A Simulation of the Sanger (Dideoxy) Method for Sequencing DNA

Robert J. Kosinski

Department of Biology Instruction and Agricultural Education
330 Long Hall
Clemson University
Clemson, SC 29634-0325
Phone: 864-656-3830
rjksn@clemson.edu

Bob Kosinski received his B.S. in Biology from Seton Hall University and his Ph.D. in Ecology from Rutgers University. He is a Professor of Biology at Clemson University, where he teaches introductory biology. His instructional interests include the use of investigative laboratories, computer simulations, and physiological interfacing in introductory biology. He has been a member of ABLE since 1989, and was host of the 2000 meeting at Clemson University.

Reprinted From: Kosinski, R. 2003. A simulation of the Sanger (dideoxy) method for sequencing DNA. Pages 245-249, in Tested studies for laboratory teaching, Volume 24 (M. A. O'Donnell, Editor). Proceedings of the 24th Workshop/Conference of the Association for Biology Laboratory Education (ABLE), 334 pages.

Abstract

This paper describes a classroom role-playing exercise that simulates the Sanger method. A mass-participation version of the method uses from 6 to more than 100 students to play the role of nucleotides and can simulate the sequencing of DNA strands up to about 6 nucleotides long. An instructor demonstration version of the method can be used in classes of any size, and takes only a few minutes to complete. Both exercises emphasize that Sanger sequencing depends on *random* incorporation of ddNTPs in growing nucleotides.

Introduction

The Sanger Method

Since its introduction in 1977, the Sanger, or dideoxy method has become the most widely practiced method for sequencing DNA. An automated, fluorescence-based variation of the method can sequence a DNA fragment with thousands of nucleotides in just a few hours (Nelson and Cox, 2000). The Sanger method is explained in many biochemistry texts (e.g., Berg *et al.*, 2002; Nelson and Cox, 2000). Space limitations prevent the method's full description here, but its essence is that when a growing DNA strand incorporates a dideoxy nucleotide, its growth must stop.

An important decision in the Sanger method (and in the simulation as well) is how much dideoxy nucleotide to include in the reaction mix. If the amount is too high, all new strands will terminate after just

one or a few nucleotides. If the amount of dideoxy nucleotide is too low, most of the strands will go to completion because they never incorporate a dideoxy nucleotide. The optimum result is to have a range of new strand sizes, with even numbers of those that terminate at the first nucleotide, those that terminate at the second nucleotide, etc. Commerical sequencing kits are carefully adjusted to give this result. For example, in an Amersham commercial kit meant for sequencing strands of about 500 nucleotides, the proportion of ddNTP to the corresponding dNTP is 1.33 µM to 400 µM — about 0.3% ddNTPs (Smith, 2002).

The Sanger Simulation

The proposed simulation emphasizes the key points of the Sanger method: Adding a ddNTP to the new strand stops new strand elongation, and if just the right *small* amount of ddNTP is used, the reaction mixture will produce a great diversity of strand lengths. The mass participation version of the simulation uses students as nucleotides, and lines of students of different lengths at the front of the room represent the fragment sizes that the Sanger method develops. An "instructor demonstration" version is very rapid and can be used with only one or two students. To simplify and speed up the exercise, a strand that contains only A nucleotides is sequenced. The only equipment necessary are cards that are marked either with "A" or "ddA." These cards will be randomly drawn from a deck and displayed by the students.

Strand Length and Replication in the Simulation

If the Sanger simulation sequencing of a "poly-A" strand is successful, there should be approximately equal numbers of student columns that stop at the first A, at the second A, etc., up to the size of a completed new strand. That is, each A in the stand generates a fragment that can be seen on an electrophoretic gel. This ideal outcome is shown in Fig. 1.

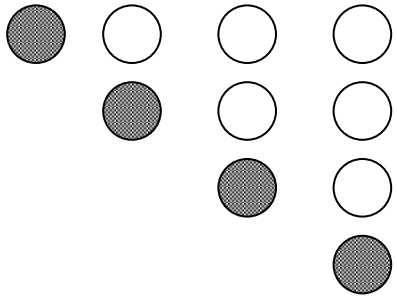

Fig. 1. The optimum outcome of an attempt to sequence a strand of 4 nucleotides. The open circles represent students with "A" cards and the filled circles represent students with "ddA" cards. The students above are standing in the front of the room, facing the class, which is at the bottom of the figure.

This is a good result because every possible strand length (e.g., one nucleotide, two nucleotides, three nucleotides, and four nucleotides) is represented. Note that the number of student vertical columns and the number of ddA cards used is equal to the length of the completed DNA strand (four). The total number of students used in this unit is 4 + 3 + 2 + 1 = 10. This is the smallest number of students that *could* show every possible strand length in this case.

However, the outcome in Fig. 1 is rarely attained because there are so many possible unsatisfactory outcomes. For example, in a simulation of a strand length of four nucleotides, it is also possible that random draws from the deck will produce the following patterns:

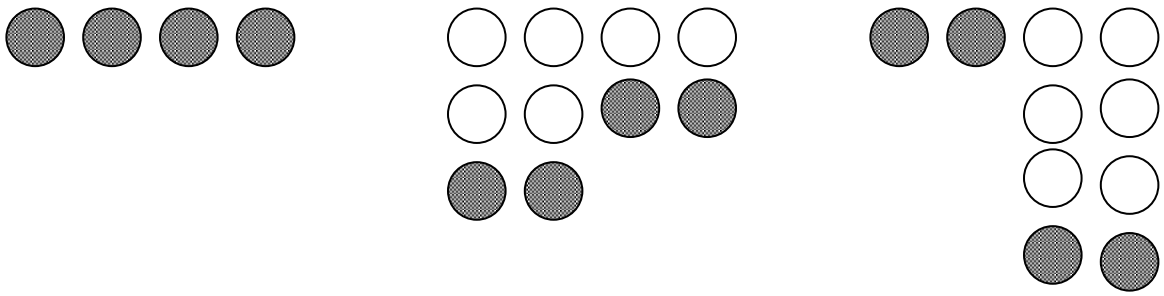

Fig. 2. Four-nucleotide simulation results that do not show all possible strand lengths. The open circles represent "A" cards and the filled circles represent "ddA" cards.

In the first case, the students drew all four ddA cards first, and no strands could be lengthened after this happened. In the second case, all the ddA cards were drawn last. In the third case, the ddA cards were drawn first and last. None of these cases show the full range of strand lengths.

In the real world Sanger method, the involvement of millions of independent molecules ensures that extreme outcomes like these do not dominate. Replication is the solution to the problem of inconvenient outcomes in the simulation as well. If we included several of the units shown in Fig. 1, we would be much more certain of seeing all nucleotide lengths in the results. For example, if we pooled results of each of the three simulations in Fig. 2 above, we would see all strand lengths represented.

The recommended number of students necessary for sequencing strands of different lengths is shown in Table 1. Note that the number of students needed rapidly inflates as strand length increases.

Table 1. The number of students and "A" and "ddA" cards required to "sequence" DNA strands of different lengths.

Length	Students	"A" Cards	"ddA" Cards
2	6	2	4
3	18	9	9
4	40	24	16
5	75	50	25
6	126	90	36
7	168	126	42
8	216	168	48

Aside from the large number of students required for greater strand lengths, space in the classroom might also become limiting. The number of columns of students in the front of the room is equal to the number of ddA cards above; the maximum number of students in each column is equal to the strand length. Therefore, sequencing of an 8-nucleotide strand would require space at the front of the room for 48 columns of students, each from 1-8 students long. Even a 4-nucleotide strand would require 16 columns of students. Finally, even if there are large amounts of room and plenty of students available, some students will always be reluctant to participate. It is wise to plan on sequencing a strand of modest length.

If space or student numbers is a problem, it is still possible to do the exercise on a transparency or a blackboard without requiring physical participation by the students. Directions for doing this are in the section entitled, "Procedure—Instructor Demonstration Option."

Materials

Aside from the students and space required (discussed above), the only materials needed are the "A" and "ddA" cards. The letters on the cards should be written large enough to be visible from the back of the class when students at the front of the class hold them up. It is convenient to print two "A" or two "ddA" (perhaps in a 200-point font) per page on heavy stock. Then laminate the pages and cut each one in half with a paper cutter, producing durable cards that are 21.6 cm x 13.8 cm.

Procedure—Mass-Participation Option

For this example, assume that the instructor plans to use a strand length of 4, which will require 40 students (Table 1).

1. Make up a card deck consisting of 16 cards with "ddA" and 24 cards with "A."
2. Make sure that the room being used has space for 16 columns of students at the front, each column 4 students long. Possibly mark the places on the front wall where each column will begin. If the number of students is large, 16 students could be devoted to holding up signs displaying the numbers from 1-16.
3. In class, explain the Sanger method. Then point out the places along the front wall of the room where each column of students will start. Shuffle the card deck. Then fan the deck out with the faces of the cards downward.
4. Ask for volunteers. Each student will approach you and randomly select a card from the deck. As the first one does, emphasize to the class that incorporation of either an NTP or a ddNTP is random, just as this card draw is random. As each one takes the card, tell him/her the column number to go to. In this example, starting at 1 and giving consecutive numbers up to 16 and then starting at 1 again would be simplest. If this step is omitted, students will tend to pile onto existing columns rather than starting new columns.
5. Each student attempts to add on to his/her column. If the last person in the existing column is holding up an "A" card or if the column is empty, the new student adds onto the column, faces the class, and holds up his/her card. Even a student with a "ddA" card does this. However, if the last student in the column is holding up a "ddA" card, that column is closed. The column is also closed if it has reached the strand length being used in the simulation (because this is the strand length of the "template strand"). A "rejected" student tries to add onto another column that is not closed. If all columns are closed, the student cannot add and must stand aside.
6. When the last card has been given out, the instructor can turn and look at the results. Optimally, there should be at least one column with one student, at least one column with two students, and so forth, up to the strand length. If (for example), there is no column with three students, explain that in a test tube with millions of DNA molecules, the chance that *no* strands will stop at three nucleotides is low. However, an imperfect outcome shows that the addition of a ddNTP to the end of a strand is a matter of probability.
7. Ask the students how the outcome would have been different if *all* the ATPs in the reaction mixture had been ddATPs, and how it would have been different if *none* had been ddATPs. The students will probably see that in the first case *no* strands would have gone to completion, and in the second case *all* strands would have gone to completion.
8. Collect the cards and send the students back to their seats.

Procedure—Instructor Demonstration Option

This option can be used where space in the classroom is limited or class size is very small. It demonstrates the same principles, but only requires active participation by two students. It is also more rapid than the mass-participation exercise. The disadvantage is it doesn't get the students up and moving. Once again, assume that the instructor wishes to sequence a four-nucleotide strand.

1. Make up a card deck consisting of 16 cards with "ddA" and 24 cards with "A."
2. On the board or on a transparency, make a grid with 16 columns and 4 rows.
3. In class, explain the Sanger method.
4. Ask for two volunteers. One of these will be "dealer" and one will draw cards. Ask the dealer to shuffle the deck and then fan it out with the faces of the cards down.
5. The other student will randomly draw cards and hold them up so the class can see them. The class then calls them out (either "A" or "ddA") to the instructor. As this is done the first time, emphasize that the incorporation of an NTP or a ddNTP is random, just as these card draws are random. After drawing, the used cards will be put aside.
6. As the class calls each card, the instructor writes either an "A" or a "D" (simpler than "ddA" in a crowded grid). The instructor fills in the top row of the grid left to right, starting at the top left. In the second, third, and fourth rows, the instructor skips a space whenever the space above it has "D." Ask the class its advice on this ("The next card is an A. Should I put an A here? No? Why not?"). Once the fourth row has been completed, all the remaining cards (if any) are ignored. A sample outcome appears below. In this case, four cards (three A and one ddA) were not used.

Table 2. Sample "instructor demonstration" outcome.

D	D	A	A	A	A	A	D	D	A	D	A	A	A	A	A
		D	A	A	D	A			A		D	A	A	D	D
			A	D		D			D			A	A		
			A									D	D		

7. Ask the students how the outcome would have been different if *all* the ATPs in the reaction mixture had been ddATPs, and how it would have been different if *none* had been ddATPs. The students will probably see that in the first case *no* strands would have gone to completion, and in the second case *all* strands would have gone to completion.
8. Collect the cards and send the students back to their seats.

Acknowledgements

I would like to thank Kerry Smith of Clemson University for his useful comments on this manuscript, and the Clemson University Biology 110 class of Fall 2001 for their participation in the first version of this exercise.

Literature Cited

Berg, J. M., J. L. Tymoczko and L. Stryer. 2002. Biochemistry. Fifth edition. W. H. Freeman and Co., New York, New York, pages 146-148.

Nelson, D. L. and M. M. Cox. 2000. Lehninger: Principles of Biochemistry. Third edition. Worth Publishers, New York. Pages 351-353.

Smith, K. 2002. Personal communication to R. Kosinski on 28 May 2002.