Sanger sequencing – a hands-on simulation

Background and Guidelines for Instructor

Jared Young

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

Contact: jyoung@mills.edu

Synopsis

This hands-on simulation teaches the Sanger (dideoxy) method of DNA sequencing. In the process of carrying out the exercise, students also confront DNA synthesis, especially as it relates to chemical structure, and the stochastic nature of biological processes. The exercise is designed for an introductory undergraduate genetics course for biology majors. The exercise can be completed in around 90-minutes, which can be broken up into a 50-minute period for the simulation and a follow-up 50-minute (or less) period for discussion. This follow-up could also take place in a Teaching Assistant (TA) led section. The exercise involves interactions between student pairs and the entire class. There is an accompanying student handout with prompts that should be invoked where indicated in these instructions.

Introduction

Sanger sequencing is an important technique: it revolutionized the field of Genetics and is still in wide use today. Sanger sequencing is a powerful pedagogical tool well-suited for inducing multiple "aha" moments: in achieving a deep understanding of the technique, students gain a better understanding of DNA and nucleotide structure, DNA synthesis, the stochastic nature of biological processes, the utility of visible chemical modifications (in this case, fluorescent dyes), gel electrophoresis, and the connection between a physical molecule and the information it contains. Sanger sequencing is beautiful: a truly elegant method that can bring a deep sense of satisfaction when it is fully understood.

Because it incorporates many different concepts, Sanger sequencing is a fairly complicated topic to learn. It is also difficult to fully comprehend from words and pictures. In this exercise, students act out the dynamic processes of DNA synthesis and gel electrophoresis, forcing them to think deeply about what is happening and discover for themselves the essence of the process.

Approach/Method

Prior knowledge

Students should have already learned the structure of DNA and nucleotides, and the biochemical reaction involved in DNA synthesis. They also should already understand gel electrophoresis. Understanding of DNA replication processes beyond the action of DNA polymerase (such as those mediated by topoisomerase, helicase, SSBs, leading vs. lagging strand etc.) is not needed. A priming reading/assignment in which they review the DNA replication reaction would be useful.

Set up

There are two parts to this exercise: a simulation, followed by a discussion. The simulation has two phases. In the first phase (DNA synthesis), students work alone or in pairs. In the second phase (electrophoresis), the whole class gathers to "run the gel." This exercise as described works best with a class of around 30-50 students, but it can be modified for a smaller or larger class. These modifications are described later.

The first phase works well in pairs, so that students have a partner to work/consult with. It can also be done individually. I will refer to pairs in these instructions. Each pair receives the following:

The following Unifix cubes (Figure 1); Each block represents a nucleotide.

- 5 each of two dull colors (representing C and G dNTPs)
- 8 each of two other dull colors (representing A and T dNTPs)
- 2 each of 4 bright colors, with an additional obstruction at the male end (representing C, G, A and T fluorescently-labeled ddNTPs)
- One six-sided die (these can be cheaply purchased at Amazon.com)
- One piece of paper with two identical double stranded DNA template strands printed on it (Figure 2)
- Two copies of the Student Handout

Note: these numbers work for the template provided and save money by limiting the number of blocks needed. To provide a more mathematically accurate representation of the simulation, each pair would receive 10 each of all 4 of the dull colors (this would reflect the concentrations simulated by the probabilities of the dye rolls – there are 5x more of the dNTPs than the ddNTPs).

Figure 1. Unifix cubes

Unifix cubes are plastic stackable cubes designed for teaching math. They work well for this exercise because they are a convenient size, are easy to stick together and take apart, and the 5’ and 3’ ends of the modeled DNA chain are fairly intuitive to identify. Unifix cubes can be purchased at Amazon.com or at classroom supply stores. Other similar colored linking units could be used, such as pop beads or Lego blocks. It is useful to add a small obstructing object on the male end of each ddNTP cube (in this image, tape flags are used) to highlight the chain termination feature of the ddNTPs.

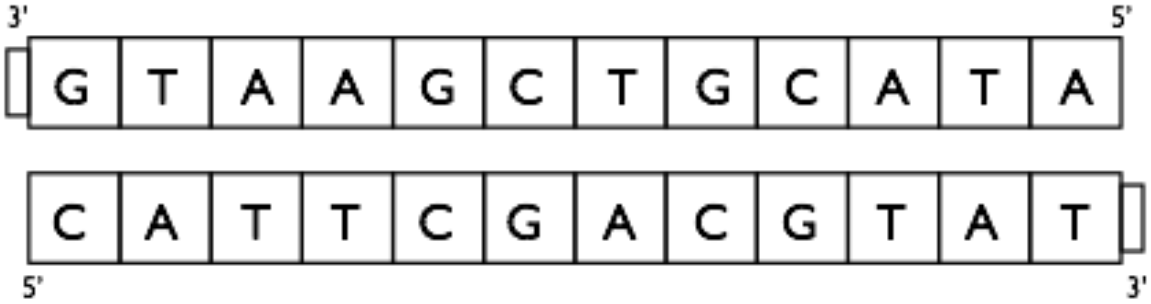

Figure 2. Template DNA

One double stranded DNA molecule is represented on a piece of paper. The images on the paper match the size and shape of a chain of Unifix blocks; chains of these blocks will be generated during the exercise.

Background information

Prior to the exercise, introduce Sanger sequencing, discussing its history, utility, and importance. Introduce the dideoxynucleotide as a chain terminator, using this as an opportunity to review nucleotide structure and the DNA synthesis reaction.

The publication of the structure of DNA by Watson and Crick in 1953 established DNA as the molecule of inheritance, with information stored in its sequence of bases. From this moment forward, stimulated by the development of the central dogma of molecular biology (information flow from DNA to RNA to protein), intense effort has been devoted toward developing techniques to determine the sequence of nucleotides in DNA. Important advances were made in the 1960s and mid-1970s, but the real game changer was published in 1977 by Fred Sanger (Sanger et al., 1977). What came to be known as dideoxy sequencing, or "Sanger sequencing" was the first practical method for sequencing, a method so robust that it dominated DNA sequencing for thirty years, serving as the main generator of sequence for the publication of the Human Genome in 2001. At that time, new "next generation" techniques began to emerge that have replaced Sanger sequencing for large-scale sequencing projects, but Sanger sequencing is still widely used for routine sequencing (for more on the history of DNA sequencing, see McGinn and Gut, 2013 and Heather and Chan, 2016).

Sanger sequencing involves a DNA synthesis reaction with both typical deoxynucleotides (the ones used in nature in DNA synthesis) and dideoxynucleotides (Figure 3 – show animation 1 after prompt 1 is completed, and animation 2 after prompt 2 is completed).

PROMPT 1 *(Students indicate the difference between a dNTP and a ddNTP)* These dideoxynucleotides differ from typical deoxynucleotides in that they have two hydrogens connected to the 3' carbon of the ribose instead of a hydrogen and a hydroxyl. Thus, they are "twice deoxygenated", missing hydroxyls on both the 2' and 3' carbons. Dideoxynucleotides can be substrates in a DNA synthesis reaction and can be

added at the 3’ end of a growing DNA strand (Figure 3). However, once a dideoxynucleotide is incorporated into the chain, the chain can no longer be extended. This is because another nucleotide cannot be added to a dideoxynucleotide due to the missing hydroxyl on the 3’ carbon.

PROMPT 2. *(Students explain to partner why ddNTPs act as chain terminators)*

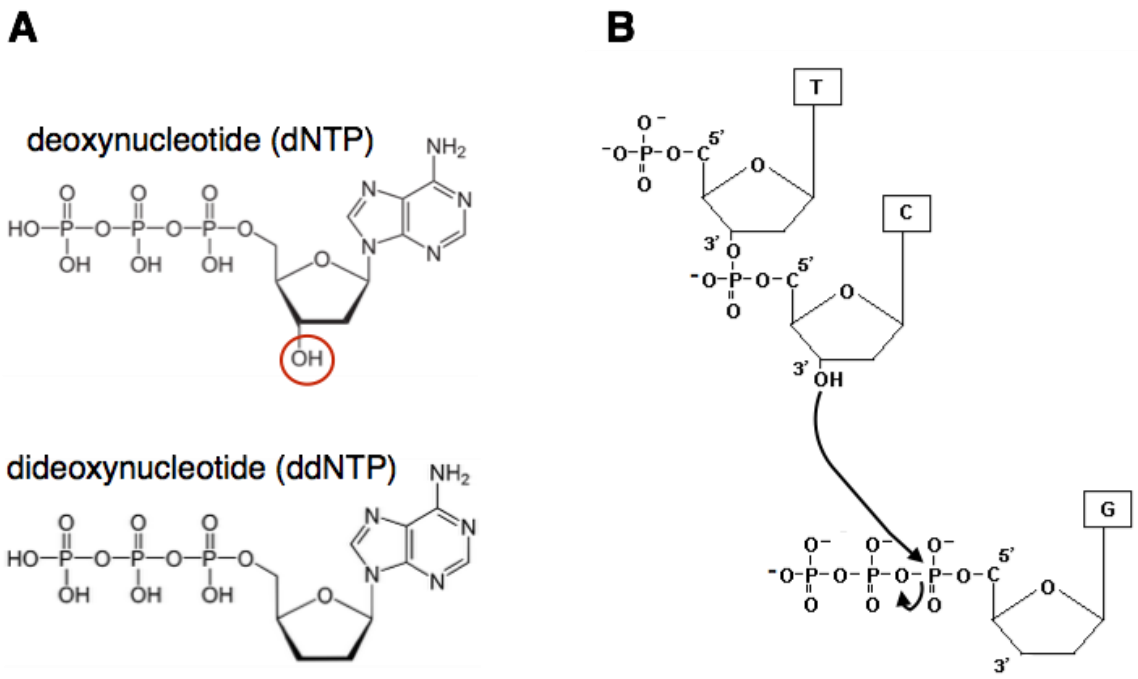

Figure 3. DNA synthesis with dideoxynucleotides

A: Deoxynucleotides and dideoxynucleotides differ in the presence or absence of a hydroxyl attached to the 3’ carbon of the ribose. The molecules shown here are the triphosphate forms of adenosine. **B:** The DNA synthesis reaction, catalyzed by DNA polymerase, forges a new covalent bond between the phosphate of the incoming nucleotide and the hydroxyl on the 3’ carbon of the existing chain. This reaction cannot take place if the terminal nucleotide of the chain is a dideoxynucleotide.

Overview

Provide a broad overview of the Sanger sequencing procedure, such as the one presented in Slide 2 (Figure 4). For more information on the actual procedure, see Smith et al., 1986 (the original protocol) and Slatko et al., 2001 and Walker and Lorsch, 2013 (more recent protocols).

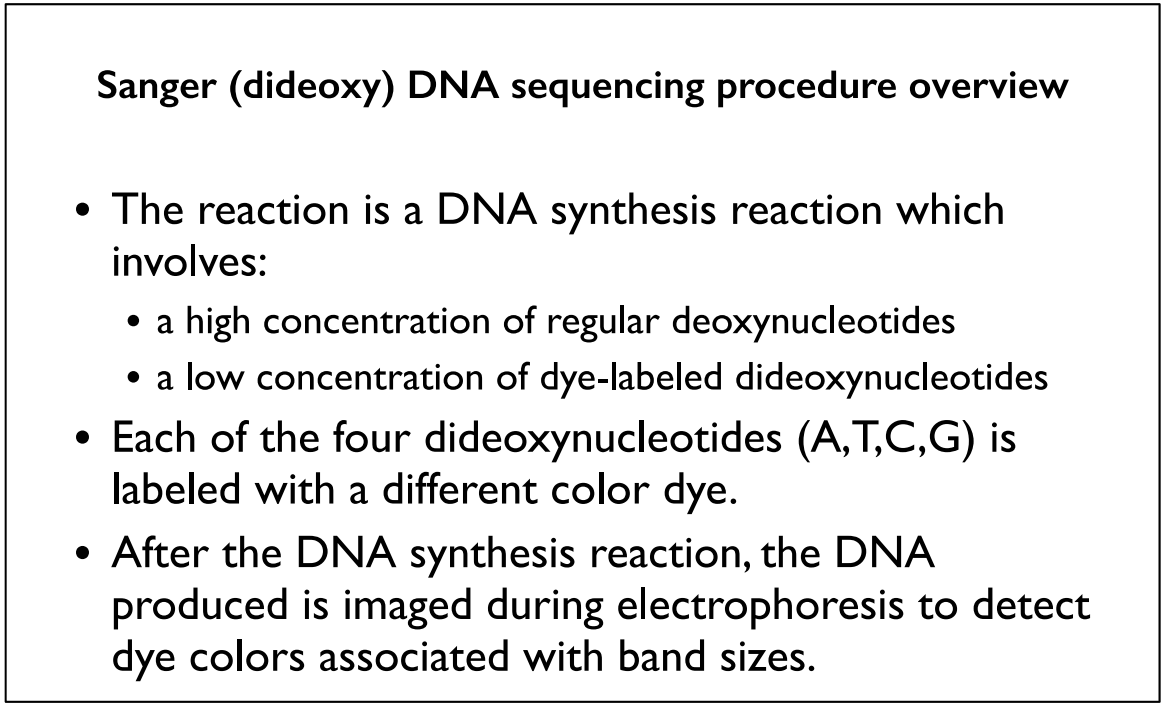

Figure 4. Slide 2 - Overview slide

This slide provides an overview of the Sanger sequencing procedure, highlighting the DNA synthesis reaction, utilizing dye-labeled dideoxynucleotides and subsequent electrophoresis.

You need not spend time explaining each step of the technique now – this will be done in the context of the simulation model.

Phase One – DNA synthesis

Explain the materials

Provide an explanation of what their materials represent, aided with a visual such as Slide 3 (Figure 5).

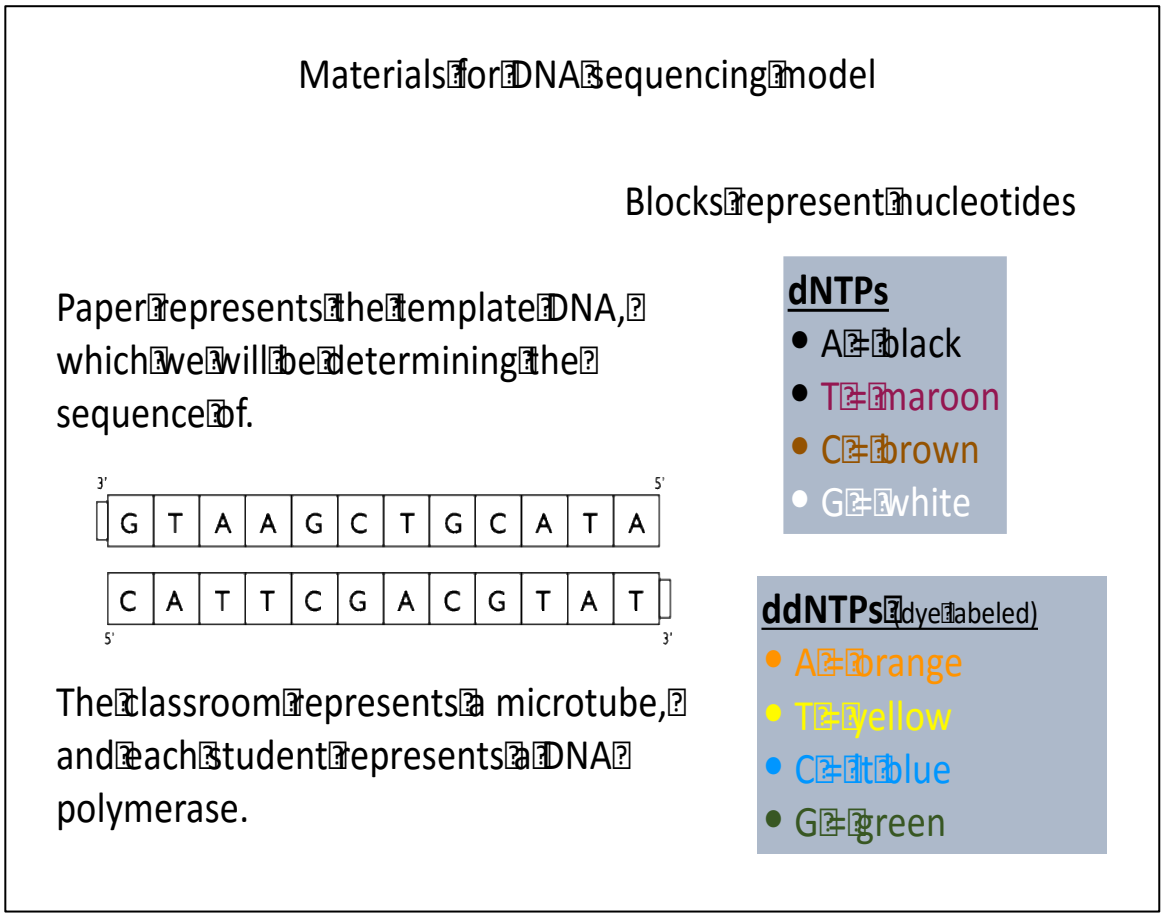

Figure 5. Slide 3 – Materials

This slide explains that template DNA and free nucleotides are represented by their paper and Unifix blocks.

As you go over their materials, also describe the situation that is being modeled. Make the following points:

- Each Unifix cube represents a specific nucleotide.
- The ddNTPs are brightly colored because they have a fluorescent dye covalently bonded (the dNTPs do not).
- The ddNTPs have an obstruction on the 3' end because they are chain terminators – another nucleotide cannot be added to the 3' end of a ddNTP.
- The printed paper represents the template DNA that we will determine the sequence of.
- Each student is playing the role of DNA polymerase.
- The entire classroom represents a microtube 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, which you will need to assemble from your nucleotides.

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 800-1000 nucleotides are routinely achievable.

First steps

Help them carry out the first steps of the simulation, aided with a visual such as Slide 4 (Figure 6). **HANDOUT PROMPT 3** *(Students figure out where primer will anneal)* Use this opportunity to reinforce concepts such as annealing, priming, and the direction of DNA synthesis. Make sure everyone has the primer properly placed before moving on.

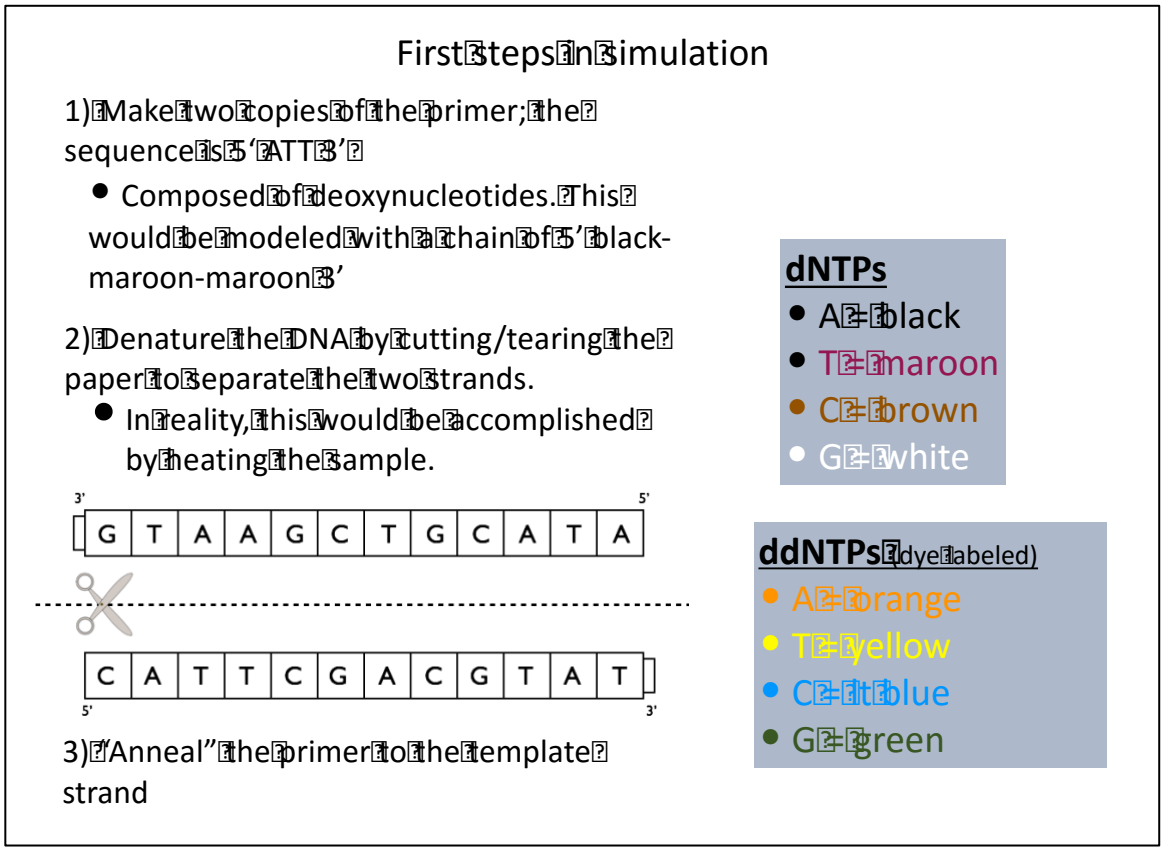

Figure 6. Slide 4 – First steps in the simulation

This slide gives instructions for the first steps the students need to carry out for the exercise. Students are instructed to assemble two primers, denature their template strands, and anneal one of the primers.

Make the following point:

- While there are many copies of template and primer in this tube (represented by the entire classroom), all templates have the same sequence and all primers have the same sequence.

Carry out DNA synthesis

Explain how they will now carry out DNA synthesis, aided with a visual such as Slide 5 (Figure 7).

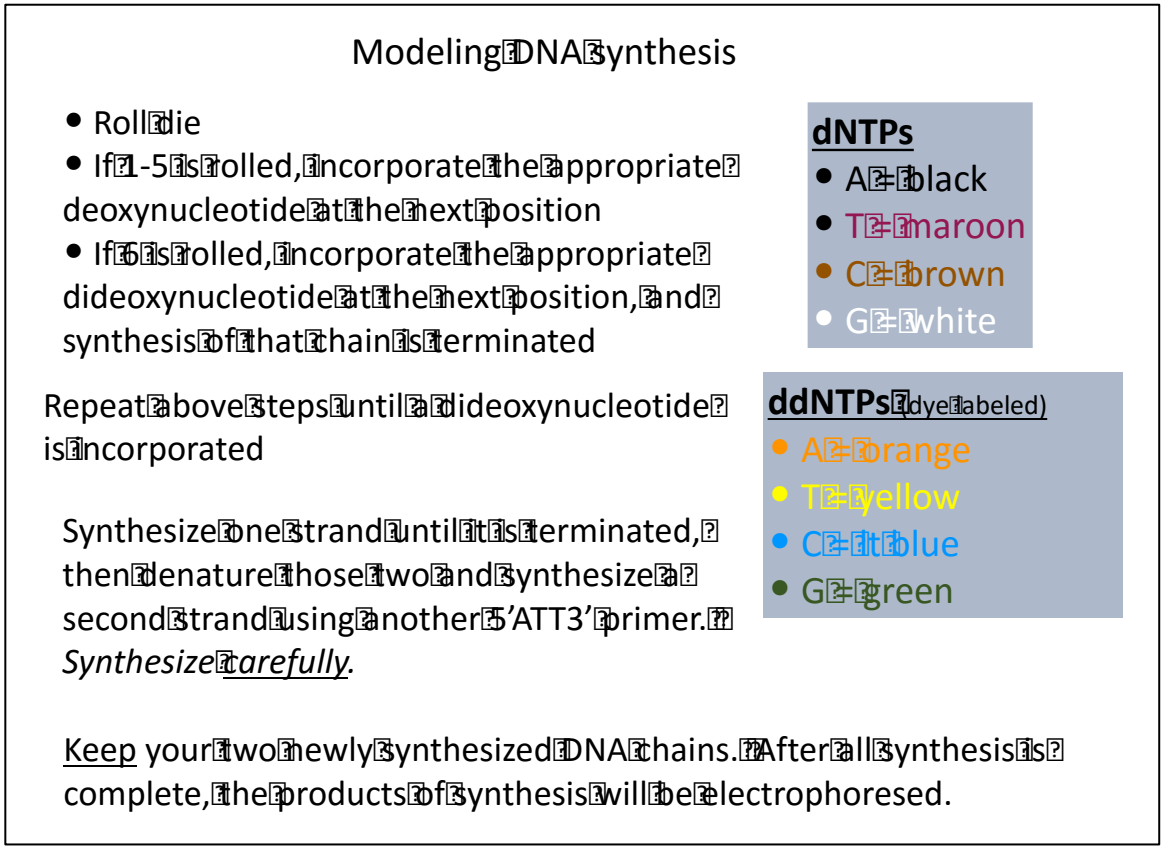

Figure 7. Slide 5 – Modeling DNA synthesis

This slide gives instructions for modeling DNA synthesis. Students should follow these instructions until they have generated two terminated DNA strands.

Before they carry out synthesis, make the following points:

- The students, in their roles as DNA polymerases, will now synthesize DNA, extending the annealed primer.
- 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. They will model this process using die rolls.
- Synthesis proceeds until a ddNTP is added.
- Each pair of students is to carry out synthesis of two new DNA strands.
- It is important to pay attention and carry out these instructions accurately in order for the simulation to work. Make sure to use the right colors.
- Make sure you understand what is being modeled and ask questions.

Carry out synthesis simulation

Now have the students perform their synthesis reactions. Instructors/TAs should rove, assist, answer questions and stimulate discussion. **PROMPT 4** *(Students write down the sequences of the strands they synthesize*). If you have enough supervisory person power, have the students raise their hands when they have completed both strands, and the instructor or TA can check their strands.

If the students produce their strands correctly, the class product should be synthesized chains of varying lengths, each chain with a brightly colored ddNTP at the 3' end (except those that synthesized to the end of the template without incorporating a ddNTP). All chains of the same length should have a ddNTP of the same color at the 3' end (except those that synthesized to the end of the template without incorporating a ddNTP). For the simulation to work optimally, you want to have at least one chain representing each possible length.

In the case where a student makes it to the last base without adding a ddNTP, have them stop there with that chain. This case will be brought up later for discussion.

Phase Two - Electrophoresis

After all synthesis is completed, gather the class up around a central area (an empty floor space or large table) that will represent a gel. Designate a space at one end of the area which will represent the well into which samples are loaded, and have all the students deposit their synthesized DNA chains into the well. Remind them that the DNA synthesis that was just modeled was happening in a single reaction tube with many template molecules and polymerases, and explain that the gel is a "denaturing gel", i.e. the buffer is such that hydrogen bonds are not formed, and all DNA is single stranded. The contents of the tube are then loaded into the well of the gel. **PROMPT 5** *(Students draw a picture of a gel with a loaded well)*

Recruit a group of 5-10 students to assist in simulating electrophoresis, and ask the rest of the class to "proofread" the work and speak out if they see an error.

Note that this electrophoresis phase is a good time to field questions.

"Apply voltage". Initially what should occur at this point is that the assisting students will start to sort through the chains and place them according to length – shorter chains further from the well, to simulate the effect of electrophoresis. As they sort by size, instruct them to keep all chains on the side of the gel close to the well (as if the gel has not run very far). Also dialogue with the class to make sure everyone understands why the chains are being sorted in that way.

Once the chains are sorted and everyone is clear on the concept, make a last check to make sure there are no errors (all chains of the same length should have the same color ddNTP at the end). Then assign one assisting student to each chain length and have those students steadily move their chains down the length of the gel. **PROMPT 6** *(Students draw a gel with the strands migrating through)*

As the bands electrophorese, explain that there is a fluorescence detector near the end of the gel that can detect the color of fluorescent dyes that pass by the detector. A computer makes nucleotide calls based on the color reported by the detector. Assign a student assistant to play the role of fluorescence detector/computer by positioning themself near the end of the gel with an arm outstretched. This student needs to be supplied with the color code (which color corresponds to which base).

As the chains pass in front of the "detector" the detector student should call out the base associated with the ddNTP color at the end of the chain. **PROMPT 7** *(Students record the nucleotide calls)*

The simulation is finished. The students should now engage in explanations with each other prompted by the "Questions for Discussion" included in their handout. Depending on timing, this Discussion session, which likely will take about 45 minutes as written, might be saved for the next class period. Of course, the Discussion length can be expanded or contracted by adding or removing questions. For this discussion, the students should be grouped; I recommend groups of three. Note that some of the images on the handout are of sequencing chromatographs and are best viewed in color. This means you should either produce color printouts, project the color images on a screen, or have students look at the electronic file on a computer/tablet/phone.

Here are the questions, with possible answers:

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

There is not a clear answer, which is part of the point of this question. One could argue that it is likely to be the correct sequence, as there is no clear indication to the contrary, and in some places, it looks quite likely that the correct base is there, but was not called by the computer. One could also argue that there are some peaks that are hardly visible and one might be reluctant to put one's faith in this result. Students could consider how critical it is for the sequence to be correct,

depending on what this sequence information is being used for, or what algorithms might be being used to generate the nucleotide calls.

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.*

This is due to the resolution limitations of electrophoresis. The longer the fragments get, the more difficult it is to separate them with single base pair resolution.

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

Most straightforward would be to do another sequencing reaction using a primer that is closer to the region of interest (and in the correct orientation).

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

Specificity comes from the primer. The primer will bind to complementary sequence with high fidelity. If there is only one locus with complementary sequence, then hopefully only that locus will be synthesized and analyzed.

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?*

If two separate regions are synthesized, then there will be two different sequences synthesized of each band length, and they may not end in the same base. For example, for the bands that are 52 nucleotides long, one sequence might end in C

and the other in T. This would yield two peaks at position 52 in the chromatograph, and it would be ambiguous which base belongs to which sequence. Because of that, little useable data can be recovered.

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.*

This is not a problem. If the end were reached, that means there is no more sequence to determine. Chains synthesized up to the end without incorporation of ddNTP have no dye associated with them, and would be invisible on the gel.

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?*

There would be a gap where there would be a missing peak, and an N call from the computer. This is avoided by synthesizing enough strands (by using enough template, primer, nucleotides, and time, and sometimes by running multiple rounds of denaturation and synthesis, as in PCR) that the likelihood of a missing strand length is vanishingly small.

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 if a low ratio were used, say to 1:1?*

ddNTPs would be incorporated too often, and it would be improbable to generate long chains of sequence. This would mean the read length would be shortened, as after a short stretch of sequence, the results would become too noisy to interpret.

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.

For the substitution, on would expect to see two peaks at the locus with the substitution. These would correspond to the two bases present in the two alleles of that locus. All the other data should look normal, with one peak. For the insertion, the data would look normal up until the insertion, and then it would become difficult to read, as there would be two peaks at most positions. This is because the two sequences from the two alleles would be one base out of register with each other, i.e. one sequence would be shifted by one position relative to the other sequence.

Notes on class size

For the simulation to work optimally, you want to have at least one chain representing each possible length. If your class is small (less than 30 students) you may want to have each student make 2 chains instead of 1; in this case have the students make their chains individually – each student should receive their own "template paper", die, and set of blocks. Alternatively, you can make the template shorter and ask them to start over if they reach the end without incorporating a ddNTP.

If you have a class larger than 50, you may want to "run multiple gels" in Phase Two – Electrophoresis. Break the class into groups of about 40 students and have each group gather around their own gel. You would need an instructor/TA with each group.

References

DiGiuseppe, M. 2003. Additional Activity: Sanger Dideoxy Sequencing: A Simulation Activity. In: *Nelson Biology 12.* Nelson Thomson Learning, Toronto.

Kosinski, R. 2003. A simulation of the Sanger (dideoxy) method for sequencing DNA. In: Tested studies for laboratory teaching, Vol. 24 (M. A. O' Donnell, editor). Proceedings of the 24th Workshop/Conference of the Association for Biology Laboratory Education (ABLE).

Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A. 1977 Dec;74(12):5463-7.

Smith LM, Sanders JZ, Kaiser RJ, Hughes P, Dodd C, Connell CR, Heiner C, Kent SB, Hood LE. Fluorescence detection in automated DNA sequence analysis. Nature. 1986 Jun 12-18;321(6071):674-9.

McGinn S, Gut IG. DNA sequencing - spanning the generations. N Biotechnol. 2013 May 25;30(4):366-72.

Heather JM, Chain B. The sequence of sequencers: The history of sequencing DNA. Genomics. 2016 Jan;107(1):1-8.

Slatko BE, Albright LM, Tabor S, Ju J. DNA sequencing by the dideoxy method. Curr Protoc Mol Biol. 2001 May;Chapter 7:Unit7.4A.

Walker SE, Lorsch J. Sanger dideoxy sequencing of DNA. Methods Enzymol. 2013;529:171-84.