Sanger Method: Determining the Termination

In 2003 the Human Genome Project project was completed. This event signified that the thirteen-year long, publicly funded, international effort to determine the DNA sequence of the human genome finished two years before it was estimated to be. While this was very exciting and began our efforts of understanding the inner workings of our genome, the real work began decades earlier.

DNA sequencing is the process of determining the sequence of nucleotide bases (As, Ts, Cs, and Gs) in a piece of DNA. **Sanger sequencing** was developed by the British biochemist Fred Sanger and his colleagues in 1977. It was the first practical method for determining the order for the nucleotides of a sample of DNA and remained the dominant method for thirty years.

But first - memory dump: The Sanger Method follows roughly the same protocol as PCR. In the space below, write down everything you remember about PCR. From how it works, to what the reagents were - anything that comes to mind, place it in the box below.

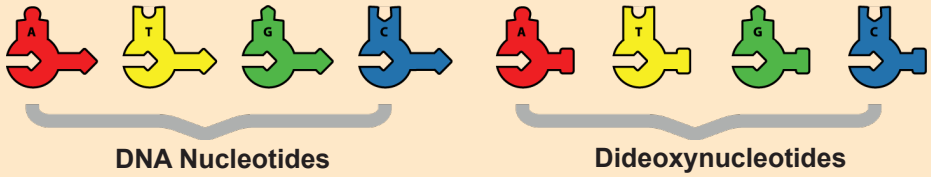
Goal: To model how Sanger Sequencing builds on the biochemical processes involved in DNA replication and how the applications of DNA sequencing have impacted industry.

Step One: Assemble the Unknown Sequence

Build the unknown sequence.

Unknown Sequence	3’		A	C	T	G	T	G	G	A	C	T	G	A	G	G	A	C	A	T	A	3’
Primer & Nucleotides	5’		G	A	C	_	_	_	_	_	_	_	_	_	_	_	_	_	_			5’

Now, while you “know” the unknown sequence for the sake of this activity, in the lab, Sanger Sequencing is used to determine the unknown sequence.



DNA Nucleotides **Dideoxynucleotides**

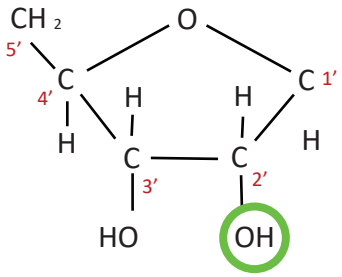
Explore the differences between the foam pieces of the DNA nucleotides and dideoxynucleotides. Compare/ Contrast their shapes. Why do you think the foam pieces were designed this way?

Sanger Method: Determining the Termination

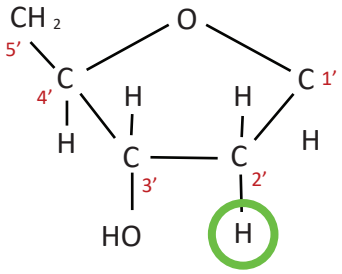
Fred Sanger was a biochemist. Before working on DNA and RNA, he investigated the structure and function of protein. He understood the chemical reactions required for elongation of a nucleotide sequence. With that in mind, he proposed a way to utilize that knowledge to terminate a sequence.

Examine the three structures of sugar molecules; ribose, deoxyribose, and dideoxyribose. Like all models, the ones you are using today don't fully visualize all differences in the biochemistry of the molecules. Consider the chemical structure of the molecules and how that will impact the function of the molecule, especially when bound to phosphates and a nitrogen base.

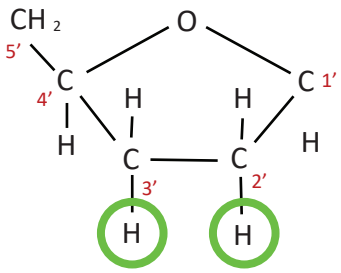
Ribose
found in RNA



Deoxyribose
found in DNA



Dideoxyribose
synthetically produced for Sanger Method



- What is the structural difference between **Ribose** and **Deoxyribose** ? Think back, what implications did that have in the differences between DNA and RNA?

- Now, examine the structure of **Dideoxyribose**. How will that impact binding?

- Think ahead, you are Fred Sanger. You have just synthetically produced dideoxyribose and produced the four dideoxy nucleotides, (ddATP, ddTTP, ddGTP, and ddCTP). What is your next step? How do you use this to sequence DNA?

Sanger Method: Determining the Termination

Steps

1. Grab your needed materials
   - Nucleotide bags
   - Primer Sequences; placed to the side in a stack
2. Using the **Unknown Sequence** as a guide and following **nucleotide base pairing rules**:
   - First add the primer to the 3' end of the unknown sequence.
   - Then, use the nucleotide bags to add free nucleotides to the 5' end of the primer.
3. When you pull a dideoxynucleotide, **ddNTP**, where N stands for any one of the nucleotide bases, then that strand is complete. Separate that single strand from the Unknown Sequence.
4. Repeat this process, starting first with adding a primer to the 3' end of the Unknown Sequence and then add nucleotides as you did in step three.

Look at the strands you created. Use this space to make observations and write down any questions that come up along the production of these strands. Where do you think this process will go next? What is the "end goal" of the Sanger Method?

Models are only as good as your memory. Try your best to draw at least two of the strands that you produced. Label the: **primer**, **deoxynucleotide (dNTP)**, and **dideoxynucleotide (ddNTP)**.

Sanger Method: Determining the Termination

What Just Happened?

The Sanger Method follows the same generalized protocol as PCR. Both of these protocols are run in a thermocycler at very specific temperatures.

- Think back. How did the PCR reaction work? What temperatures were used? Why were they used? What enzyme (polymerase) was utilized? Why did it have to be this enzyme, not a "regular" human DNA Polymerase III?

- Is it logical that these two protocols would be similar in nature? Why/Why not?

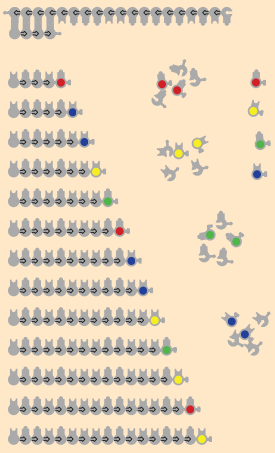
- In truth, Kary Mullis, who is credited with inventing PCR while working at the Cetus Corporation, did not "discover" PCR until 1983, six years after Fred Sanger invented PCR in 1977. How might Mullis have utilized Sanger's protocol and experimental findings in his creation of PCR? If you were Mullis, what information would you be most curious about looking into?

Looking at these two images of the models for both Sanger Sequencing and PCR to answer the following question:

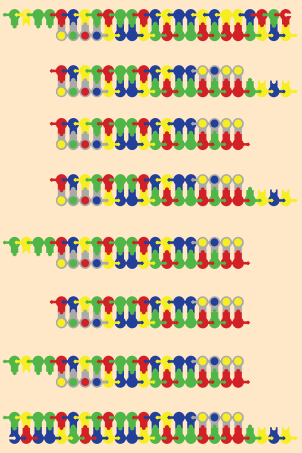
- What are the differences between the products of these two protocols?

- How do the differences relate to the scientific or investigative purpose of the protocol/experiment?

Sanger Sequencing



PCR



Sanger Method: Determining the Termination

What Happens Next?

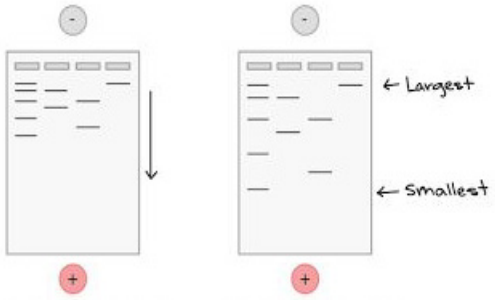
The Sanger Method utilized **chain terminated sequences** to determine the sequence of DNA in a specific fragment. Look at all of the single stranded, chain terminated sequences that you have produced.

- Why is the Sanger method also known as the **chain termination method**?

- Propose a way to organize these strands that makes the most logical sense. Use the lines below to explain how/why you organized your bands. Draw a sketch of how you organized it in the blank area.

- What would happen if these fragments were run on a gel? How would you organize the gel / utilize the wells in the gel so that the result could be used to give you the sequence of bases in the DNA sample?

Remember, with DNA electrophoresis, the smaller fragments will be able to travel faster through the pores in the agarose gel. This will mean that they will appear farther down the gel than the longer fragments that will travel slower.



Sanger Method: Determining the Termination

Using the space below, create both a table in the space below that organizes your data as well as a predicted gel for your results.

Now, utilize both the physical model and the picture that you just produced to answer the following questions:

- Given your model, what is the sequence of the DNA that you analyzed, from 5' to 3'?

  _____

- What questions would a scientific investigator still be faced with after your gel would run? How could we change the procedure to make sure all of the questions were answered?

  _____
  _____
  _____
  _____
  _____
  _____
  _____
  _____

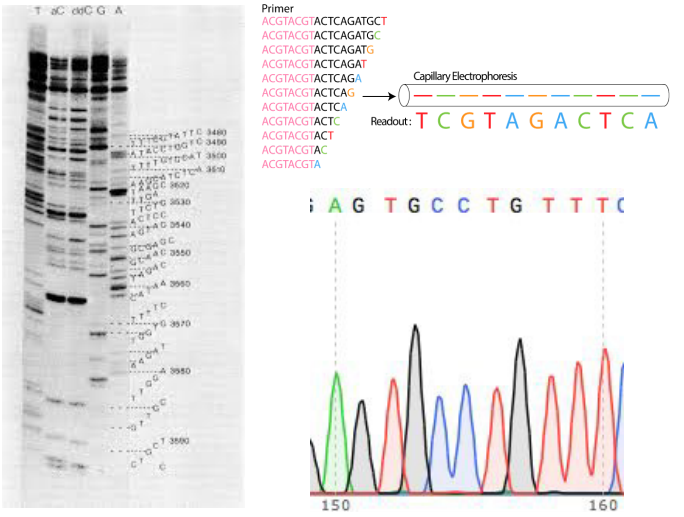
- What if you got multiple copies of a single chain termination sequence?
  - How do you model this with your physical model?
  - How would this appear on the gel?

  _____
  _____
  _____

Sanger Method: Determining the Termination

Next Generation Sequencing

The Sanger form of sequencing was widely used for forty years. However, it was not without issues. The Sanger Method works for small sections of DNA (>500bp), but with scientists looking to sequence larger genomes, it became clear that the scale and efficiency of each step within the protocol needed to be evaluated and investigated for possible improvements. The 1990s had many transformations for sequencing protocol, including an automated, color-coded method. Dideoxynucleotides were synthetically treated once again to have a fluorescent tag added to them. Now dubbed **dye-terminator sequencing**, the overall protocol remained the same, however, instead of the final products being run on a gel, they were now read by a computer. This process, called **capillary electrophoresis**, greatly improved the quality of the sequencing data as well. **Chromatograms** were now utilized to represent the output of sequencing data. When looking for patterns in data, this new form of output greatly aided scientists in their quest to understand genomes of various organisms.



***Clockwise from above**: Gel electrophoresis was the first way that sequencing data was portrayed to scientists. Capillary electrophoresis not only eliminated the pouring of gels, but also simplified both the extraction and interpretation of fluorescent signals given off by dyed ddNTPs. Chromatograms displayed the results in an easy to read graph that both color coded the bases as well as displayed the sequence data*

- Looking at a Sanger gel, why do you think scientists moved away from this practice?

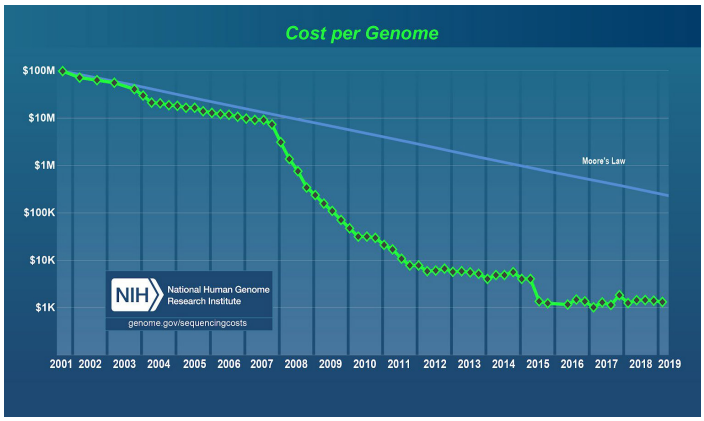
- Sanger sequencing is still utilized today for small bp length samples. Why do you think this is the case?

- What are some other conditions/required improvements that you believe scientists wanted to improve on with genome sequencing? What infrastructure or technological improvements would be needed to make these gains?

Sanger Method: Determining the Termination

Faster is not Better... Enough

As with everything in life, we search for ways to continue improving. DNA sequencing was not void of this intrepid mindset. One of the biggest drivers with this progression has been the cost. It is clear the scope and scale of genomics research is driven by the associated costs of understanding the genome itself. In 2001, the Human Genome Project (HGP) cost **$2.7 billion** over the **13 years** it took to sequence. Currently, the price, as seen in the NIH graph to the right, hovers around **$1000**, a considerable cut in costs from 2001. In addition, a whole genome can now be sequenced in as little as a day.



In the summer of 1986, James Watson called a meeting at Cold Spring Harbor Laboratory located on the North Shore of Long Island, NY. One of the biochemical researchers there was Kary Mullis.

- What did Mullins work on?

_____

- How could his research help initiate / propel the HGP?

_____

_____

- The first genomes sequenced were of small, model organisms. In fact, in 1998 C. elegans was the first animal genome completely sequenced. Why would scientists advocate for model organisms to be sequenced first? How does this parallel pre-clinical trials?

_____

_____

_____

In May 2000, President Bill Clinton grew tired of the apparent arms race between the two groups racing to sequence the human genome and the money/tense nature it was creating. He sent a two word note to a member of the Department of Energy saying simply, “Fix this!”. On June 26, 2000, a press conference was conducted to announce the completion of the HGP. In truth, the project would not finish until February 2001 when competing articles were simultaneously published in both Science and Nature.

- Why would the President want this “arms race” cut short?

_____

_____

_____