KEY CONCEPTS AND LEARNING OBJECTIVES

1. Restriction enzymes cleave double-stranded DNA at specific sequences to produce smaller pieces, and gel electrophoresis can be used to separate the DNA fragments based on their size.
   a. Define the term palindrome and identify palindromic sequences in the DNA.
   b. Identify the types of DNA ends generated by different restriction enzymes.
   c. Determine the length of a DNA fragment from its electrophoretic migration.

2. DNA cloning refers to the ability to select a specific nucleic acid sequence from all the sequences in a cell, and produce it in a purified or enriched form. One way to accomplish this is to insert the selected DNA into a self-replicating genetic element (plasmid) using restriction enzymes. The plasmid containing the selected DNA is introduced into bacteria (transformation) or other appropriate host cell and replicated to produce millions of copies.
   a. Describe how a specific DNA molecule can be introduced into a plasmid for its subsequent propagation in a bacterial cell.
   b. Explain how recombinant DNA can be used in clinical medicine.

3. Nucleic acid hybridization uses a labeled, single-stranded DNA probe to detect a complementary nucleic acid sequence within a complex mixture. Southern blotting is a technique for detection and quantifying a specific DNA. Northern blotting is a technique for detecting and quantifying a specific RNA.
   a. Describe the principles behind Southern and Northern blotting.
   b. Explain the term restriction fragment length polymorphism (RFLP) when comparing DNA from different individuals.
   c. Interpret the results from an RFLP experiment (you can use the example given in class or in the associated small group problem set).
   d. Draw a simple family pedigree based on the results of an RFLP analysis.
   e. Describe in situ hybridization and discuss how it might be used in the diagnosis of disease.

4. The polymerase chain reaction (PCR) is a sensitive molecular technique that amplifies a specific nucleic acid sequence within a complex nucleic acid mixture. This method exponentially amplifies the sequence to generate millions of copies that are readily detectable by gel electrophoresis. This makes it possible to detect the presence of a rare nucleic acid sequence in a complex sample.
   a. Explain the principles behind PCR and design a PCR reaction to amplify a desired DNA fragment.
   b. Describe how PCR exponentially amplifies a PCR product.
   c. Describe how PCR can be used to diagnose a specific genetic disease, such as hemophilia.
   d. Define a microsatellite sequence, and explain why they are useful in forensic science.
   e. Explain how a microsatellite sequence can be used in PCR to obtain an individual’s genetic fingerprint (paternity testing, forensic analysis).

5. Detection of an RNA target by PCR can be accomplished by preparing a cDNA copy of the RNA and using the cDNA as the template for PCR. cDNA synthesis is performed by reverse transcriptase, a polymerase found in retroviruses that catalyzes the formation of DNA from RNA. Reverse transcription – PCR (termed RT-PCR) is important because it allows us to determine
which genes are being transcribed in a cell. It also allows the detection of RNA viruses, such as HIV and influenza A.

a. Describe the primary differences between DNA and cDNA.
b. Explain how cDNA is prepared and when cDNA would be used in PCR instead of DNA.
c. Explain how RT-PCR might be useful in the diagnosis of disease.

6. Microarray analysis uses hybridization techniques to monitor expression of thousands of genes simultaneously. Microarrays are powerful tools for research and personalized medicine (healthcare where medical decisions and treatments are customized for an individual patient).

a. Describe the primary advantage of a microarray experiment compared to other hybridization techniques such as Northern blotting or reverse-transcription PCR.
b. Describe how a microarray experiment would be performed.
c. Explain how the information generated from a microarray might be utilized to better understand disease processes, identify new therapeutic targets, or help select the most appropriate therapy for an individual patient.

7. DNA sequencing identifies the bases of a DNA fragment. It is essentially a modified PCR reaction with a single primer where labeled nucleotides are inserted into the DNA fragment as it is replicated.

a. Explain the theory behind chain terminator sequencing.
b. Explain how a DNA sequence can be used to predict the amino acid sequence of a protein that it encodes.
c. Describe how DNA sequencing can be used in the diagnosis or treatment of disease.

8. Next Generation Sequencing (NGS) allows millions of small DNA fragments to be sequenced at the same time (termed high throughput) compared to a single DNA sequence in traditional sequencing.

a. Describe how NGS is performed (Illumina platform).
b. Describe the major differences between NGS and traditional sequencing.
c. Explain how NGS might be used in clinical medicine in the future.

9. Putting it all together!

a. Given some basic information, determine the molecular technique most appropriate for a particular application. Example: You need to perform microsatellite analysis on a specimen obtained from a small blood sample found at a crime scene. What molecular test would be most appropriate?
b. Given some basic information, decide if the starting material for a molecular test should be DNA, RNA or cDNA. Example: As a medical oncologist, you want to know if your patient’s tumor cells express matrix metalloproteinases, enzymes that degrade extracellular matrix and promote tumor metastasis. You decide to run a PCR experiment. What type of starting material would you want for your analysis?

Resources:
Alberts Molecular Biology of the Cell, 6th edition is an excellent place to go if you need clarification. While I try to explain things clearly, each individual is unique with respect to how they learn. The text is outstanding, and it may explain things in a better way for you.
I. Cutting and Rejoining DNA Fragments.
A. Restriction enzymes cut double-stranded DNA at specific sequences.
   The sites recognized and cleaved by restriction enzymes are often palindromic. A palindrome is a sequence that reads the same after being flipped by 180° around its central axis.

B. Restriction enzymes can cut symmetrically or asymmetrically. Some enzymes, such as HaeIII, cut symmetrically (straight across the DNA) leaving blunt ends. Other enzymes, such as EcoRI and HindIII, produce staggered ends (called sticky ends). These short, single-stranded DNA overhangs help cut DNA molecules join back together.

Blunt ends

Sticky ends

Figure 8-24. Restriction nucleases cleave DNA at specific nucleotide sequences.

C. DNA fragments can be joined together (ligated) to produce new combinations.
When two DNA fragments are cut by the same restriction enzyme (and thus have complementary sticky ends), they can readily join to each other through base pairing of the single-stranded DNA overhangs. This is called annealing. In the adjacent figure, a target DNA (red) is mixed with four different DNA fragments (black) with slightly different single-stranded overhangs. The target DNA will anneal to the DNA with a complementary overhang and is much less likely to anneal to the other DNA fragments.
II. Creating Recombinant DNA

A. Joining DNA fragments from different sources.
DNA fragments from two different sources can be linked together using the sticky ends produced by restriction enzymes. Once the two fragments have annealed (via complementary base pairing), the cut ends are sealed with DNA ligase, and the new DNA is termed a recombinant DNA.

B. Multiplying a recombinant DNA sequence.
It is often necessary to obtain many copies of a specific DNA in order to manipulate it in the lab. One way to do this is to create a recombinant DNA by inserting the specific DNA sequence into a bacterial plasmid using restriction enzymes. A plasmid is a double-stranded DNA molecule that remains separate from the chromosomal DNA. Plasmids are introduced into bacteria through a process called transformation. When the bacteria are grown in culture, the plasmid (along with the inserted DNA) replicates along with the bacteria. The plasmid can be isolated from the chromosomal DNA. The inserted DNA can be cut from the plasmid using the same restriction enzyme and purified for use in a variety of applications, such as preparation of probes for blotting and hybridization.

Figure 8–27. The insertion of a DNA fragment into a bacterial plasmid. The plasmid and the DNA fragment to be
C. Recombinant DNA can be used to produce clinically relevant proteins.

Some plasmids contain the appropriate regulatory elements to force transcription and translation of the inserted DNA. Thus, cells can be forced to express large quantities of clinically or experimentally useful proteins. This table contains examples of clinically important proteins that are produced using recombinant DNA technology.

Do not memorize this table.

III. Separating and detecting specific nucleic acids.

A. DNA (and RNA) molecules can be separated based on size by gel electrophoresis.

The nucleic acid fragments are separated by running them through an electrical current within a gel matrix (agarose or polyacrylamide). The gel is submerged under an electrophoresis buffer that carries the electrical current and provides a buffer to maintain the pH. The DNA sample is placed into the wells at the top of the gel. Because DNA is negatively charged, it will migrate toward the positive electrode (anode). The DNA is invisible in the gel, but can be seen by staining with ethidium bromide, which intercalates between the bases of double-stranded DNA.

RNA can also be separated by size using gel electrophoresis. The primary difference between DNA and RNA gels is that special buffers must be used with RNA to inhibit its degradation.

A cartoon of a gel is pictured below (far left). The largest DNA fragments are located at the top of the gel and the smallest toward the bottom, because the larger fragments have more difficulty moving through the matrix than the smaller ones. The size of the DNA fragments can be estimated by comparing them to a set of DNA fragments of known size (markers). Next to the cartoon are photographs of gels. There are several types of gels. The first picture is of an agarose gel, which is the most common gel used to analyze nucleic acids, and the DNA has been stained with ethidium bromide. The second picture is a polyacrylamide gel. This type of gel is used to evaluate small DNA fragments [<250-300 base pairs (bp)]. The third picture is a pulse field gel. This is an agarose gel, but it is run under special conditions to separate very large DNA fragments (>50,000 bp). Normally, the electrical field is held constant while running a gel, but in pulse field gels, the electrical field is periodically shifted. This forces the DNA fragments to re-orient themselves toward the “new” anode. This re-orientation takes time, and it takes longer for larger pieces of DNA than smaller ones. Through multiple, brief shifts of the electrical field, the large pieces of DNA slowly separate from one another.
B. **Specific nucleic acid sequences can be detected by nucleic acid hybridization.**

Double-stranded DNA can be **denatured** by heating the DNA or exposing the DNA to alkali conditions. This breaks the hydrogen bonds holding the double helix together producing single-stranded DNA. Cooling the DNA slowly allows the complementary sequences to re-base pair and re-form the double helix. This is termed **hybridization.**

**Figure 8-33.** DNA can undergo denaturation and renaturation.

C. **Southern blotting combines electrophoresis and hybridization to detect specific DNA molecules in a complex mixture of DNA.**

Single-stranded DNA molecules used to detect complementary sequences are called **probes.** These molecules carry radioactive or chemical markers so they can be detected. Hybridization using DNA probes is so sensitive that it can detect complementary sequences at a concentration of one molecule per cell.

The figure to the left demonstrates high and low stringency hybridization conditions. High stringency ("strict") conditions do not tolerate mismatch between the probe and the target sequence. Low stringency conditions will allow some mismatch. This allows detection of related, but not identical DNA fragments.

**Figure 8-25.** DNA molecules can be separated by size using gel electrophoresis.
**Southern blot** analysis is used to detect a specific DNA in a complex mixture of DNAs. It begins with cutting of isolated DNA with a restriction enzyme to obtain small pieces and then using gel electrophoresis to separate the pieces based on size. The DNA is made single-stranded by soaking the gel in a NaOH solution and is then transferred to a membrane. Because the DNA is mobile within the gel, it is transferred to a membrane where it can be permanently affixed in its position. A radioactive single-strand DNA probe is incubated with the membrane. If a complementary sequence was present in the original DNA sample, the probe will bind via base pairing. Unbound radioactive probe is washed away and the membrane exposed to x-ray film. Any place where radioactivity is bound will be evident on x-ray film when it is developed.

Similar analysis can be done with RNA and is called a **Northern blot**. The specific RNA fragment is detected by hybridization with a single-stranded DNA probe same as in Southern blotting.
D. **Restriction Fragment Length Polymorphism (RFLP) analysis.**
Restriction digestion and Southern blotting can be used to detect a disease-causing gene. Sickle cell anemia is often caused by a single base change in the coding region of \( \beta \)-hemoglobin (also called \( \beta \)-globin). The mutation not only results in an amino acid substitution, but it also destroys a \textit{MstII} restriction enzyme cut site. This results in a change in the restriction pattern produced when DNA is cut with \textit{MstII} and is referred to as a **Restriction Fragment Length Polymorphism or RFLP**. One way to diagnose sickle cell anemia is to isolate DNA from a patient, cut the DNA with \textit{MstII} and perform a Southern blot. The loss of the internal \textit{MstII} cut site in this region of the \( \beta \)-globin gene can be identified when the size of the DNA fragment bound by the DNA probe is larger than normal – the probe binds to a 1.3 kb DNA fragment instead of the normal 1.1 kb fragment. Note that the probe hybridizes to the 1.3 kb and 1.1 kb bands on the Southern blot, but the 0.2 kb band is not visible because the probe does not cover that region of the DNA.

E. **Hybridization techniques can locate specific DNAs within cells and genes on chromosomes.**
Southern blot requires a large amount of DNA. In cases where sample DNA is limited, such as a patient biopsy, alternative hybridization techniques performed in cells or on chromosomes are utilized. Here, the samples are heated to break the hydrogen bonds in the double-stranded DNA so the single-stranded DNA probe has access to bind it’s complementary sequence, if present.

In the adjacent figure, six different probes have been used to mark their complementary DNA sequences on human chromosome 5. The chromosomes in this figure are undergoing mitosis and have already replicated their DNA so that each chromosome contains two identical DNA helices. Thus, each probe binds twice to each chromosome. This technique is called **FISH: fluorescent in situ hybridization**.

FISH can be used to detect specific genes in a tissue sample, such as in the diagnosis of Her2-positive breast cancer.

**Figure 8-34.** In situ hybridization locates genes on isolated chromosomes.
One type of breast cancer overexpresses Her2 (epidermal growth factor receptor 2), which promotes tumor cell growth. Her2 overexpression is often caused by gene amplification. That is, the region of chromosome encoding the Her2 gene has been abnormally replicated so the cancer cells make more Her2 mRNA and protein than normal. Her2-positive breast cancer is treated differently than other breast cancers; thus, oncologists need to know if their patient has Her2 overexpression, and FISH can be used to detect it. In this figure, there are two probes. The green probe binds to a sequence in chromosome 17 (Cep17, positive control). The red probe binds to the Her2 gene, and the nucleus is stained blue. A positive result for Her2 amplification is shown on the right where the Her2 gene copy number > 6.0 (ratio of Her2:Cep17 is >2.2).

IV. Polymerase Chain Reaction
Polymerase chain reaction (PCR) is a powerful tool to amplify DNA. This power lies in the selectivity of hybridization of the primers to a specific sequence and the ability of the DNA polymerase to faithfully copy DNA through repeated rounds of synthesis in vitro. Because the PCR product is amplified exponentially, billions of copies of a DNA fragment can be obtained in only a few hours.

A. Molecular Diagnosis with PCR.
PCR has become an important tool for detecting genetic differences between individuals and for detecting disease-causing genes.
1. **Hemophilia**

Hemophilia is a sex-linked disorder; the Factor VIII gene is on the X-chromosome. One mutation associated with hemophilia occurs in intron 18 of the Factor VIII gene and results in the loss of a \textit{Bcl} restriction enzyme cut site. This mutation can be detected using PCR to generate a 142 bp product that spans the \textit{Bcl} polymorphic site. The PCR product is treated with \textit{Bcl} and analyzed with gel electrophoresis. The amplified product from the normal gene is cleaved into 99 and 43 bp fragments (lanes 1, 3, 6 in gel cartoon to the left) by the enzyme, while the mutant gene product lacks the \textit{Bcl} site and is uncleaved (lanes 2, 4, 5, 7).

2. **Forensic science.**

**Microsatellites** [(also called short tandem repeats (STR) or simple sequence repeats (SSR))] are 2-6 bp sequences found in various positions (loci) in non-coding regions of the genome. The number of repeats varies between individuals (usually 4-40 repeats), but the sequence that is repeated does not change. Because of the high variability in the number of repeats, microsatellites can be used for identification, particularly in paternity testing and criminal investigations. PCR using primers that bind sequences on either side of the microsatellite produce a two bands of amplified DNA: One represents the maternal microsatellite and the other the paternal microsatellite. The length of the PCR product, and thus, its position after electrophoresis, will depend on the number of repeats.

In the figure, three microsatellite regions are examined (red, green, and blue). PCR amplification of each region produces two bands, and the results for individuals A, B, and C are shown on the gel along with the forensic sample (F). By comparing the lengths of the PCR products for each locus, the result shows only individual B matches at all three loci with the forensic sample.

*Figure 8-39.* PCR can be used to distinguish different individuals from one another.
B. Reverse Transcription – PCR (RT-PCR) is used to amplify RNA templates by PCR.

1. Preparation of complementary DNA (cDNA) from RNA.
RNA is single-stranded and is relatively unstable making it unsuitable for PCR. Instead, cDNA can be used. Reverse transcriptase is used to prepare cDNA that serves as a template in PCR to analyze gene expression or RNA virus infection.

Reverse transcription begins by annealing a short primer to the mRNA poly-A tail (called oligo-dT or poly-T). The reverse transcriptase enzyme uses the primer to begin synthesizing a DNA copy of mRNA. The resulting DNA-RNA hybrid is treated with RNase H and then a DNA polymerase copies the single-stranded cDNA into double-stranded cDNA. Because cDNA is prepared from mRNA, it contains no introns and the nucleotide sequences at the 5’ ends of the original mRNA are often absent.

![Figure 8-31. The synthesis of cDNA.](image)

2. RT-PCR in disease diagnosis.
Because PCR can amplify the signal from a single DNA molecule, it is an exceptionally sensitive method for detecting trace amounts of virus in the blood or tissues.

HIV and influenza are RNA viruses. Detection of these virus by PCR requires that the RNA first be converted to cDNA.

![Figure 8-38. PCR can be used to detect the presence of a viral genome in a sample of blood.](image)
V. Microarrays
A. Microarrays can analyze the expression of multiple genes simultaneously and are based on hybridization techniques. Microarrays are a collection of single-stranded DNA fragments representing a select group of genes that are fixed to a solid surface (usually a slide or membrane). Microarrays range in size from 10-20 genes focused on a specific pathway (example below) to 200,000+ genes across a handful of slides.

To run a microarray, RNA is isolated from a tissue sample, and the mRNA reverse transcribed to cDNA. The cDNA is made single-stranded and fluorescently labeled. The cDNA is then applied to the microarray and allowed to hybridize to any complementary DNAs on the array. Unbound cDNA is washed away and the array scanned for fluorescence. Each fluorescent spot identifies a gene expressed in the tissue sample. The results are quantitative. Thus, signal intensity correlates with the amount of mRNA in the tissue.

B. DNA microarrays can be used for diagnosis, monitoring therapeutic effectiveness, and assessing risk.
1. Example of a current clinical application for microarrays. Oncotype Dx for breast cancer was one of the first microarrays to enter the market for clinical use. It provides oncologists with prognostic information regarding the risk of breast cancer recurrence. It measures expression of 16 cancer-associated genes and 5 reference genes in patient samples. A score is generated based on which genes are expressed and the level of their expression in the tumor tissue.

   Gene targets measured by Oncotype Dx for breast cancer.
   Do not memorize these targets.
Scores from 0-18 indicate low risk of recurrence. These women are unlikely to benefit from chemotherapy. In fact, the risks associated with chemotherapy may outweigh any potential benefit. Scores of >31 indicate a high risk of recurrence and a benefit from chemotherapy.

Oncotype Dx has been well received by clinicians, and its use has resulted in a 30-40% decrease in under-treatment / over-treatment of breast cancer patients.

VI. Sequencing

A. Dye-terminator or chain terminator sequencing.

Chain terminator sequencing (a form of Sanger sequencing), is the most commonly used sequencing method. The sequencing reaction is similar to a PCR reaction, except there is only one primer (complementary to the template DNA strand) and the nucleotides (dNTPs) are spiked with a small amount of fluorescently labeled dideoxynucleotides (ddNTP, with a different color used for each of the four bases). The ddNTPs are incorporated readily into the DNA during synthesis, but because they have no reactive hydroxyl group on the sugar moiety, once incorporated, DNA synthesis stops. As only a small portion of the nucleotides are ddNTPs, they are randomly inserted into growing DNA strand. Because hundreds of thousands of DNA copies are made, a terminator is inserted at every base throughout the sequence. The sequencing reaction is run on a polyacrylamide gel and the labeled ddNTPs identified and recorded. The result given is the sequence of the complementary strand.
B.  Genetic diseases caused by DNA mutations can be diagnosed with DNA sequencing.
   1.  Cystic Fibrosis. Seventy percent of cystic fibrosis cases are caused by a 3 bp deletion in the gene for a chloride ion transporter called the cystic fibrosis transmembrane regulator (CFTR). The deletion results in the loss of a phenylalanine (F508) from CFTR. F508 is part of an alpha-helix, and loss of this hydrophobic amino acid results in protein mis-folding. The cell recognizes the protein as mis-folded and targets it for degradation.

C.  Next Generation Sequencing (NGS).

NGS is a “catch all” phrase referring to several different technologies (Illumina, Roche 454, Ion Torrent) that allow sequencing of millions of short DNA fragments simultaneously. NGS is preferred for large-scale genome analysis, because it is cheaper, faster, requires less DNA, and is more accurate than traditional sequencing.

1.  Illumina platform. This description is overly simplified, but is sufficient for our purposes. For more information, visit https://www.illumina.com/science/technology/next-generation-sequencing.html The video “How Does Illumina NGS Work” is well done, but has more details we require.

The Illumina method is based on chain terminator sequencing. Here the DNA to be sequenced is sheered into small pieces, and primers of a known sequence are added to the ends of each piece. The DNA is made single-stranded and then attached to a special slide called a flow cell. The synthesis reaction is unique in that each nucleotide has (1) an attached removable fluorescent molecule (a different color for each base) and (2) a removable chemical block on the 3’ hydroxyl of the sugar such that the nucleotide functions as a ddNTP. Sequencing is carried out as follows: A single-stranded DNA complementary to the
primer sequence is added so that it binds to each of the billions of pieces of DNA template immobilized on the flow cell. The fluorescently labeled “ddNTPs” and DNA polymerase are added and a nucleotide is added to the primer in the first position complementary to the template DNA. Only a single base will be added as the chemical block prevents addition of a second nucleotide to the DNA chain. The unincorporated “ddNTPs” are washed away and a camera takes a picture to register the color of the added “ddNTP”. The fluorescent label and the 3'-OH blocking group are then removed enzymatically and washed away. Now the nucleotide bound to the first position in the template is a functional dNTP – ready for the next base to be added that is complementary to the second nucleotide on the template. This series of steps is repeated over and over until the entire DNA fragment has been sequenced.

The power of NGS is that billions of pieces of DNA are sequenced simultaneously. Because they came from the same DNA sample, there are many copies of each DNA sequence bound to the flow cell. Thus, an NGS experiment will give multiple, independent sequencing results for the same DNA sequence. By lining up the results (see below), each base in the gene will have been sequenced many times. This is called depth of coverage. NGS is more error prone than traditional sequencing, but because of depth of coverage, NGS is more accurate.

Left: Individual clusters of PCR-generated DNA molecules using Illumina Sequencing.

Right: A computer will determine the nucleotide sequence of each of the DNA fragments and align the overlapping sequences. The results can be compared to a reference sequence and any differences identified. The depth of coverage in this example is 4 as the highlighted “T” was independently sequenced in 4 different DNA fragments. This base was a “G” in the reference sequence, but all four fragments show a “T” residue. Because the change was seen in most of the sequenced products, we are assured this is a real result documenting a mutation from a “G” to “T” in the DNA sample.
2. Key differences between traditional and NGS.

- **Faster**: NGS is faster because the chemical reaction is combined with signal detection (these are separate in traditional methods) and NGS generates data from thousands of reactions simultaneously (compared with one DNA fragment sequenced in traditional sequencing).

- **Accuracy**: NGS is more accurate. Because it relies on reading short overlapping fragments, each section of DNA is sequenced multiple times. This is called depth of coverage. Greater depth of coverage = more accurate data.

- **Less DNA**: NGS requires less DNA than traditional sequencing.

- **Cost**: NGS may cost more in terms of initial investment, but the reduced time, manpower and reagents mean NGS is more cost effective, particularly for large scale projects.

Scientific progress in sequencing technology has come a long way since the human genome project was initiated. The first human genome to be sequenced took 13 years compared to current NGS technology where it can be completed in 8 days for a fraction of the cost. However, it is still too slow and too expensive to transition into the clinical lab.

Oxford Nanopore Technology in the UK has developed a technology called the MinION. It identifies DNA bases by measuring the changes in electrical conductivity generated as DNA strands pass through a biological pore. It is portable, affordable, reads longer sequences than second generation NGS (a few hundred thousand bases compared to a few hundred base pairs) and is fast; thus, there is significant interest in its use in clinical applications. Current issues with transition into the clinic include an accuracy rate of 65-88% and issues with instability that slow data throughput.