28

BIOCHEMISTRY

Faculty of Medicine – JU2018

◉ Sheet ○ Slides

DONE BY

Dena Kofahi

CONTRIBUTED IN THE SCIENTIFIC CORRECTION

Ibrahim Elhaj

CONTRIBUTED IN THE GRAMMATICAL CORRECTION

Ibrahim Elhaj

DOCTOR

Dr. Mamoun

Note: This sheet will discuss the last two methods from the "Protein Analysis" slides and then continue with the beginning of the "Enzyme-Based Molecular Techniques" slides.

Protein Analysis

We will discuss two techniques that are used to determine protein structure. It is very important to know the structure of a protein because once you know the structure you can:

1. Design drugs against the protein.
2. Modify existing drugs – Remember the enzymes COX1 and COX2. These enzymes were both targeted by Aspirin. Knowing their structure allowed us to design a specific drug, Celebrex, to target only COX2.
3. Predict the structure of other proteins using their primary sequence – This means you may be able to predict the presence of a motif, domain, or active site from the primary sequence on one protein based on what you already know from the structures of other proteins.

1) X-ray Crystallography

Video Link Dr. Mamoun placed in the slides:

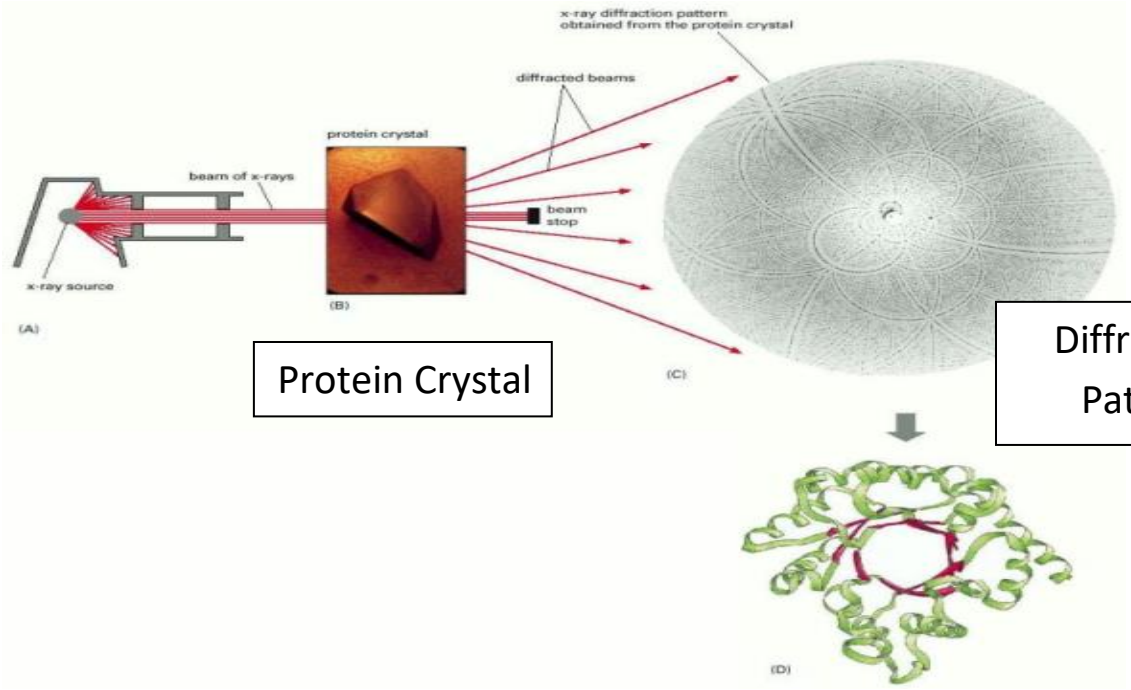
http://www.dnatube.com/video/279/Protein-Structure-Revealed-xray-crystallography

X-ray crystallography is a technique used to determine the three-dimensional structure of proteins. The basic steps of the process are:

1. The sample must first be purified to have one single protein in the sample (by single we mean one specific type *not* only one molecule).
2. This purified sample is placed in a freezer for a long time to allow the sample to crystallize.
3. The small crystal is then exposed to a beam of x-rays. When the x-rays hit the atoms of the crystal, the electrons cause the x-rays to scatter and create diffraction patterns on a metal sheet. This pattern can then be used to determine what the structure of the protein is.

Issues with X-Ray Crystallography:

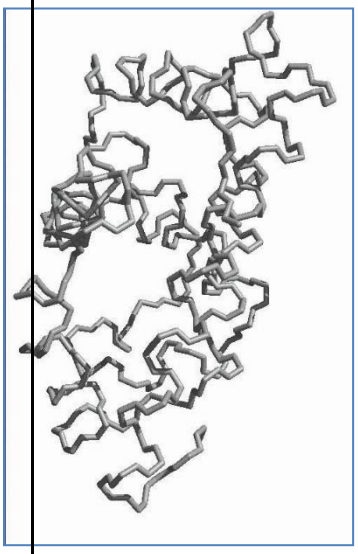
1. It is a laborious, tedious technique, which means it takes a lot of effort and time.
2. When you take a protein out of its natural environment, the protein loses its structure (an example is a protein that normally exists in a hydrophobic environment like the plasma membrane). Therefore, this technique will not necessarily give the complete, accurate structure of the protein, but it is a good start.



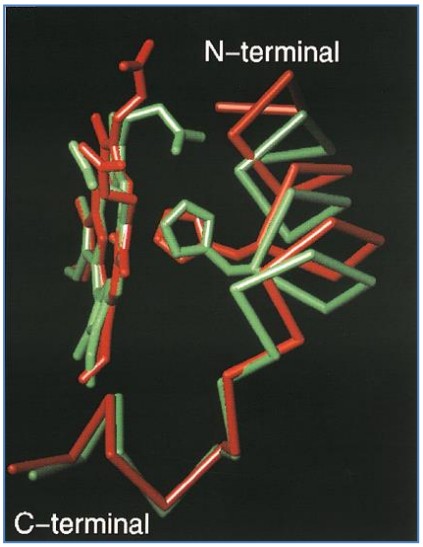
2) Nuclear Magnetic Resonance (NMR) Spectroscopy

- Nuclear magnetic resonance (NMR) spectroscopy reveals the structure and dynamics of proteins in solution (not crystals).
- So NMR has the advantage of placing the protein in a solution similar to its natural environment so the result is more accurate.
- This is important for proteins that bind to other molecules like enzymes to their substrates, receptors to their ligands, etc. This is because you can see the change in the structure of the protein when the molecule binds to the protein and when it is released.

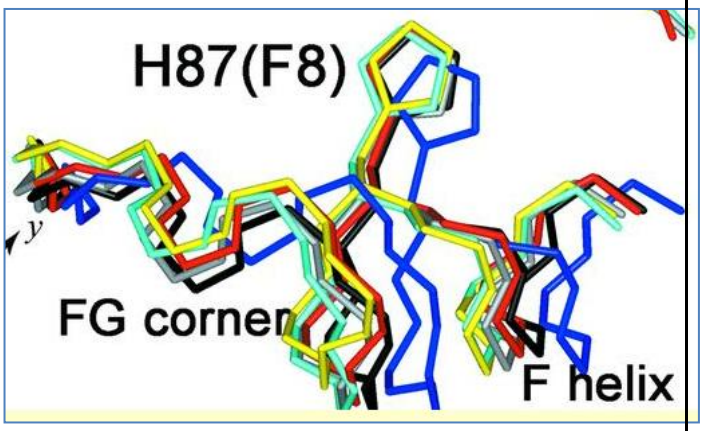
For proteins like hemoglobin and myoglobin, these two techniques (both crystallography and NMR) can allow you to see the change in structure when oxygen binds to the molecules. These techniques are very specific, so you can even see when atoms shift by a few angstroms. This works by superimposing the different structures on top of each other so you can see how the different atoms move.



Myoglobin



Myoglobin and Hemoglobin



Hemoglobin – Notice how the different structures are superimposed to see the atomic movement.

Enzyme-Based Molecular Techniques

DNA Sequencing

- DNA sequencing is the process of determining the exact order of nucleotides in a genome.
- <u>Importance:</u>
  - Identification of genes and their localization – Genes can be identified by looking for consensus sequences, like promoters. Localization means on what chromosome is the gene located on.
  - Identification of protein structure and function – You can determine the mRNA transcript that would be transcribed from the DNA sequence and then determine the amino acid sequence that would be translated from the codons. As we said earlier, once you know the primary transcript of the protein, you can predict the structure and function of the protein.
  - Identification of DNA mutations – Done by comparing the DNA sequences of the normal and abnormal individuals.
  - Genetic variations among individuals in health and disease – Such as between ethnic groups
  - Prediction of disease-susceptibility and treatment efficiency – Disease susceptibility: what diseases their genome indicates they are at a greater risk of having. Treatment efficiency: Like what medication the patient should or shouldn't take.
  - Evolutionary conservation among organisms

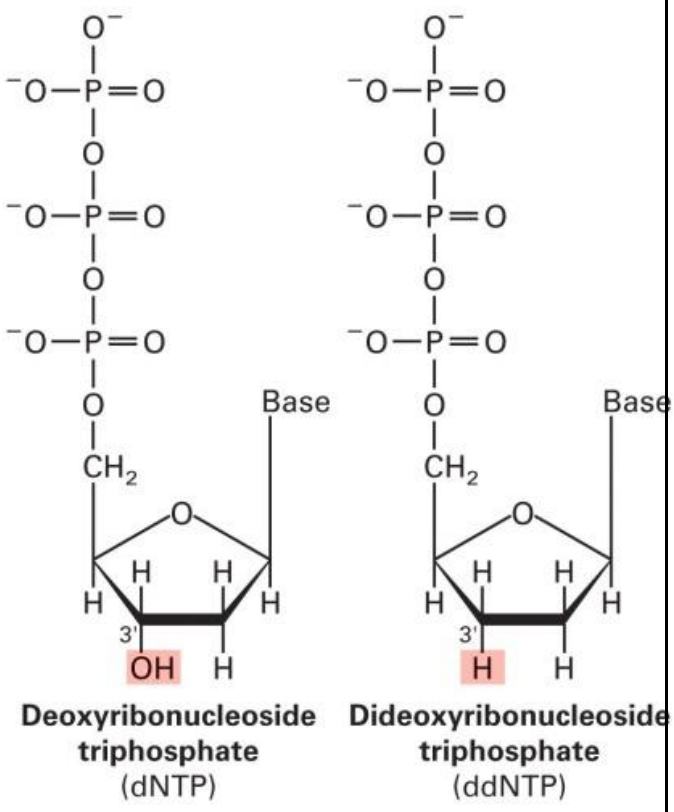
- History of DNA Sequencing of Different Organism's Genomes: **(We do not need to know this, but it is here for your information)**
  - Viruses and prokaryotes first
  - Then human mitochondrial DNA
  - The first eukaryotic genome sequenced was that of yeast, Saccharomyces cerevisiae.
  - The genome of a multicellular organism, the nematode Caenorhabditis elegans.
  - Determination of the base sequence in the human genome was initiated in 1990 and completed in May 2006 via the Human Genome Project

Note: There are two charts in slides 6+7 that are not included in this sheet. The professor said there is no need to memorize any of the information in the charts. Just know that there is a relationship between the number of base pairs and the number of genes

Methods of DNA Sequencing

This sheet will discuss two methods of DNA Sequencing.

1) An old technique based on the premature termination of DNA synthesis by dideoxynucleotides.



<u>The Basic Idea</u>

- Deoxynucleotides are the monomers of DNA and they have a hydrogen instead of an OH group at carbon #2.
- Dideoxynucleotides have <u>*two*</u> hydrogens that replace OH groups, one at C2 and one at C3.
- In normal DNA replication, the next nucleotide is added at the OH group of the 3' end.
- However, when a dideoxynucleotide is added, DNA replication is forced to stop because there is no longer the needed reactive group (OH) on the 3' end.

The Process

Dideoxynucleotides can be used to our advantage in sequencing with their ability to stop DNA synthesis. This method has **four separate reactions** occurring, and each will need:

1. DNA polymerase
2. The 4 deoxynucleotides (dATP, dCTP, dGTP, dTTP)
3. The template strand
4. Primers – The primers are labelled with a radioisotope so that the primer can give a signal.
5. **One** specific dideoxynucleotide – This is why we have four separate reactions, each one will have their own dideoxynucleotide. One reaction will have ddGTP, one ddATP, one ddTTP, and the last ddCTP.

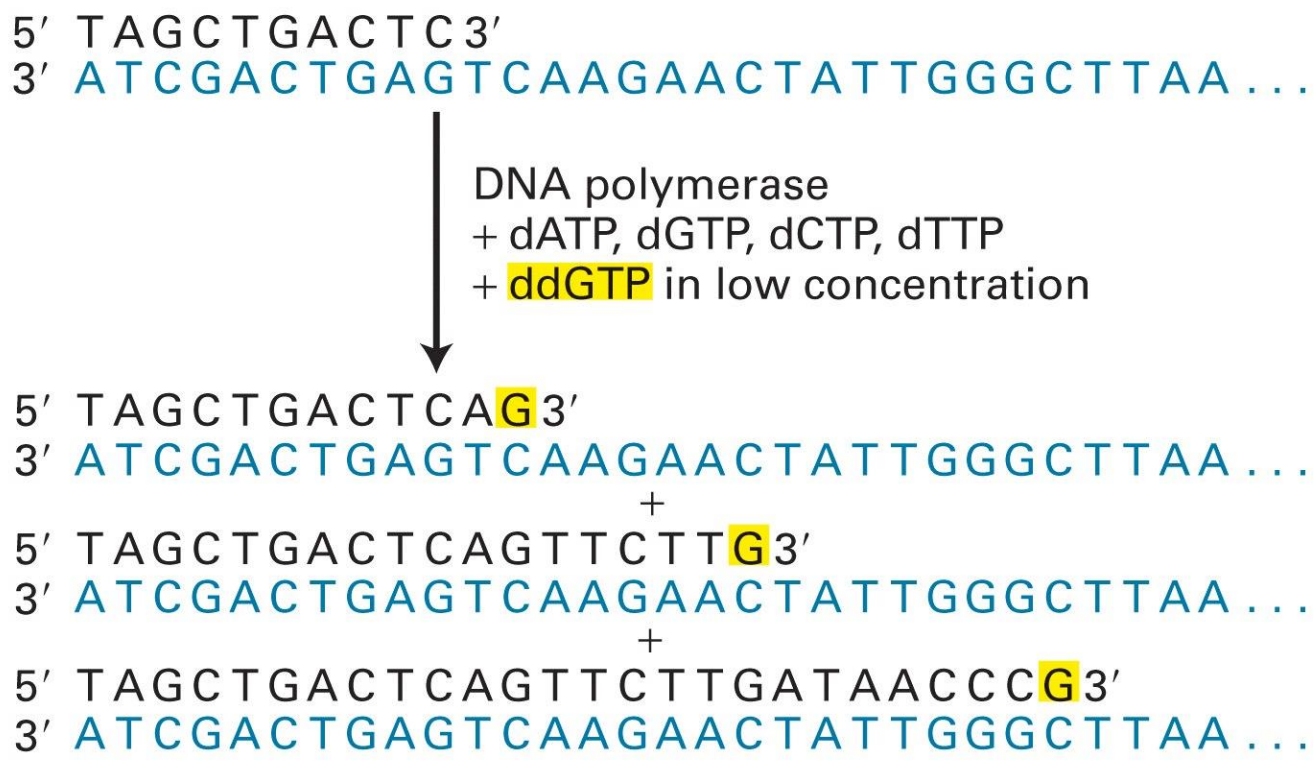
There are two things you should note:

a) Since we need a primer, we need to have some prior knowledge on the sequence of the template strand so we can know what specific primer we should use.
b) Remember in these reactions, we do not have only one DNA strand, or one primer, or one DNA polymerase in the test tube, but thousands of each.

Now let's get into the actual process:

1. Synthesis and Generation of Fragments

- Four separate reactions are run, each including deoxynucleotides, plus one dideoxynucleotide.
- DNA synthesis is initiated from a primer that has been labeled with a radioisotope.
- Then the polymerase can continue adding complementary deoxynucleotides until a dideoxynucleotide is added. Incorporation of a dideoxynucleotide stops further DNA synthesis because no 3' hydroxyl group is available for addition of the next nucleotide.
- A series of labeled DNA molecules are generated, each terminated by the dideoxynucleotide in each reaction.



Let's refer to this figure to see how this part of the process works. In this reaction the dideoxynucleotide is ddGTP. DNA polymerase keeps adding nucleotides until it sees a C. Now, there are two probabilities. Either it adds a normal dGTP, or a ddGTP. If it adds a ddGTP, DNA synthesis is terminated. If it adds a dGTP, DNA synthesis continues until DNA polymerase reaches the next C and the same two probabilities are faced again.

Let's say we started with 1000 strands of the DNA transcript. A portion of the strands would have DNA synthesis stopped with a ddGTP added to the first C. It may be 100 strands, so there would be 900 strands left. Then replication continues on these 900 strands. Once the DNA polymerase reaches the next C, it faces the same two possibilities again. If 100 strands have their DNA synthesis terminated, then 800 remain. This continues on until all strands have their synthesis terminated.

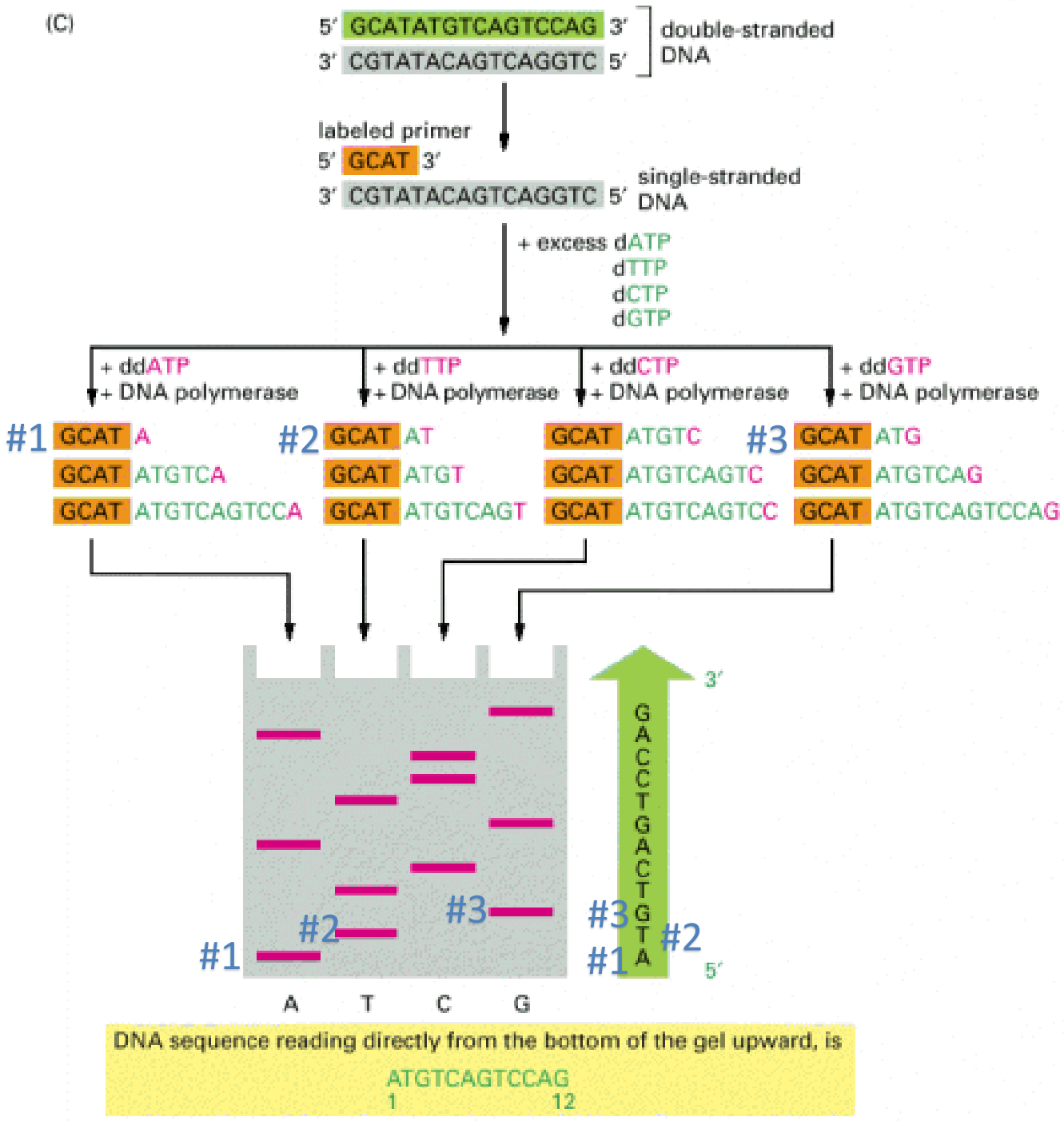
What is the result of all this? We will have multiple DNA fragments of different lengths. Some will differ by only one nucleotide. Next, we will see how we can use gel electrophoresis to determine the sequence of DNA using these fragments.

Note: Notice that in the previous figure it says "ddGTP in low concentrations." The reason why we need ddGTP in low concentration is because if we had a high concentration of ddGTP in comparison to the other nucleotides, the ddGTP will be able to compete more strongly with dGTP, which means that it will immediately stop replication in the beginning, and we won't have the same generation of fragments of different lengths.

Basically, the low concentration makes sure that replication doesn't always stop all synthesis at the beginning.

2. Gel Electrophoresis and Sequencing

- These fragments of DNA are then separated according to size by gel electrophoresis and detected by exposure of the gel to X-ray film
- The size of each fragment is determined by its terminal dideoxynucleotide, so the DNA sequence corresponds to the order of fragments read from the gel



Remember that the primer that is used is labelled (it gives a signal). So all these strands of different lengths give a signal.

We take each one of the four reactions and place them into their own well. Acrylamide gel is used. This gel is known for having **high resolution**, which means that it can separate DNA fragments even if they differ by only one nucleotide.

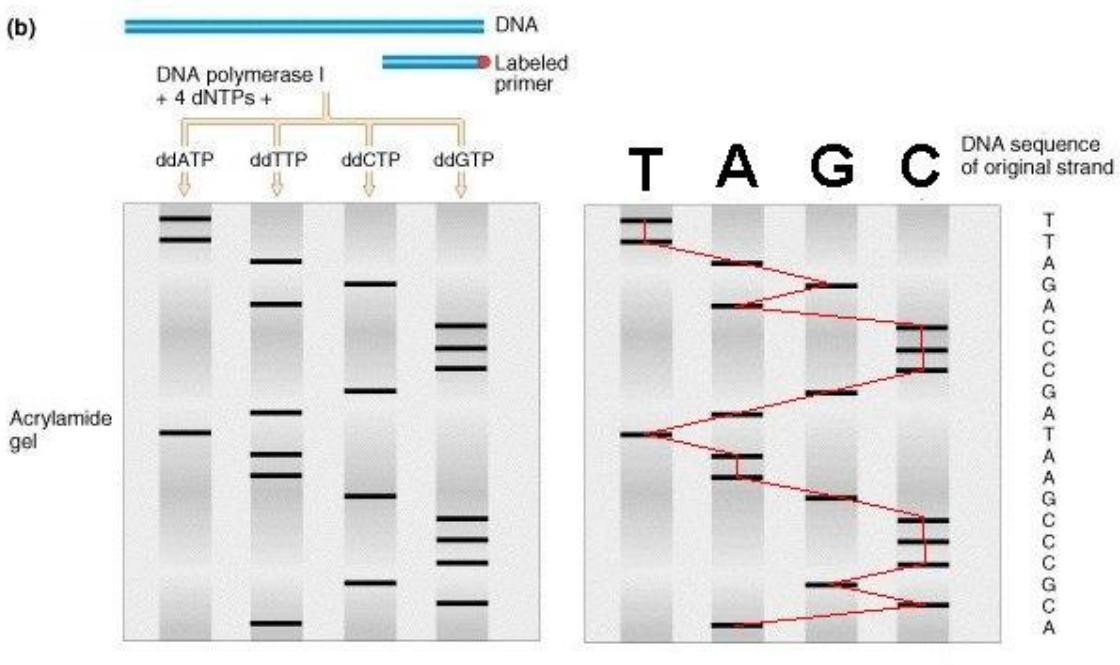
*(Agarose gel is specific for horizontal analysis)

Since the primers are labelled, we can expose the gel to X-ray film to see the placement of each fragment. The fragment at the bottom is the smallest as it travels with the highest speed. The fragment after it differs by only one nucleotide, and since the gel has high resolution, it will travel slightly slower so that the difference can be seen on the gel. This continues on for all the strands, with the largest strand being the topmost strand, because it is the strand that moves the slowest.

In order to know the DNA sequence, we start with the smallest fragment from the bottom then read to the top. Why the bottom? The one at the bottom is the one with the nucleotide closest to the 5' end, and remember the DNA sequence is read from 5' to 3'. So if the fragment at the bottom is in the well for A, then the first nucleotide is A. Then if the strand above it is in the well for T, then the next nucleotide in the sequence is T. Then you continue this process for the rest of the fragments.

But don't forget that this is the sequence of the newly synthesized strand! To know the sequence of the original template strand, you need to do the opposite because, remember, the two strands of DNA are complementary but antiparallel (the strands run in opposite directions of each other). So, the 5' end of the template strand is at the top, and the 3' end on the bottom.

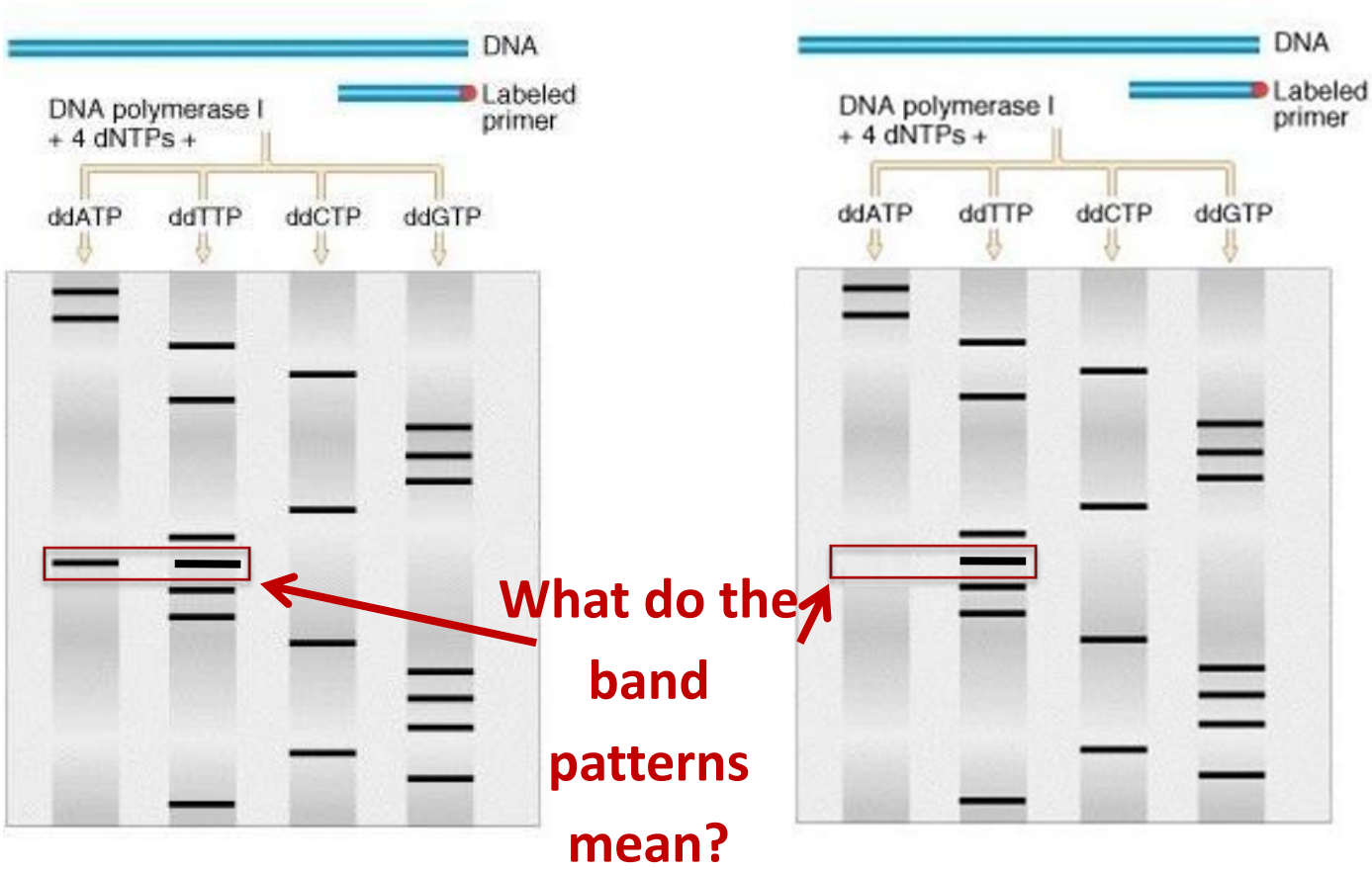
Below is a visual for how to read the strand. You can use the figure on the left to read the synthesized strand from the bottom (5') to the top (3'). On the other hand, you can use the figure on the right to read the template strand from the top (5') to the bottom (3'). In the end, no matter the figure, you can follow this rule to determine the sequence of either the newly synthesized strand or the template strand.



Remember that we are diploid; we have two copies of every chromosome. So the primer binds to the DNA you got from your mother and the DNA you got from your father. As replication occurs, the majority of the nucleotides added should be the same, because human DNA is highly similar. However, we can be heterozygous at a certain location. So we can have different DNA sequences on the same position on the two different chromosomes. It can be a single nucleotide polymorphism, a mutation, or some other variation. It can indicate a disease (like sickle cell disease), or we can compare the results of someone who has and someone who does not have a disease.

How does this appear on the gel? **(refer to the figures on the next page)**

a) If we see that we have two fragments, in two different reactions (wells), that have exactly the same size **(red box in figure on the left)** then this person is heterozygous for that nucleotide at that position. They have two different nucleotides at the same place in the sequence. So, as in the figure below, on one of their chromosomes they have an A at that position, and on the other chromosome there is a T at that position.
b) If we know that the heterozygous version exists, yet we see that someone has only one band **(red box in figure on the right)** at that specific position, we know that the person is homozygous.



This whole technique, however, is not used anymore to sequence DNA. This is because of a few issues:

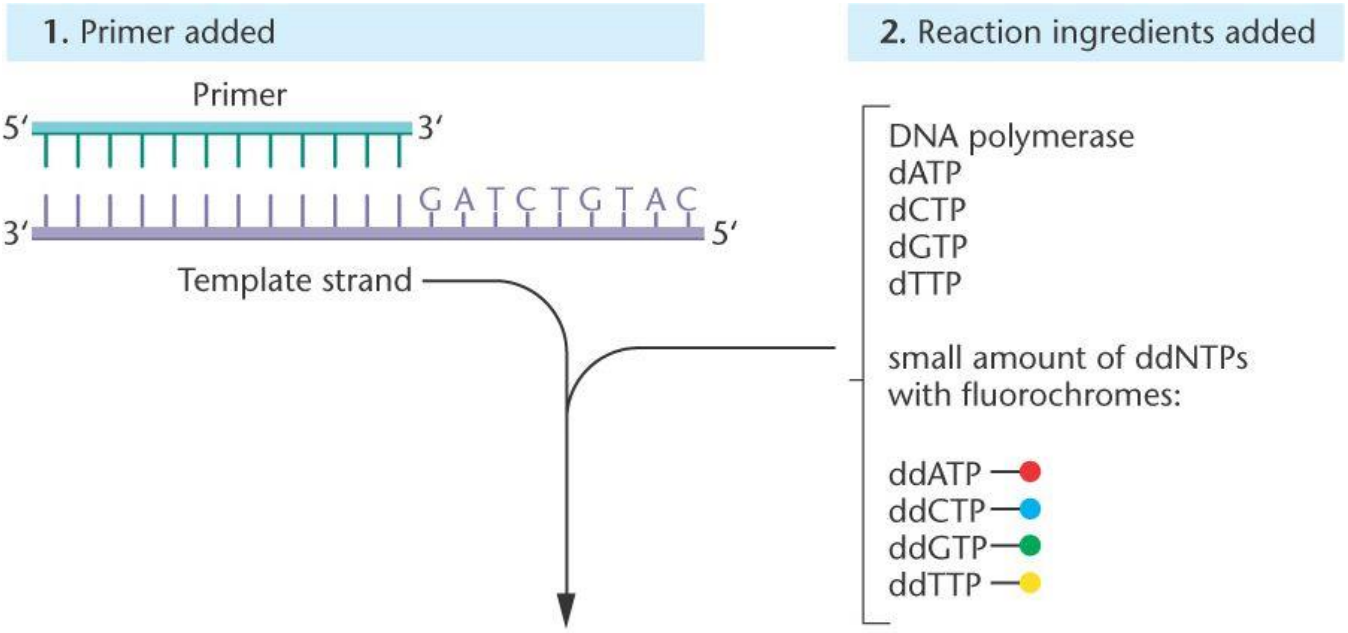
a) This process cannot be automated – Scientists prefer to have computers do all the work, and in this process you have to manually determine the sequence, which takes too much time.
b) The primers are labelled radioactively – this is dangerous because radioactivity can lead to mutations.

This led to the creation of fluorescence-based sequencing.

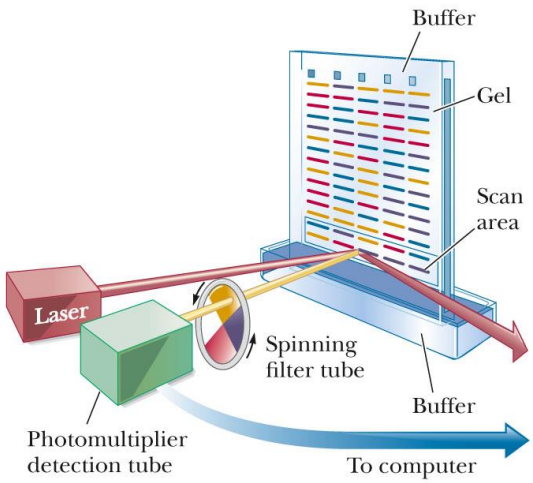
2) Fluorescence-Based Sequencing

This method needs the same exact materials: the template strand, the primer, the DNA polymerase, and the four normal deoxynucleotides. The difference is that the dideoxynucleotides are each labelled with their own **unique** fluorescent tag, so that each ddNTP would give its own color. Because of this, all four reactions can be minimized to one single reaction and everything can be placed in one test tube.

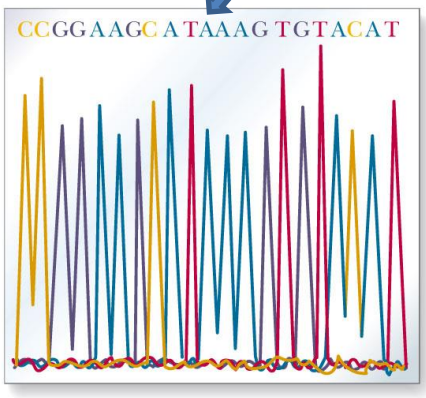
Note that when the primer is added, we can say that the primer is *annealed*, or *hybridized* to the template strand.



3. Primer extension
Chain termination
Product recovery



Chromatogram

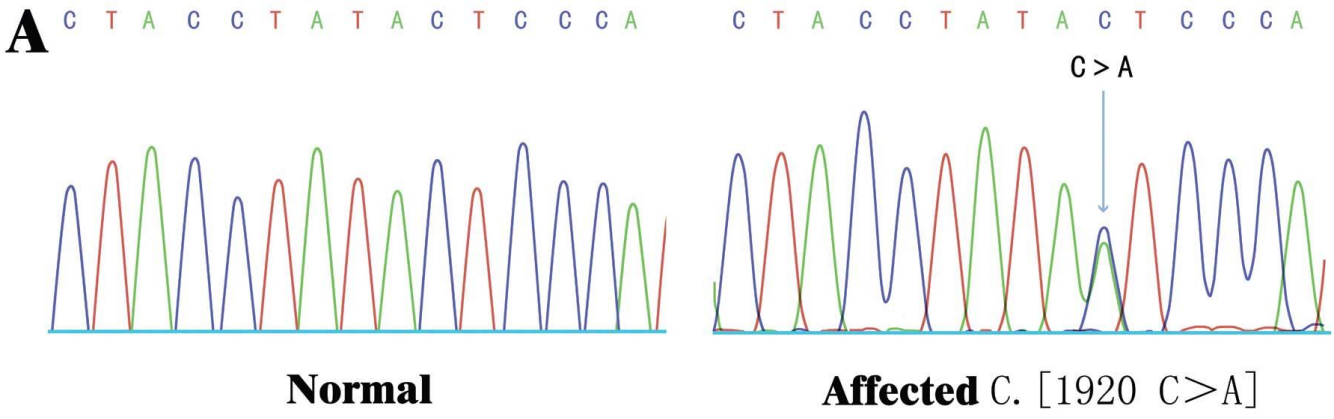


The result of synthesis is the generation of different fragments of different sizes that differ by only one nucleotide. These fragments each give a certain signal based on the dideoxynucleotide that terminated the reaction. So, let's say that ddCTP gives a yellow signal, then any fragment that had its synthesis terminated by ddCTP will give a yellow signal.

This sample can then be added to the gel, and can be added to only one lane, and the fragments will be separated according to size. Each fragment will differ by one nucleotide from the one before it. As the fragments migrate down the gel, there will be a detector that can read the color signal, and then the computer can translate this signal into a letter. The final result is a chromatogram, which the computer can read to give the sequence.

Analyzing the Chromatogram

On the chromatogram, each peak represents a signal. Each color corresponds to a specific nucleotide. When you view the chromatogram, there should normally be one specific peak for each position. There are two cases, however, that we must consider.



1. Two Colors on the Same Position (as seen in the chromatogram above)

If there are two colors at the same position, the computer is going to read both, and show that there are two nucleotides at the same position on the DNA sequence. This simply means that the individual is heterozygous.

2. A Whole Color Change

Let's say one sample gives a blue signal at a specific position seen on its chromatogram. Another sample gives a green signal at that same position on a different chromatogram. This indicates that there is a change, or a mutation that occurred at this nucleotide, as the two samples indicate two different DNA sequences.

Comparing The Two Methods

Early Termination By Dideoxynucleotides	Fluorescence-Based Sequencing
Labelled with radioactivity	Labelled with fluorescent tags
The primer is labelled	The dideoxynucleotides are labelled
Only one signal is emitted (radioactivity emits the same signal no matter the molecule)	4 different signals (colors) are emitted depending on the dideoxynucleotide
4 different reactions in 4 different test tubes	1 reaction in 1 test tube
Reactions need to be placed in 4 different wells	The reaction can be placed in one single well
Manual	Automated

Goodluck 😊