A Demonstration of Automated DNA Sequencing

Sandra Latourelle Bonnie Seidel-Rogol

In science, the development of new methodologies is often required before significant advances in knowledge can be made. DNA sequencing (Maxam & Gilbert 1980; Sanger et al. 1977) is one technological advance that has led to a huge "bloom" in the amount of genetic information that is now available. Methods for sequencing are continually being refined, and thanks to the impetus provided by the Human Genome Project (DeLisi 1988), researchers have been successful in automating much of the process (Smith et al. 1986). The purpose of this exercise is to use a pencil and paper model to demonstrate the principles behind automated DNA sequencing. This simulation is a revision of a model first developed by Nancy Ridenour (1994) and later incorporated into a laboratory activity manual distributed by Cornell Institute of Regional Biology Mentors (CIRBM) under the direction of Joseph Novak (1995).

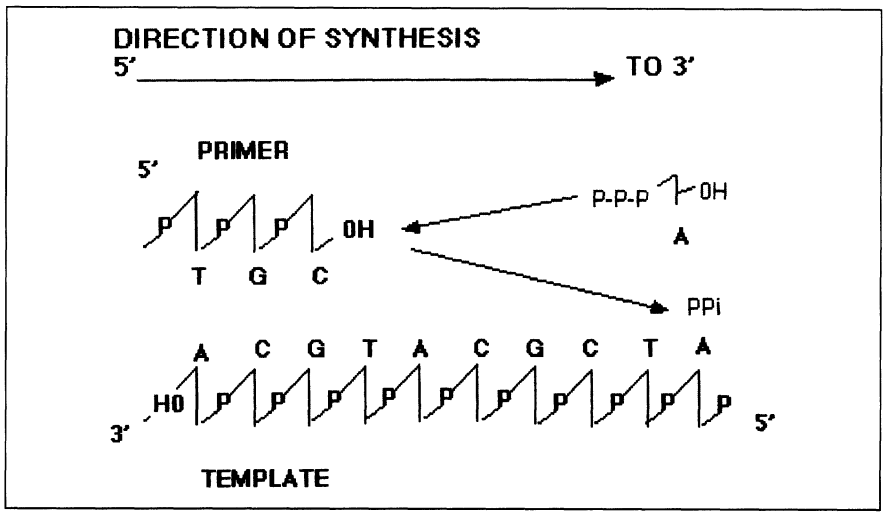

Figure 1. Mechanism of DNA polymerase action. Synthesis is in a 5' to 3' direction. DNA polymerase binds to the 3'-OH group of the primer and incorporates a base that is complementary to the base on the template strand. Base complementarity requires that A pairs with T and C pairs with G. Note that two phosphate groups are cleaved from the incoming deoxynucleotide triphosphate, providing energy for the reaction.

Background Information

Automated DNA sequencing is based upon the Sanger di-deoxy method (Sanger et al. 1977). This technique is based upon a standard reaction that occurs during DNA replication. Provided with a primer and a template strand of DNA, the enzyme DNA polymerase will sequentially add nucleotides to the 3'-OH end of the primer so that the bases being added are complementary to the template (see Figure 1). Normally the nucleotides that are added are in the form of deoxynucleotide triphosphates (dATP, dCTP, dGTP, dTTP), with the cleavage and removal of two phosphate groups providing "energy" for the reaction (Griffiths et al. 1993). Deoxynucleotides lack the 2'-OH group on the ribose (see Figure 2). If the 3'-OH group is absent as well, this results in a di-deoxynucleotide (ddATP, ddCTP, etc.) or nucleotide analog. This causes termination of the DNA synthesis reaction. DNA polymerase requires the presence of the 3'-OH group in order to work. Thus, polymerase reactions that incorporate di-deoxynucleotides result in a mixture of DNA fragments of different lengths that can then be separated by gel electrophoresis. This separates the fragments on the basis of their size, so that the smallest fragments migrate the fastest through the gel. If the di-deoxynucleotides are labeled with a fluorescent dye—a different colored dye for each nucleotide—fragments can be detected as they come off the gel. This is done by exposing them to a laser beam that causes the dye to fluoresce (see Smith et al. 1986 for a diagram of the sequencing apparatus). This fluorescence is then detected, information is stored in a computer, and a printout of the sequence of colors coming from the gel is obtained. Analysis of this information provides the nucleotide sequence.

The advantages of automated sequencing are that it is rapid, fairly inexpensive, and avoids the use of radioisotopes. It should be noted that in practice, the polymerase chain reaction (PCR) is used to generate the DNA fragments that will then be separated on the gel (Mullis 1990). This simulation does not model the use of PCR, but the technique has been included in the concept map which accompanies this article (Figure 3).

Sandra Latourelle is an Instructor and **Bonnie Seidel-Rogol** is Associate Professor in the Department of Biological Science, SUNY-Plattsburgh, NY 12901-2681.

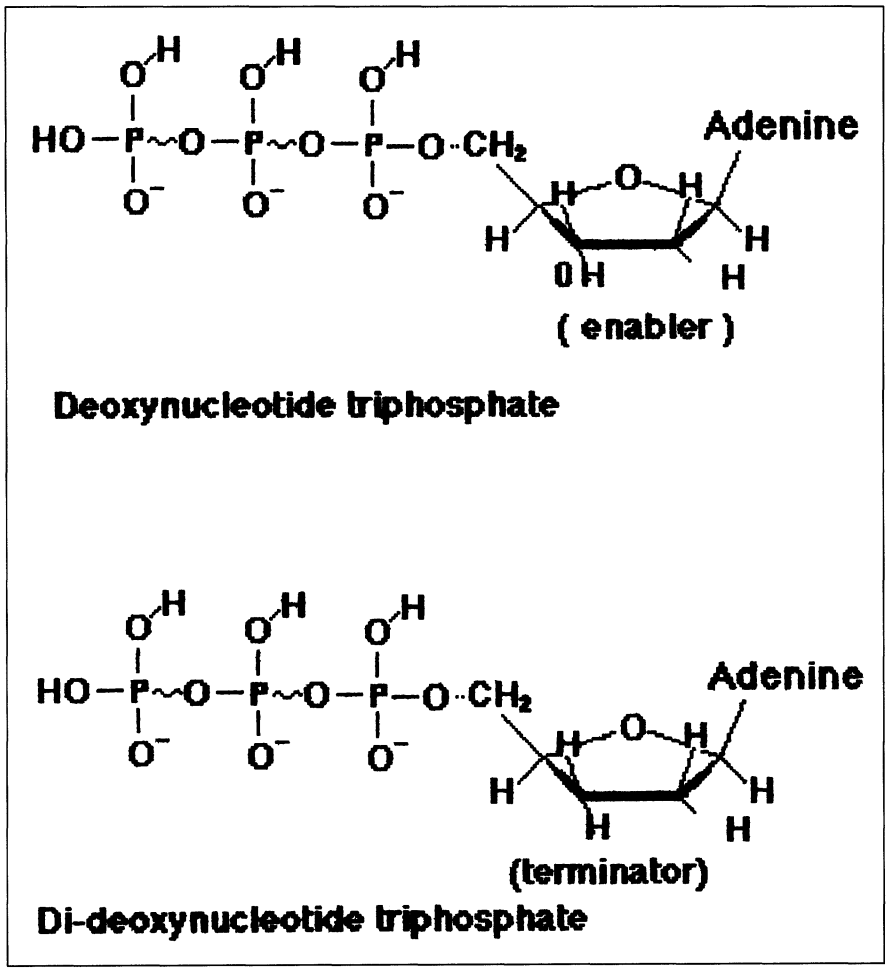

Figure 2. Comparison of deoxy and di-deoxynucleotide structure. Deoxynucleotides contain an OH group at the 3′ position of the ribose molecule while di-deoxynucleotides contain only a hydrogen atom. The 3′-OH group is required for synthesis of the DNA daughter strand.

Materials per Group

- Two pages of the primer-template sequence*
- A page of A's duplicated on (for example) yellow fluorescent paper*
- A page of C's duplicated on blue fluorescent paper*
- A page of G's duplicated on green fluorescent paper*
- A page of T's duplicated on pink fluorescent paper*
- A page of each of the above nucleotides on white paper
- Scissors
- Tape
- An 11″ × 14″ sheet of paper—to simulate gel
- Paper bag or other opaque container for mixing nucleotides

**Templates included in Figure 4*

Procedure

Running the Reaction

1. Students work in groups of two. Each group should start by cutting the primer/template sequence sheet into individual primer-template pairs. **Do not separate the primer from the template at this point.** At least seven primer/template strips are required and can be placed on a table or desk.
2. The pages containing letters must also be cut up into small sections, each containing one letter. For each letter, each group will need 20 white and 5 colored pieces of paper. These are all mixed together in a container.
3. One student holds the container, and the other student acts as the DNA polymerase. The "DNA polymerase" reaches into the container and pulls out a nucleotide. This should be a totally random event, so no peeking!
4. Refer to the flow diagram (Figure 5) which outlines the general decisions to be made when synthesizing the daughter strand. What is the nucleotide that was just pulled out? If it was an "A" it is complementary to the first available template base (a "T"). Tape the "A" next to the "C" on the primer. If the nucleotide was not an "A," throw the base back into the container and try again. (Note: This is exactly what a DNA polymerase enzyme would do. It has a 3′–5′ exonuclease activity that removes any bases that are not complementary.) Repeat until an "A" is pulled out. Tape it down and go on to Step 5.
5. Was the "A" white or colored? If it is colored, you have just incorporated a di-deoxynucleotide and DNA synthesis can no longer continue on this strand. Lay the strand aside and start again at Step 4 with a new primer/template strand. If the "A" was white, DNA synthesis can continue, and the "DNA polymerase" can reach into the container to pull out another nucleotide.
6. Repeat Steps 4 and 5 until you have at least seven **different** primer/template strips, each containing a di-deoxynucleotide at a different position. You may obtain duplicate sequences as you are doing this. These can be pooled together, in which case you'll have more than seven sequence strips, but they will be classified into only seven categories.

Running the Gel

1. The electrophoresis gels that are run are denaturing gels, which means they separate the DNA into single strands. DNA strands are separated because the gels are run at high temperatures (greater than 70°C) in the presence of urea which interferes with hydrogen bonds. Take your primer/template strips and cut them horizontally to separate the template and daughter strands.
2. Now arrange the daughter (taped) strands according to size on the 11″ × 14″ sheet of paper, with the smallest strand positioned at the bottom.
3. Start at the bottom of the gel and record the color of the di-deoxynucleotide that is contained in

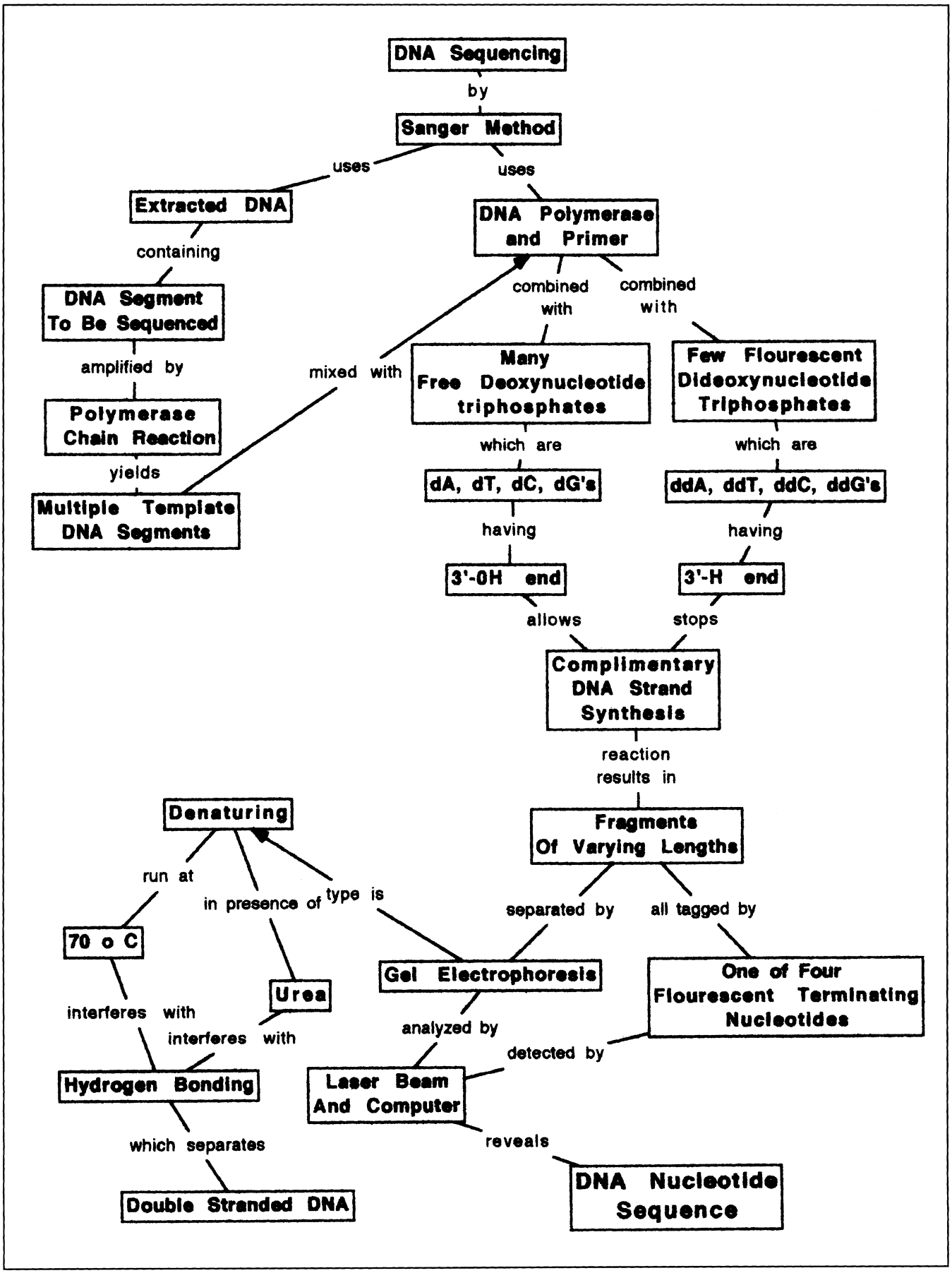

Figure 3. Concept map detailing automated DNA sequence. Figure made using CMap software for Macintosh. Upon completion of the laboratory exercise, students could be asked to design their own concept map.

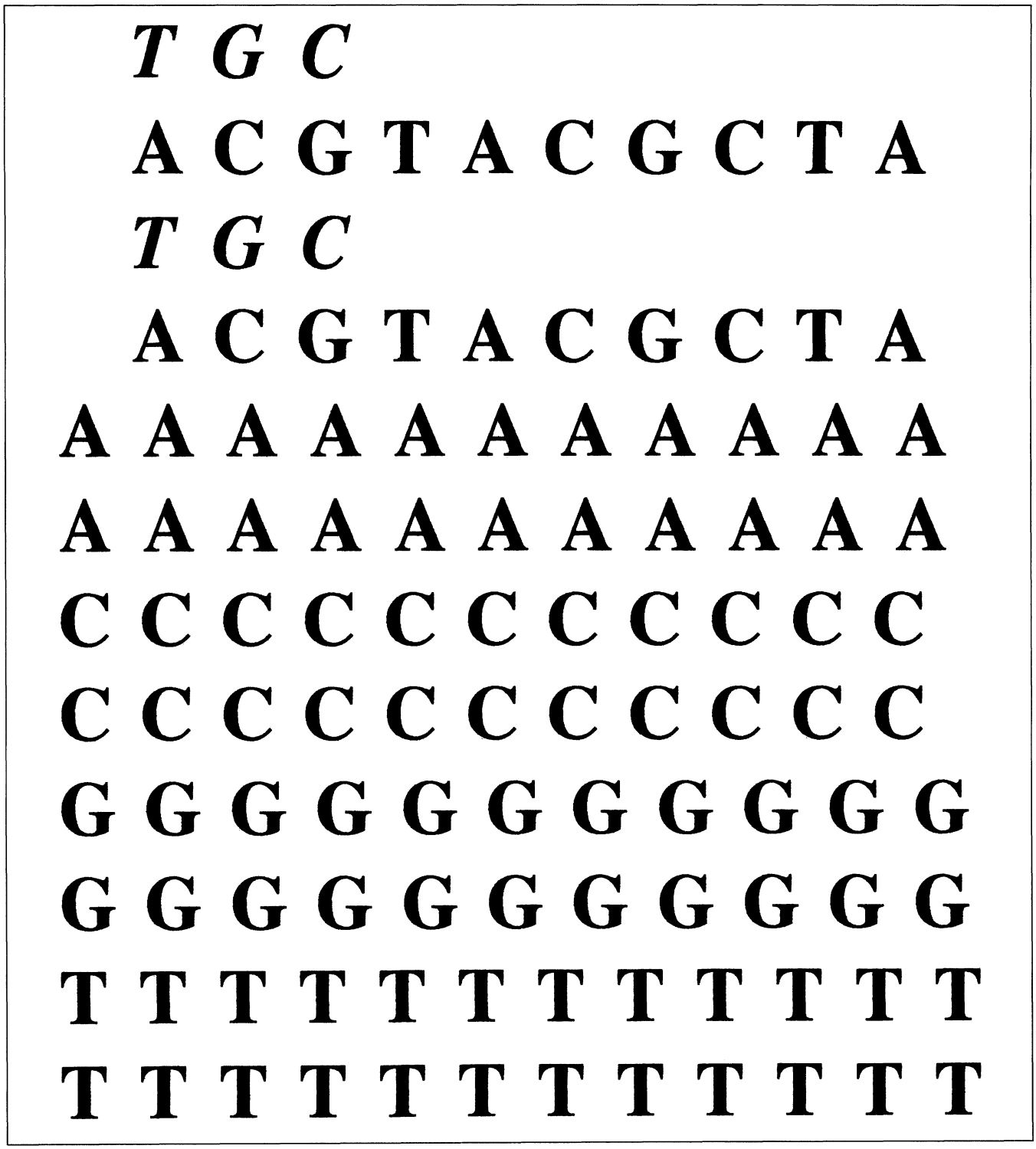

Figure 4. Templates that can be used to reproduce sequence and bases needed for materials list.

the daughter strand in each position on the gel. Work your way up the gel from smallest to largest fragment.

4. Now write the letter that corresponds to each color:
   A = yellow
   C = blue
   G = green
   T = pink

You now have your sequence.

Discussion

The usefulness of pencil and paper models for helping students to visualize some of the complex concepts in molecular biology has become well established (Kreuzer & Massey 1996). In fact, a number of DNA sequencing simulations have already been developed (Kreuzer & Massey 1996; Contolini 1996). To our knowledge, this is the first simulation that models automated DNA sequencing. In addition this exercise demonstrates two important concepts: the functioning of the DNA polymerase including its ability to edit, and the principle behind running a

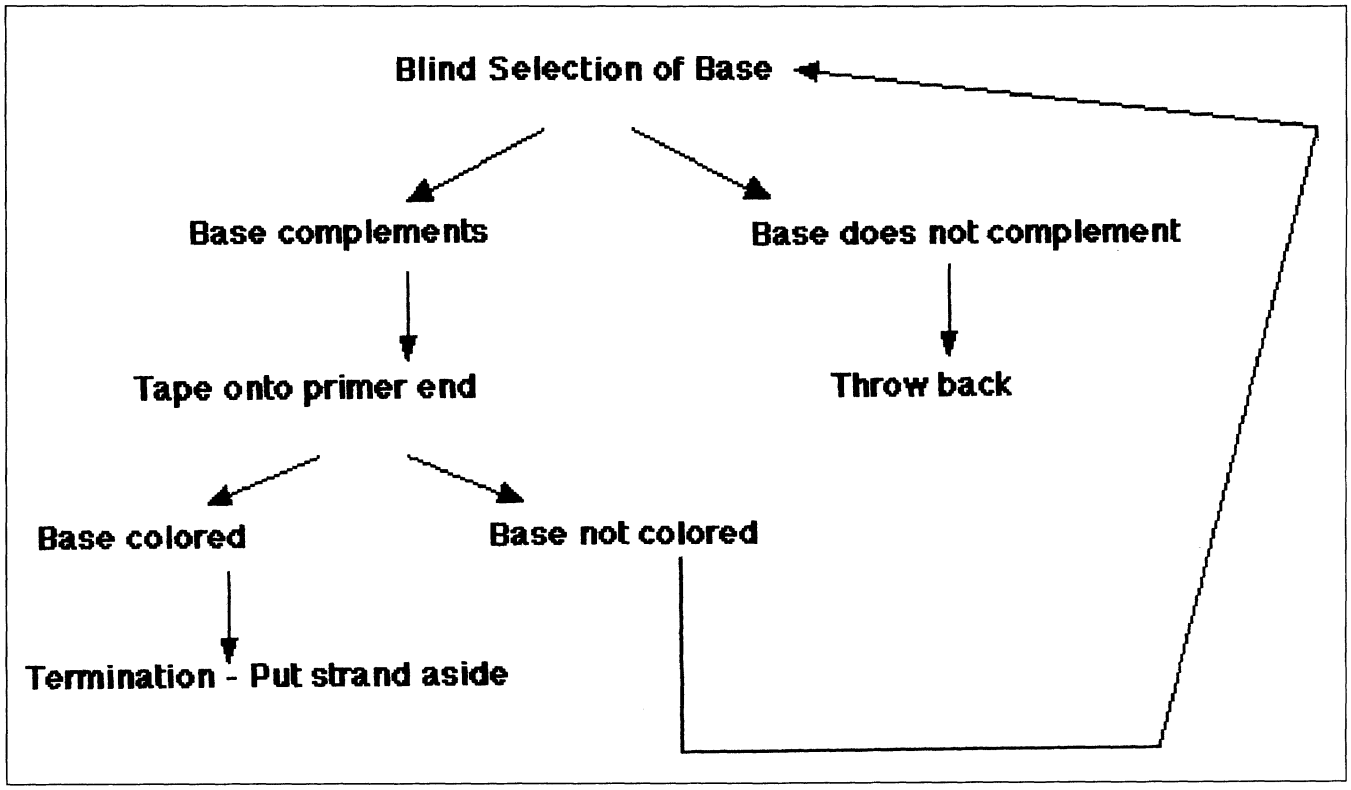

Figure 5. Flow diagram showing sequence of decisions DNA polymerase makes during DNA synthesis.

denaturing gel. Some previous classroom discussion of DNA replication and the Sanger method of DNA sequencing was found to be helpful. Follow-up discussions with students focused upon a comparison of the automated and the Sanger methods and the fact that the Sanger method requires four lanes on a gel, while the automated method requires only one, since all four fluorescently labeled nucleotides can be mixed together in the same tube. Figure 6 shows the results that would be obtained if the same sequence was subjected to the original Sanger protocol and compares this to the expected student results as well as a model computer output.

Additional class discussions focused on the importance of the concentrations of both the normal and di-deoxynucleotides. In this exercise, a ratio of 20:5 normal (white) to di-deoxy (colored) respectively was used. One possible variation would be to try other ratios such as 15:10. It is expected that this would make the analysis more difficult, because the concentration of di-deoxy nucleotides are too high and the sequence would be terminated rapidly, prohibiting the sequencing of regions located far from the primer end. On the other hand a ratio of 40:5 would have too few di-deoxynucleotides and it would be difficult to obtain sequence information close to the primer end.

Another possible variation is to alter the sequence that is being analyzed. Sequences containing two identical bases in a row (i.e. "CC" or "TT") were found to be more difficult to sequence, because the concentration of normal (white) C's or T's decreases rapidly, making it difficult to sequence past this region. Any sequence with an even

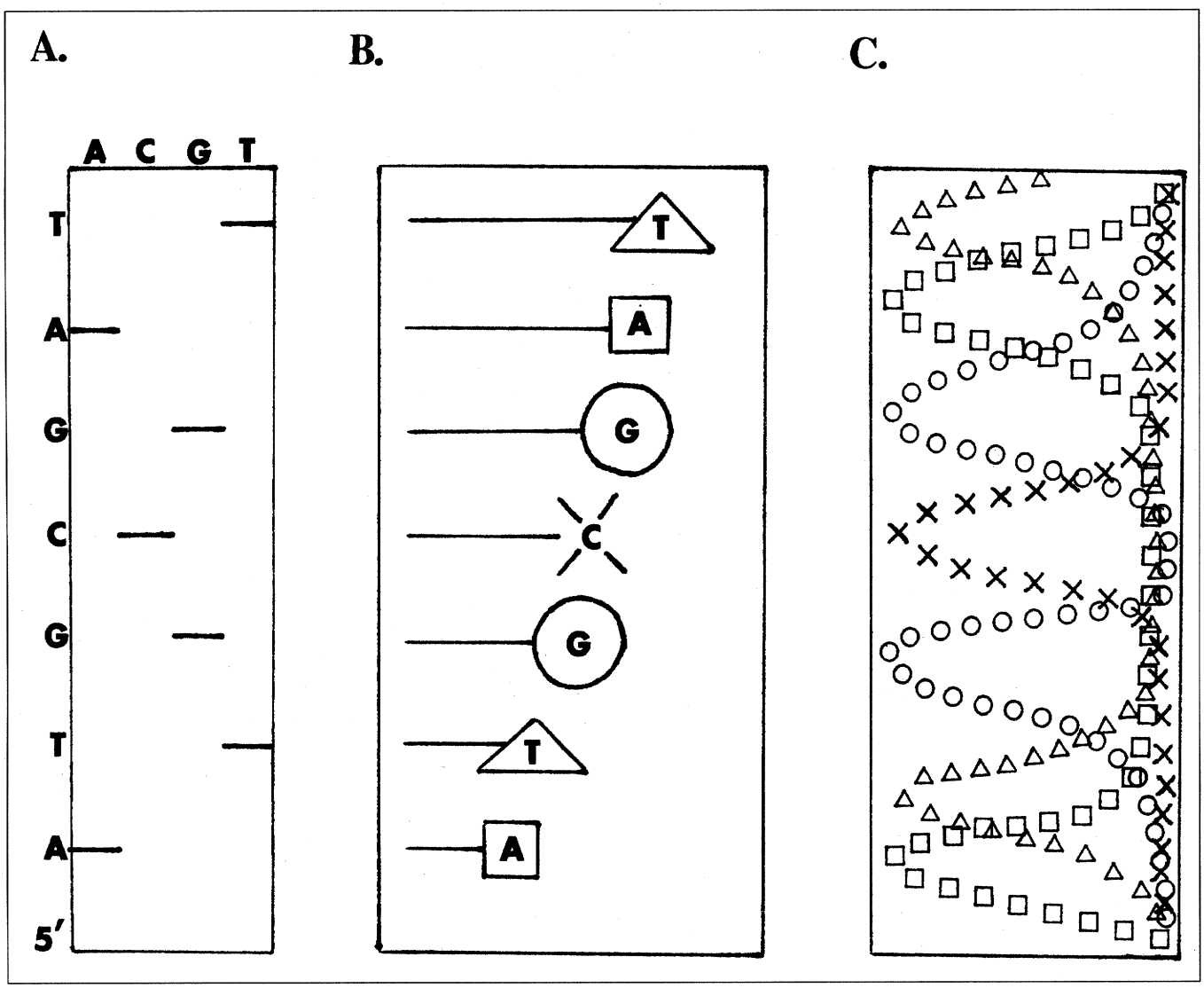

Figure 6. Comparison of Sanger and automated sequencing. A. Results of Sanger sequencing for template provided. B. Appearance of student generated gel simulation. Triangles represent ddT with a pink fluorescent label. Squares represent ddA with a yellow fluorescent label. Circles represent ddG with a green fluorescent label. Finally, blue fluorescent ddC is represented by an X. C. Diagram illustrating computer output for same sequence. Symbols and colors are the same as described in B above. Colored pencils could be used to make a more realistic diagram of the computer output.

distribution of the four nucleotides should work fairly well. It could be mentioned, as part of the discussion, that DNA containing a large amount of G's and C's is difficult to sequence because it is more difficult to unwind.

Additional pedagogical materials have been developed to accompany this lab, and will be made available upon request. Please include a self-addressed, stamped, 8″ × 11″ envelope with your written request.

References

Contolini, N. (1996). DNA sequencing simulation. *The Science Teacher, 63*(4), 26–29.

DeLisi, E. (1988). The human genome project. *American Scientist, 76*, 488–493.

Griffiths, A.J.F., Miller, J.H., Suzuki, D.T., Lewontin, R.C. & Gilbert, W.M. (1993). *An Introduction to Genetic Analysis*. 5th ed. New York: W.H. Freeman and Co.

Kreuzer, H. & Massey, A. (1996). *Recombinant DNA and Biotechnology: A Guide for Teachers*. Washington, DC: ASM Press.

Maxam, A.M. & Gilbert, W. (1980). Sequencing end-labeled DNA with base-specific chemical cleavages. In L. Grossman and K. Moldave (Eds.), *Methods in Enzymology* (pp. 499–560). New York: Academic Press.

Mullis, K. B. (1990). The unusual origin of the polymerase chain reaction. *Scientific American, 264*(4), 56–65.

Novak, J. (1995). *Laboratory Activity Manual*. Ithaca, NY: Cornell Institute of Regional Biology Mentors.

Ridenour, N. (1994). DNA sequencing designed for Honors Biology. Ithaca, NY: Ithaca High School.

Sanger, F., Nicklen, S. & Coulsen, A.R. (1977). DNA sequencing with chain terminating inhibitors. *Proceeding of the National Academy of Sciences USA, 74*, 488–493.

Smith, L.M., Sanders, J.Z., Kaiser, R.J., Hughes, P., Dodd, C., Connell, C.R., Heiner, C., Stephen, B.H.K. & Hood, L.E. (1986). Fluorescence detection in automated DNA sequence analysis. *Nature, 321*, 674–679.