Sanger sequencing – a hands-on simulation

Student Handout

Jared Young

Correspondence concerning this article should be addressed to Jared Young, Mills College 5000 MacArthur Blvd, Oakland, CA 94613

Contact: jyoung@mills.edu

Introduction

The dideoxy method of DNA sequencing (also known as the "Sanger" sequencing method after the main developer of the method) is an important technique: it revolutionized the field of Genetics and is still in wide use today. It is also a beautifully designed procedure. In this exercise, you will carry out a simulation of this method that will enable you to appreciate how the steps of the procedure enable determination of a DNA sequence.

Learning Outcomes

After completing this exercise, you should be able to do the following:

- Explain the process of Sanger sequencing and how it produces sequence information.
- Explain the process of DNA synthesis *in vitro*, including the role of template and primers.
- Explain the process of nucleic acid electrophoresis.
- Interpret a chromatograph readout of a Sanger sequence result.

Instructions

The simulation has two phases. In the first phase, you will carry out a DNA synthesis reaction. In the second phase, the class gathers to put the synthesized DNA through gel electrophoresis.

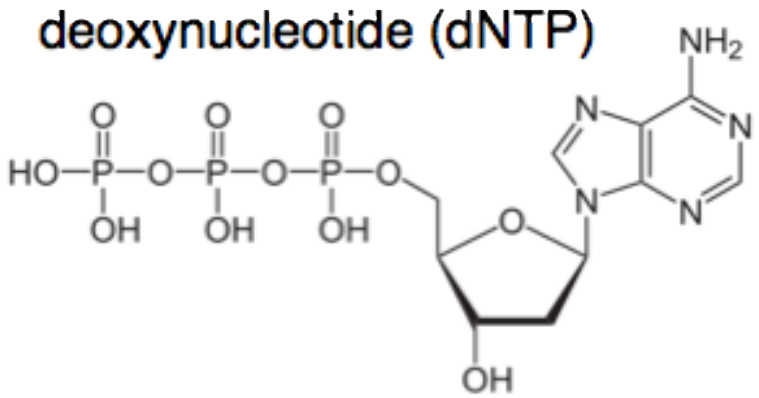
You will work with a partner. Each pair should get the following:

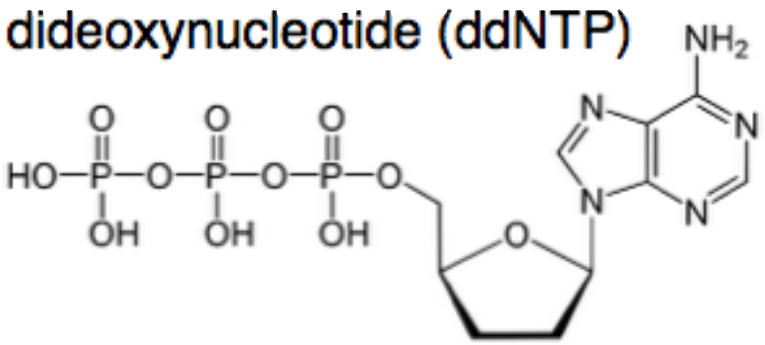
- One six-sided die
- One piece of paper with two identical double stranded DNA template strands printed on it (see below)
- The following Unifix cubes. Each block represents a nucleotide.
    - 5 each of brown and white cubes (representing C and G dNTPs)
    - 8 each of maroon and black cubes (representing A and T dNTPs)
    - 2 each of the 4 bright colors (representing C, G, A and T fluorescently-labeled ddNTPs)

Lecture

Follow along with the lecture, filling in the handout when prompted.

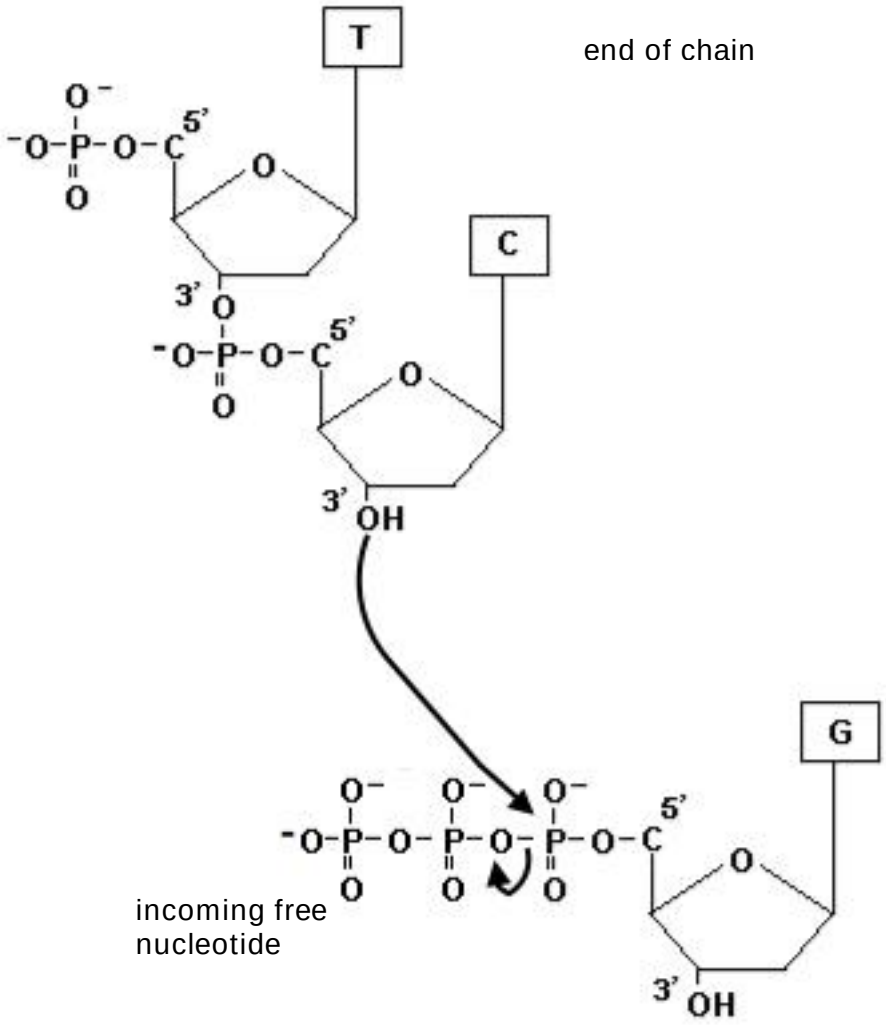
PROMPT 1: In the nucleotides to the right, there is one difference in chemical structure. Circle this difference.





PROMPT 2: The consequence of the difference circled above is that dideoxynucleotides act as chain terminators during DNA synthesis. Pair up, and have the person sitting on the left side of the partnership explain to the person on their right why dideoxynucleotides act as chain terminators. In the explanation, make references to the structure of DNA, to the covalent bonds formed during DNA synthesis, and to the picture on the right.

Once the first person has finished their explanation, the person who was explained to should turn to the person on *their* right (not the person who just explained to them), and give their own explanation of the same concept. (If you are on the extreme right end of a row, loop around and explain to the person at the extreme left end of the row.)



Phase One – DNA synthesis

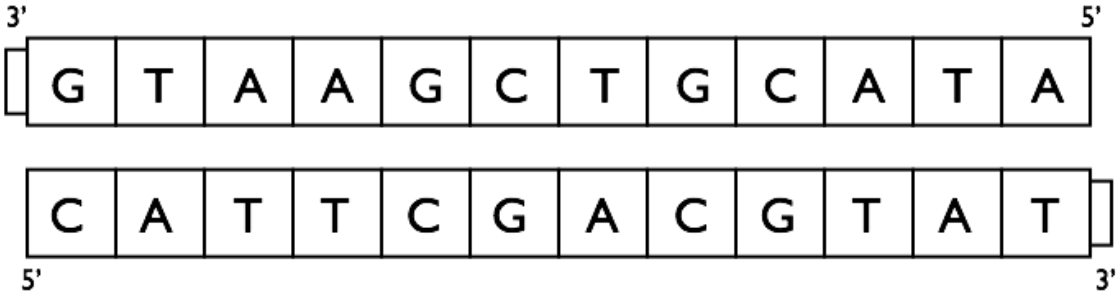
The first phase of Sanger sequencing is a DNA synthesis phase. We will model this with the following representations:

- Each student is playing the role of a DNA polymerase enzyme.
- Each Unifix cube is a specific nucleotide.
- The ddNTPs are brightly colored because they have a fluorescent dye covalently bonded (the dNTPs do not).
- The printed paper represents the template DNA that we will determine the sequence of.
- The entire classroom is a test tube filled with templates, dNTPs, dye-labeled ddNTPs, and DNA polymerases. The environment within the tube provides the right conditions for the DNA polymerases to be active.
- Also included in the tube are primers. Make two of these with the sequence 5'ATT 3', out of dNTPs.

Note: there are many copies of template and primer in this tube, but all templates have the same sequence and all primers have the same sequence.

Note: the model is conceptually accurate, but not quantitatively accurate for practical reasons. In a real Sanger sequencing reaction, the primer would typically be around 20 nucleotides (the same as in PCR). The ratio of dNTP:ddNTP varies from around 10:1 to 300:1, depending on desired read length, buffer conditions, the polymerase used, and the electrophoresis conditions. Read lengths around 800-1200 nucleotides are routinely achievable with today's technology.

PROMPT 3: Anneal your 5'ATT3' primers to your template. Where will they anneal? Indicate where the primer will anneal on the image below and discuss your reasoning with your partner.



Carry out synthesis simulation

DNA synthesis will now proceed, catalyzed by you, in the role of DNA polymerase. In this role, you will faithfully add, at each position, the nucleotide that is complementary to the partner nucleotide on the template strand. However, each base that is added could be **either the dNTP or the ddNTP** that is complementary to the template base at that position. Which one is incorporated is random, but it is more likely to be the dNTP because of the relative concentrations present in the tube. You will model this probabilistic process using die rolls.

For each position:

- Roll die
- If 1-5 is rolled, incorporate a deoxynucleotide at the next position
- If 6 is rolled, incorporate a dideoxynucleotide at the next position

Note that when a dideoxynucleotide is incorporated synthesis of that chain is terminated!

Synthesize two new strands (synthesize one until it is terminated, then denature those two strands and synthesize a second, using another 5'ATT3' primer).

PROMPT 4: Write down the sequences of the two strands produced by you and your partner.

Strand 1:

Strand 2:

Phase Two - Electrophoresis

You will now electrophorese your synthesized DNA on a "denaturing gel" (denaturing gels disrupt weak bonds, rendering DNA single stranded and lacking intramolecular base pairing).

PROMPT 5: Draw a picture of a gel with a loaded well, thus representing what is being modeled.

Voltage will be switched on, and strands electrophoresed.

PROMPT 6: In mid-electrophoresis, before any strands have reached the bottom of the gel, draw a picture of the gel with the strands migrating through, thus representing what is being modeled.

A flourescence detector reports the intensity of the colors associated with the four dyes at a rapid sampling rate. This color report is analyzed by a computer program that identifies nucleotides based on the color reported.

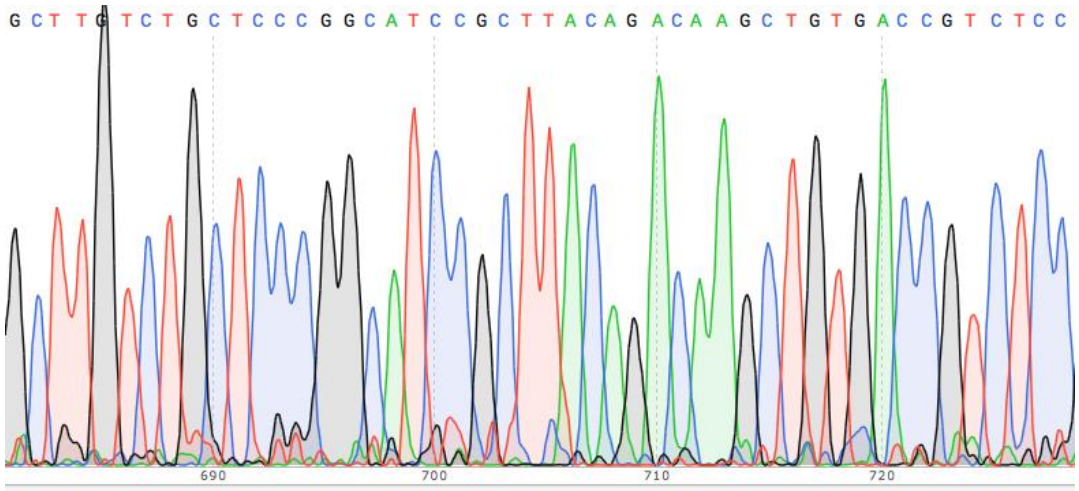
PROMPT 7: As our fluorescence detector/computer calls out nucleotides, record them here:

Now compare the sequence read by the detector with the template sequence that was originally present in the tubes in the synthesis phase. How do they compare?

Questions for Discussion

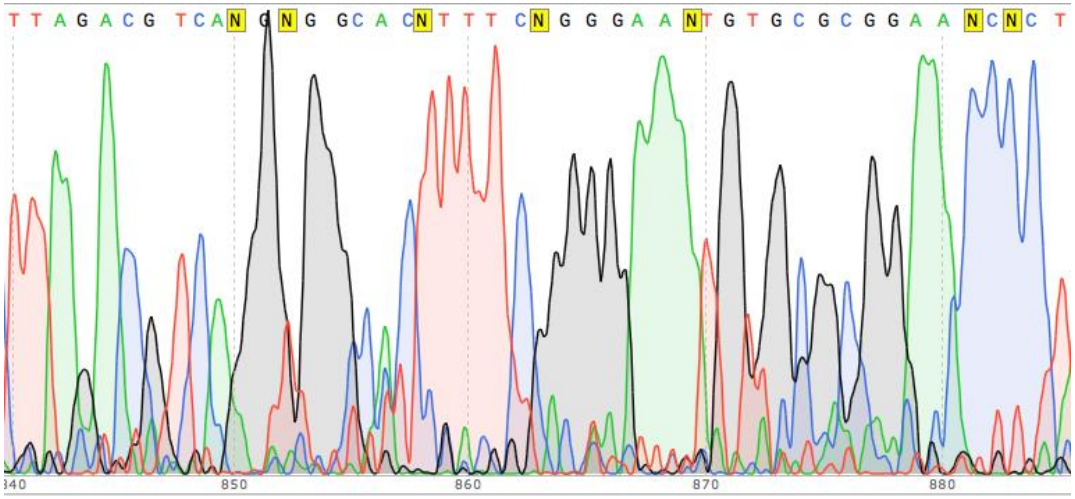
You will now form discussion groups. Within your group, discuss each of the questions below. Make sure to encourage participation from all in the group, and give everyone space to articulate what they understand and what they are confused about. This kind of discussion is an effective way to learn. Ultimately, come to agreement on a response to each question (with assistance from instructors/TAs as needed) and express your responses in this handout.

The image below, called a chromatograph, is a data display of a Sanger sequencing result. Notice how the peaks are distinct and regularly spaced, and the computer is making nucleotide calls, which appear above the peaks. At the bottom is listed the position, downstream of the primer, that corresponds to these peaks (you are looking at a stretch of sequence about 700 nucleotides downstream from the primer).



CHROMATOGRAPH A

Below is another chromatograph, taken from the same Sanger sequencing run as the image above. Notice how the peaks are less distinct, and the computer is struggling to make calls at several positions (in such cases an "N" call is returned).



CHROMATOGRAPH B

Examine chromatograph B. Consider whether you think you can make nucleotide calls at the loci where the computer returned an "N".

Now consider that you are the scientist who produced this data, and your purpose in doing this sequencing run is to check whether there are any mutations in a plasmid you are cloning. Here is what the sequence of the plasmid is supposed to be, without any mutations, for the region represented in chromatograph B:

ttagacgtca ggtggcactt ttcggggaaa tgtgcgcgga acccct

Note that the spaces that separate every 10 bases are only for visual convenience; there are not actually gaps between those bases.

DQ1: Provide your interpretation of the Sanger sequencing results shown in chromatograph B. Would you conclude that the sequence is correct, or that there are mutations? How certain are you about your conclusion? Explain your reasoning.

DQ2: It is typical for Sanger sequencing reactions to return clear, interpretable fluorescence curves through the first 800-1000 bases, then for the data to get increasingly ambiguous. Provide a hypothesis for why this is the case and explain your reasoning.

DQ3: Propose a follow-up experiment that should provide a less ambiguous sequence for the region represented in chromatograph B.

In practice, the template is a much longer sequence than that presented in the simulation we acted out – it could be a plasmid of thousands of bases, or an entire genome. Yet it is imperative that only a single region be sequenced.

DQ4: Where does specificity come from in terms of which locus is sequenced?

DQ5: Why is it imperative that only a single region be sequenced? If two regions are synthesized during the same sequencing reaction, what problem does this create?

DQ6: Some chains in our simulation did not incorporate a ddNTP before coming to the end of the sequence. Do you think this would be an issue in a real Sanger sequencing procedure? Explain your reasoning.

DQ7: What would you see in the chromatograph if no strands of a specific length are generated during the synthesis reaction (e.g. there are no strands that are 585 nucleotides long)? In practice, how do you think that issue is avoided?

DQ8: The dNTP/ddNTP ratio needs to be high for a successful reaction (typical ratios are 10:1 up to 300:1). What would the problem be a low ratio were used, say to 1:1?

DQ9: Consider the sequencing result if the template sample is the genome of an organism that has variation in the sequence being determined. That is, the genome of an individual who is heterozygous at that locus. What would the chromatograph look like if the two variants differ by a single nucleotide substitution? What would the chromatograph look like if the two variants differ by a single base pair insertion? Explain your reasoning in both cases, incorporating drawings in your explanations.