046...    48  2e-04
gb|AAB49720.1| (U89984) transformation-sensitive protein homolog...    47  4e-04
pir||T16689 hypothetical protein R05F9.10 - Caenorhabditis elega...    47  4e-04
gb|AAD33400.1|AF129085_1 (AF129085) carboxy terminus of Hsp70-in...    46  6e-04
.
.
.
.
.
.
emb|CAA61595.1| (X89416) protein phosphatase 5 [Homo sapiens]          43  0.007
pdb|1A17|    Tetratricopeptide Repeats Of Protein Phosphatase 5       43  0.007
ref|NP_006238.1|| protein phosphatase 5, catalytic subunit >gi|1...    43  0.007
pir||S52570 phosphoprotein phosphatase (EC 3.1.3.16) 5, catalyti...    43  0.007
gb|AAB18614.1| (U12203) phosphoprotein phosphatase [Rattus norve...    43  0.007
sp|P53042|PPP5_RAT SERINE/THREONINE PROTEIN PHOSPHATASE 5 (PP5) ...    43  0.007
gb|AAB60384.1| (U25174) serine-threonine phosphatase [Homo sapiens]    43  0.007
sp|Q60676|PPP5_MOUSE SERINE/THREONINE PROTEIN PHOSPHATASE 5 (PP5...    42  0.009
gb|AAB70573.1| (AF018262) protein phosphatase 5; PP5 [Mus musculus]    42  0.009
```

Entries removed for clarity

Alignment Score

Identifier Line: database | accession # | name or locus

Entries removed for clarity

Examples of Alignments

```
>sp|P50503|HIP_RAT HSC70-INTERACTING PROTEIN >gi|4379408|emb|CAA57546.1| (X82021)
Hsc70-interacting protein [Rattus norvegicus] (Length = 368)

 Score =  106 bits (261), Expect = 5e-22
 Identities = 60/224 (26%), Positives = 97/224 (42%)

Query: 1   MDIEKIEDLKKFVASCEENPSILLKPELSFFKDFIESFGGKIKKDKMGYXXXXXXXXXX 60
           MD  K+ +L+ FV  C ++PS+L   E+ F ++++ES GGK+
Sbjct: 1   MDPRKVSELRAFVKMCRQDPSVLHTEEMRFLREWVESMGGKVPPATHKAKSEENTKEEKR 60

           (SDEEEEDEEEEEEEEEDDDPEKLE)
Query: 61  XXXXXXXXXXXXXXXXXXXXXXXXXXXXAVECPPLAPXXXXXXXXXXXXXXCKLKEEA 120
                                      + P                   + K  A
Sbjct: 61  DKTTEDNIKTEEPSSEESDLEIDNEGVIEADTDAPQEMGDENAEITEAMMDEANEKKGAA 120

Query: 121 VDLVENKKYEEALEKYNKIISFGNPSAMIYTKRASILLNLKRPKACIRDCTEALNLNVDS 180
           +D + + + ++A++ +   I      A++Y KRAS+ + L++P A IRDC  A+ +N DS
Sbjct: 121 IDALNDGELQKAIDLFTDAIKLNPRLAILYAKRASVFVKLQKPNAAIRDCDRAIEINPDS 180
```

Filtering--Glu rich

```
Query: 181 ANAYKIRAKAYRYLGKWEFAHADMEQGQKIDYDENLWDMQKLIQ 224
           A  YK R KA+R LG WE A  D+    K+DYDE+   M + +Q
Sbjct: 181 AQPYKWRGKAHRLLGHWEEAARDLALACKLDYDEDASAMLREVQ 224
```

Middle row = matches and similar residues (+)

```
>pir||T24865 hypothetical protein T12D8.8 -Caenorhabditis elegans (Length = 422)

 Score = 86.2 bits (210), Expect = 5e-16
 Identities = 44/101 (43%), Positives = 60/101 (58%), Gaps = 2/101 (1%)

Query: 119 EAVDLVENKKYEEALEKYNKIISFGNPSAMIYTKRASILLNLKRPKACIRDCTEALNLNV 178
           +A +   N  ++ AL  +   I     SAM++ KRA++LL LKRP A I DC +A+++N
Sbjct: 121 KAQEAFSNGDFDTALTHFTAAIEANPGSAMLHAKRANVLLKLKRPVAAIADCDKAISINP 180

Query: 179 DSANAYKIRAKAYRYLGKWEFAHADMEQGQKIDYDE--NLW 217
           DSA  YK R +A R LGKW  A  D+    K+DYDE  N W
Sbjct: 181 DSAQGYKFRGRANRLLGKWVEAKTDLATACKLDYDEAANEW 221


 Score = 41.4 bits (95), Expect = 0.016
 Identities = 16/34 (47%), Positives = 23/34 (67%)

Query: 9  LKKFVASCEENPSILLKPELSFFKDFIESFGGKI 42
          LK+FV  C+ NP++L  PE  FFKD++ S G  +
Sbjct: 7  LKQFVGMCQANPAVLHAPEFGFFKDYLVSLGATL 40
```

Gaps to maximize alignment

A second high-scoring segment