Instructor

- Steve Wunderly
  - PhD. organic chemistry
  - Research Scientist for 32 years with emphasis on:
    - Organic synthesis
    - Molecular-biology
    - Nuclear detection
    - Regulatory/Quality/Radiation Safety
Acknowledgement

• Many of the graphics slides were taken with permission from Beckman Coulter training graphics or from the web site Graphics Gallery.

• *Graphics Gallery* provides a series of labeled diagrams with explanations representing the important processes of biotechnology.
Outline of Course

• Goal: To understand DNA, how the human genome was measured and how DNA is used in genetic engineering

• Pathway:
  • Understand what DNA is and its importance to how our body functions
  • How do we isolate and amplify DNA
  • How do we determine the sequence of DNA (human genome determination)
  • How do we engineer micro-organisms (following series)
Within every living organism are cells.
# The Cell - What is it?

- The Cell is the most basic functional unit of life
- It can be compared to a well planned city

<table>
<thead>
<tr>
<th>Workers</th>
<th>Proteins/Enzymes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Power Plant</td>
<td>Mitochondria</td>
</tr>
<tr>
<td>Roads</td>
<td>Actin fibers, microtubules</td>
</tr>
<tr>
<td>Trucks</td>
<td>Kinesin, dynein, myosin</td>
</tr>
<tr>
<td>Factories</td>
<td>Ribosomes</td>
</tr>
<tr>
<td><strong>Library</strong></td>
<td><strong>Genome (DNA)</strong></td>
</tr>
<tr>
<td>Recycling center</td>
<td>Lysosome</td>
</tr>
<tr>
<td>Police</td>
<td>Chaperones</td>
</tr>
<tr>
<td>Post office</td>
<td>Golgi apparatus</td>
</tr>
<tr>
<td>Communications</td>
<td>Signaling networks</td>
</tr>
</tbody>
</table>
A cell is the simplest reproductive element of life.

A cell with 6 pairs of chromosomes. (Humans have 23 pairs.)

Chromosomes are made up of long chains of DNA.
A single chromosome is like a book of recipes (the nucleus of the cell is like a library of books)

- A Gene represents an individual recipe
The chromosome is like a DNA “recipe book”
The gene is like a single recipe (example a protein)
Average of 1,000 ‘letters’ to make one gene “recipe”
A closer look at one portion of the gene “recipe”
What is a gene?

- A gene contains the genetic construction plan for an organism. The information in DNA consists of instructions how to produce proteins.
- So a gene is like a recipe composed of the DNA letters A, T, C, and G in a specific order. Just like English words depend on the specific order of letters for their meaning.
- Scientists have broken the “code”. We know which 3 letters (bases) code for each of the 20 amino acids.
DNA: What is it?

- **DNA,** Deoxyribose *N*ucleic *A*cid, is one of the fundamental molecules of Life. It is found in the nucleus of every living cell. It contains all the **information** (blueprints, instructions) for making all of the proteins in the body. It also contains the control levers for turning on and off the manufacturing line in the cell.

- **Proteins** are also fundamental molecules of life and are found throughout the body. They are the **building blocks and machinery** (enzymes) of the body.
DNA is like a twisted ladder. Sugar-phosphate spiral backbone make up the rails of the ladder. The rungs holding the two rails together are nucleic acids.
The DNA rungs are composed of long stretches of four chemical ‘bases’, A, T, C and G (rungs of the ladder).
DNA is a collection of subunits that makes a long molecule (polymer). The subunit is called a nucleotide.

The backbone of the molecule is a sugar-phosphate chain:

```
<table>
<thead>
<tr>
<th>BACKBONE</th>
<th>RUNGS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phosphate</td>
<td></td>
</tr>
<tr>
<td>Sugar</td>
<td>Base (A,C,T or G)</td>
</tr>
<tr>
<td>Phosphate</td>
<td></td>
</tr>
<tr>
<td>Sugar</td>
<td>Base (A,C,T or G)</td>
</tr>
<tr>
<td>Phosphate</td>
<td></td>
</tr>
<tr>
<td>Sugar</td>
<td>Base (A,C,T or G)</td>
</tr>
<tr>
<td>Phosphate</td>
<td></td>
</tr>
</tbody>
</table>
```

...
Three subgroups of the nucleotide.

- Phosphate

\(-\text{O-P-O-}\)
Nucleotide Sugars

PENTOSE
a five-carbon sugar

two kinds are used

β-δ-ribose
used in ribonucleic acid

β-δ-2-deoxyribose
used in deoxyribonucleic acid

Each numbered carbon on the sugar of a nucleotide is followed by a prime mark; therefore, one speaks of the "5-prime carbon," etc.
Third subgroup of the nucleotide is the nucleic acid (or base).

- **THE BASE** (The alphabet)
  - A = Adenine
  - T = Thymine
  - G = Guanine
  - C = Cytosine
- RNA substitutes U for Uracil in the place of Thymine

The order of the bases in the chain is of utmost importance!
The bases are nitrogen-containing ring compounds, either pyrimidines or purines.

Cytosine (C), Thymine (T), Uracil (U), Adenine (A), Guanine (G)
Nucleotides – the three components together

A nucleotide consists of a nitrogen-containing base, a five-carbon sugar, and one or more phosphate groups.

Nucleotides are the subunits of the nucleic acids.
More on the sugar component

- In any given nucleic acid, DNA or RNA, all the sugars are the same molecule
  - Nucleic Acids with Ribose are called RNA
    - *We will introduce the function of RNA shortly*
  - Nucleic Acids with Deoxyribose are called DNA
Another DNA representation

- Same structure occurs in simple single cell yeast as well as human cells
Protein: What is it?

- Proteins are very long chains of smaller sub-units called amino acids.

- Amino acids consist of a cluster or chain of carbon, hydrogen, oxygen and nitrogen atoms (example on next slide)

- There are only 20 amino acids found in the biological proteins of life.

- Both ends of the amino acid contains a reactive chemical group so that it can form chains. Long chains of amino acids are proteins.
Amine (ammonia)  Carboxylic acid (acetic acid)

Figure 4-2  Essential Cell Biology, 2/e. (© 2004 Garland Science)
Primary protein structure is the sequence of a chain of amino acids.
DNA short hand representation

• The double helical strands of DNA can be represented in multiple ways.
  • Strings of letters, where each letter represents a base connected to a sugar-phosphate strand

• AACTAGGTCTATCTTAGGCC - Single strand of DNA

• AGACTTACCGTTAACACATTG
  TCTGAATGCCAATTGTAAC - Double strand of DNA
DNA: Representations

- DNA can be represented by lines when the base order is not important to the teaching process.

- ______________________ Single stranded
- ______________________ Double stranded
- ______________________

- _____________CATCATCATCAT_________ single stranded DNA with region of interest specified
DNA CODE – the most incredible part

• There are 4 letters in the DNA language: A, C, G, T

• Three letters used together correspond to a word, which represents an amino acid

• There are $4^3 = 64$ possible combinations of four letters in sets of three.

• There are only 20 amino acids used to make proteins -- redundancy
DNA: What is it?

• The long strands of DNA code direct the order that amino acids are to be added to make proteins. Thus they are blueprints for all the proteins in the body.
• There is even a start sequence the defines the reading frame.

• GOD IS NOWHERE
• GOD IS NOW HERE
• GOD IS NO WHERE
• GOD I SNOW HERE
The DNA CODE

Khorana and Nirenberg, along with Robert Holley, won the 1968 Nobel Prize in Physiology or Medicine for their interpretation of the genetic code and its function in protein synthesis.
<table>
<thead>
<tr>
<th>First Base Position</th>
<th>All possible 2nd base Positions Here Last Base Position</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>T</strong></td>
<td><strong>C</strong></td>
</tr>
<tr>
<td>Phe</td>
<td>Ser</td>
</tr>
<tr>
<td>Phe</td>
<td>Ser</td>
</tr>
<tr>
<td>Lue</td>
<td>Ser</td>
</tr>
<tr>
<td>Lue</td>
<td>Ser</td>
</tr>
<tr>
<td><strong>C</strong></td>
<td></td>
</tr>
<tr>
<td>Lue</td>
<td>Pro</td>
</tr>
<tr>
<td>Lue</td>
<td>Pro</td>
</tr>
<tr>
<td>Lue</td>
<td>Pro</td>
</tr>
<tr>
<td>Lue</td>
<td>Pro</td>
</tr>
<tr>
<td><strong>A</strong></td>
<td></td>
</tr>
<tr>
<td>Ile</td>
<td>Thr</td>
</tr>
<tr>
<td>Ile</td>
<td>Thr</td>
</tr>
<tr>
<td>Ile</td>
<td>Thr</td>
</tr>
<tr>
<td>Met</td>
<td>Thr</td>
</tr>
<tr>
<td><strong>G</strong></td>
<td></td>
</tr>
<tr>
<td>Val</td>
<td>Ala</td>
</tr>
<tr>
<td>Val</td>
<td>Ala</td>
</tr>
<tr>
<td>Val</td>
<td>Ala</td>
</tr>
<tr>
<td>Val</td>
<td>Ala</td>
</tr>
</tbody>
</table>
How does DNA direct the manufacture of proteins

Step 1: the DNA strands separate

DNA Strands Separate

In preparation for the manufacture of mRNA, a DNA molecule in the nucleus separates into two strands in the region of a gene carrying instructions for a specific protein. Each sequence of three bases in a DNA strand is called a triplet, which is a code for one of 20 amino acids, the building blocks of protein.
Protein manufacture step 2 - mRNA is created from the DNA template by a set of enzymes

Transcription

One DNA strand acts as the template, or pattern, for the construction of mRNA. In this process, called transcription, free-floating RNA nucleotides traveling in the cell nucleus pair with complementary bases on the DNA template strand. RNA nucleotides use the base uracil (U) instead of thymine (T). As RNA nucleotides pair with DNA bases, uracil (U) from the RNA pairs with adenosine (A) on the DNA strand,
Step 3 - mRNA carries message of DNA to protein building machines (ribosome)

Messenger RNA Binds to Ribosome

Once mRNA is completely formed, the mRNA strand leaves the cell nucleus to enter the cytoplasm, where it attaches to a cellular organelle called a ribosome. Protein synthesis occurs in the ribosomes.
Step 4 External to the nucleus of the cell are tRNA molecules. They carry a specific amino acid on one end and RNA code on the other end. These link with the mRNA in the Ribose.
Step 5 - Ribosome enzyme the connects the tRNA with the mRNA and squeezes the amino acids together to form a protein strand.

Translation

After a tRNA binds to an amino acid, it carries the amino acid to the mRNA-ribosome complex. The anticodon of the tRNA binds to a codon on the mRNA. The sequence of bases in the codon code for the type of amino acid carried by the tRNA. A second tRNA attaches to the mRNA-ribosome complex. The first tRNA transfers its amino acid to the amino acid of the second tRNA before detaching from the
Step 6 - The process continues coupling multiple amino acids together.

The ribosome continues to move the mRNA strand as the polypeptide chain is built. The polypeptide chain is completed when the ribosome comes to an mRNA codon known as the stop codon, which instructs the ribosome that the manufacture of the protein is finished.
Protein formation complete

Released from the ribosome, the newly formed protein is an exact replica of the structure encoded in the original DNA strand.
I just told you the results of vast experiments – how were the results obtained?

- Isolating **sufficient quantity** of DNA and **sequencing** it was a major hurdle solved by two brilliant ideas.
- The first hurdle was resolved by a Nobel Prize winning technique called Polymerase Chain Reaction or PCR for short.
- And second hurdle, by an award winning sequencing strategy developed by Sanger and Coulsen
PCR

- keywords
  - primer
  - template
  - Nucleotides
  - polymerase

- 3 steps in process
  - denature
  - anneal
  - Extend

- This **breakthrough** process is key to sequencing, fingerprinting and genetic engineering
Polymerase

- Protein, enzyme, that adds building blocks of nucleotides to form a chain. It is an enzyme that makes a polymer.
  - DNA polymerase, RNA polymerase,

- In order to form a polymer chain the building blocks (monomers) must have two reactive sites on each end of the molecule

- DNA polymerase joins individual nucleic acids (building blocks) of DNA together
Nucleotide and Template

- **Nucleotide** – The basic building block of DNA it is a molecule consisting of Base-Sugar-Phosphate

- **Template** – The DNA chain under experimental study. It is the target gene or other portion of DNA to be studied
**Primer**

- Short single stranded DNA (small piece of synthetic DNA 17 to 30 nucleotides), synthesized by automated synthesizer, machine/instrument
- The primer matches the initial portion of a strand of DNA in the area or gene of interest. (the template or target)
- Prior knowledge of sequence of the primer DNA is required as well as the sequence of the target DNA.
The goal of PCR is to **increase quantity** of a **specific sequence** DNA so that it can be studied.
Step 1: Denature – Separating the strands

• Heat the DNA sample isolated from a biological source in solution containing primer, nucleotides and polymerase.
• Heat causes the double stranded DNA to separate into single strands.

-----------------------------------
heat

-------------------------------

Step 2: Anneal- laying down the primer to the desired target or template

The reaction begins when a primer lays down on a DNA Template

Solution contains many different strands of DNA
Step 3: Extension – making the rest of the 2\textsuperscript{nd} strand

An enzyme then assembles a chain of bases that corresponds to the bases on the DNA Template.
Extension Continues template is copied

The process is repeated many times
Review of the PCR cycle

- The process is controlled by changing the reaction temperature and consists of **3 steps**:  
  - **Denaturation** (96 degrees C - 20s), separates chains  
  - **Annealing** (50 degrees C - 20s), attaches primer  
  - **Extension** (60 - 70 degrees C - 4min), activates enzyme

- To generate enough copied DNA for detection, we repeat the process 30-60 times.
POLYMERASE CHAIN REACTION

1. DNA is denatured. Primers attach to each strand. A new DNA strand is synthesized behind primers on each template strand.

2. Another round: DNA is denatured, primers are attached, and the number of DNA strands are doubled.

3. Another round: DNA is denatured, primers are attached, and the number of DNA strands are doubled.

4. Another round: DNA is denatured, primers are attached, and the number of DNA strands are doubled.

5. Continued rounds of amplification swiftly produce large numbers of identical fragments. Each fragment contains the DNA region of interest.
The award winning Sequencing Process
This is the process that was used to determine the entire human genome

• Modify the PCR method to randomly terminate the extension phase with fluorescent dideoxy nucleotides

• Separate each distinct chain length and detect by the fluorescent marker
Sequencing Process-Four Steps

- Cycle sequencing
  - Denature double stranded helix
  - Anneal primer to template
  - Polymerase binds complex and extends the primer to match the template.
  - Fluorescently labeled dideoxy-nucleotides, ddNTPs, randomly stop extension
What is a Dideoxy-base

- A Dideoxy-base is a DNA building block with only one hook.

- The second hook is replaced with a non-reactive visualization (or flag) molecule.
Extension with random termination
Termination Ends Replication

A dideoxy base prevents the amplification from going any further
Termination Products – fragments of DNA

When we run enough reactions we get a series of DNA copies
• each has a fluorescent dideoxy base at the end

27 BASES

20 BASES

13 BASES
Final results – a fragment at each chain length with a termination flag to detect the end.

We repeat the process 30-60 times to produce enough DNA pieces to detect.
The Sequencing (Detection) Process

• Separation of the multiple DNA chains of different length by the process of gel electrophoresis.

• This process separates the DNA chains on the basis of size (length).

• The motion of the DNA fragment molecules is driven by application of voltage across the gel, driving the molecule to the positive charged end.
Gel-filled Capillaries or plates

- The physical gel called **LPA** (*linear polyacrylamide*).
Separation of fragments by size

Gel electrophoresis

LPA

detector

(Andy Vierstraete 1999)
Capillary Electrophoresis
Base Calling

Each dideoxy base is identified by its unique fluorescent color

G A T C T

Detection Window
Gene assembly with the ultimate result the whole human genome

Assembling a gene

(Andy Vierstraete 1999)
Genetic engineering of DNA (modify the sequence of the DNA in the Gene)

Bacterial DNA

____________________________

Cut the bacterial DNA

_________              ___________

Insert desired Gene into bacterial DNA

_________ insulin Gene_________

The bacteria now makes insulin
Genetic Engineering

The next two sessions, Jan 11 and Jan 18 will focus on the techniques of genetic engineering and its applications to synthetic biology.
Nucleotides

G Guanine

C Cytosine

A Adenine

T Thymine

U Uracil

replaces Thymine in RNA