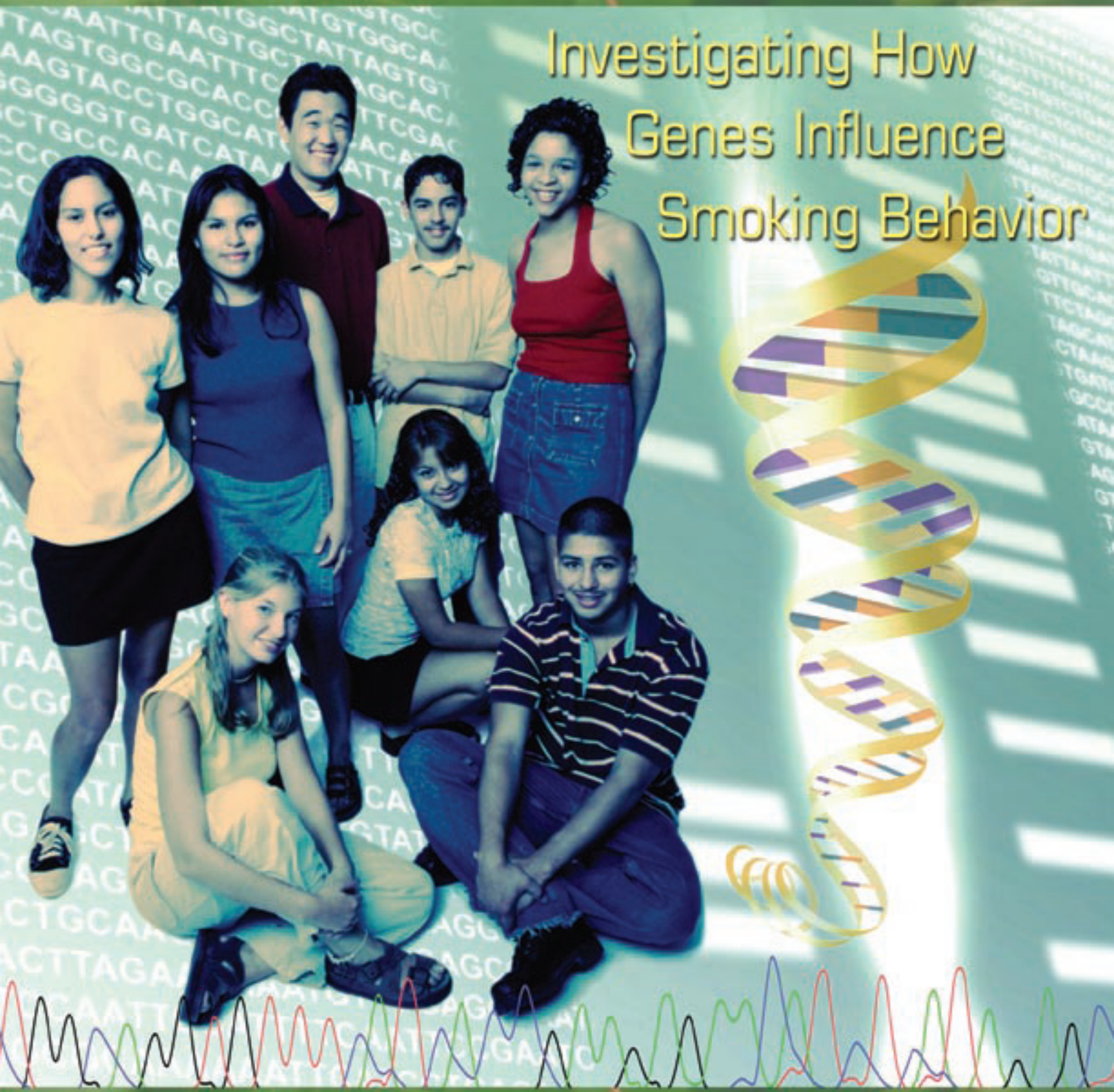
Genetic Variation and Nicotine Addiction

Investigating How Genes Influence Smoking Behavior



Education Outreach
Department of Genome Sciences
University of Washington

BSCS

Student Edition

Genetic Variation and Nicotine Addiction:

Investigating How Genes Influence Smoking Behavior

A Human Genome Curriculum Supplement for
High School Biology

Student Edition

Education Outreach
Department of Genome Sciences
University of Washington

Maureen Munn, Ph. D.,
University of Washington
Box 352145
Seattle, Washington 98195-2145
(206) 616-4538
mmunn@u.washington.edu

StarNet Administrative Staff
Debbie Nickerson, Ph.D., Principal Investigator, University of Washington, Seattle, Washington
Maureen Munn, Ph.D., Project Director, University of Washington, Seattle, Washington
Kristi Martinez, Project Manager, University of Washington, Seattle, Washington

Authors
Maureen Munn, Ph.D., University of Washington, Seattle, Washington
Kristi Martinez, University of Washington, Seattle, Washington

Contributing Author
Mark Bloom, Ph.D., BSCS, Colorado Springs, Colorado

Lead Teachers
The following teachers made major contributions to the development of the High School Human Genome Program and/or this curriculum unit

Janet Bowersox, Nathan Hale High School, Seattle, Washington
Bonnie dos Remedios, formerly of Roosevelt High School, Seattle, Washington
Patrick Ehrman, Institute for Systems Biology, Seattle, Washington
Edward Hopfner, formerly of Nathan Hale High School, Seattle, Washington
Mary Glodowski, Juanita High School, Kirkland, Washington
Sharon Masse, Garfield High School, Seattle, Washington
Peggy O'Neill Skinner, The Bush School, Seattle, Washington
Barbara Schulz, formerly of Lakeside School, Seattle, Washington
Mary Margaret Welch, Mercer Island High School, Mercer Island, Washington

Illustrations
Cindy Desmarais, University of Washington, Seattle, Washington
Kristi Martinez, University of Washington, Seattle, Washington
Justin Miller, Formerly University of Washington, Seattle, Washington

Web Site and Graphic Design
Megan Brown, Ph.D., University of Washington, Seattle, Washington
Cindy Desmarais, University of Washington, Seattle, Washington
Kristi Martinez, University of Washington, Seattle, Washington

BSCS Production Staff
Barbara Perrin, Production Manager, BSCS, Colorado Springs, Colorado
Lisa Rasmussen, Graphic Designer and Illustrator, BSCS, Colorado Springs, Colorado
Stacey Luce, Production Specialist, BSCS, Colorado Springs, Colorado
Pamela Warren, Project Assistant, BSCS, Colorado Springs, Colorado
Barbara Resch, Editor, Colorado Springs, Colorado

Thanks to
Maynard Olson, Director of the University of Washington Genome Center, for providing laboratory and office space for this project, and for his ongoing intellectual and technical support of project activities.

Participants of the 2003 and 2004 StarNet Summer Institutes for providing for valuable comments on this manuscript.

This project was supported by a Science Education Drug Abuse Partnership Award from the National Institute on Drug Abuse. Its content is solely the responsibility of the authors and does not necessarily represent the official views of NIDA or NIH.

Table of Contents

Student Edition

Section I:

Smoking and the *CYP2A6* Gene

Engage

What Factors Influence Our Behaviors?

Procedure

1. Look at the following photos. What behaviors do you see depicted in each photo?

Figure 1. What factors influence our behaviors?

2. Working with a partner, choose one photo and discuss factors that contribute to the person's involvement in the depicted behavior. Be prepared to share your ideas with your classmates.

3. Now discuss the following questions as a class:

   a. Are there similarities in the factors that contribute to these different behaviors?

   b. Consider people who smoke. Do people vary in their smoking habits? In what ways do they vary? What factors contribute to this variation?

Explore

No Accounting for Taste, or Is There?

Procedure

Part A: Can You Taste Phenylthiocarbamide Paper?

To a geneticist, a trait refers to the outward expression (**phenotype**) of the underlying genetic instructions (**genotype**). Some traits are influenced by a single gene, while others are influenced by many genes. Behaviors are described as **multifactorial traits** because they are influenced by many genes as well as environmental factors. In this activity, you will explore a single gene that influences the complex trait of food preference.

1. Obtain one piece of control paper and one piece of phenylthiocarbamide (PTC) paper from your teacher.

2. Taste the piece of control paper and record your perception of its taste in your lab notebook.

3. Taste the piece of PTC paper and record your perception of its taste in your lab notebook.

   If the PTC paper tastes different from the control paper, then you are a taster. If the PTC paper tastes the same as the control paper, then you are a non-taster.

4. If you are a taster, describe the taste of PTC to someone who is a non-taster.

5. As a class, assemble a histogram on the board that depicts the relative numbers of tasters and non-tasters. To do this, you will be given a yellow Post-it note if you are a taster and a pink Post-it note if you are a non-taster. You and your classmates will take turns placing Post-it notes in the appropriate columns as shown in Figure 2. Draw the completed class histogram in your lab notebook.

6. Prepare a data table in your lab notebook like the one shown in Figure 3. Calculate the percentage of the class who are tasters and the percentage that are non-tasters and enter the data in your table.

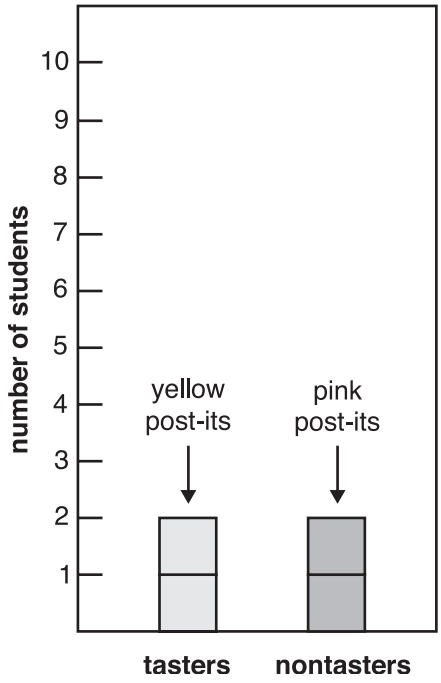


Figure 2. Number of tasters and non-tasters in class

	number of people	percentage of total
tasters		
non-tasters		
total in class		100%

Figure 3. Percentage of tasters and non-tasters in class

Part B: How Does Your Ability to Taste Affect which Foods You Like?

Does your ability to taste PTC have an effect on your food preferences? In what way? In this experiment you will test whether there is a relationship between the ability to taste PTC paper and whether or not a person likes the taste of black licorice.

1. Make a prediction: Are people who can taste PTC paper more or less likely to enjoy the flavor of black licorice? Why or why not? Write your prediction and reasoning in your lab notebook.

2. Do you like black licorice? Taste the piece of licorice offered by your teacher and record whether or not you like its taste in your lab notebook.

3. As a class, assemble a histogram on the board that depicts how tasters and non-tasters of PTC paper feel about the taste of black licorice. As before, you will be given a yellow Post-it note if you are a taster and a pink Post-it note if you are a non-taster. You and your classmates will take turns placing Post-it notes in the appropriate columns, as shown in Figure 4. Draw the completed class histogram in your lab notebook.

4. Prepare a data table such as that shown in Figure 5 in your lab notebook. Calculate the percentage of tasters who like black licorice and the percentage of nontasters who like black licorice.

5. Describe your results. What percentage of the people who can taste PTC like the flavor of black licorice? What percentage of the people who cannot taste PTC like the flavor of black licorice? Do these results agree with your prediction? Why or why not?

6. With a partner, discuss factors that contribute to whether you like or dislike the taste of black licorice. Which of these factors are genetic? Which are environmental?

7. Think about how your sense of taste affects your health. In what ways does it contribute to health? Does it ever work against good health?

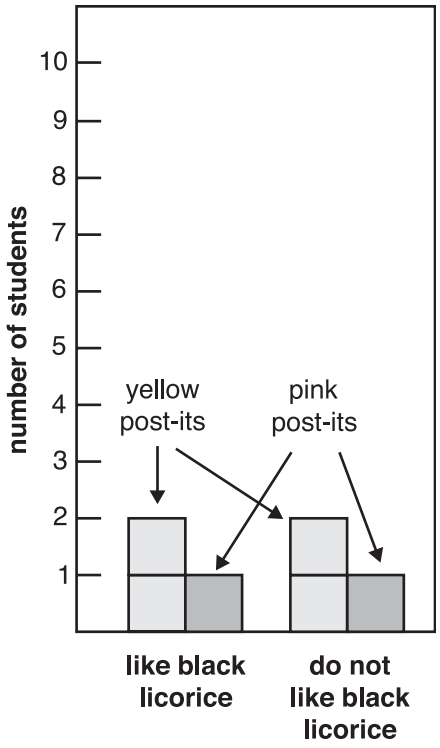


Figure 4. Effect of PTC tasting ability on black licorice preference

Figure 5. Relationship between ability to taste PTC and liking or not liking black licorice

	Taster	Non-taster
PTC tasting ability (number of students)		
Like black licorice (number of students)		
Do not like black licorice (number of students)		
Percentage who like black licorice*		
Percentage who do not like black licorice		

*Percentage of tasters who like black licorice = (number of tasters who like black licorice ÷ total number of tasters) x 100

Explore/Explain

Exploring What We Know about Smoking and Addiction

In the previous activity, you investigated how a single gene contributes to the ability to taste a specific chemical. Now you will consider the more complicated trait of smoking. Smoking is a complex behavior that involves your sense of taste as well as other body systems.

Procedure

1. Think about the following questions, share your ideas with a partner, and then participate in a class discussion:

   a. What is a **drug**?

   b. What is **drug abuse**?

   c. What is **drug addiction**?

   d. What is **nicotine**?

   e. What evidence do you have that smoking tobacco is addictive?

   f. Discuss the health hazards of using tobacco (smoking cigarettes, cigars, or pipes or chewing tobacco).

   g. What factors contribute to smoking being addictive?

2. Think about where nicotine goes once it enters the body. Your teacher will give you a diagram of the human body (Figure 6). Use it to trace the pathway that nicotine follows from the time the person inhales it until it leaves the body.

3. Circle places in the pathway where individuals might vary in the way they react to nicotine.

4. When nicotine reaches the brain does it have any influence on the brain, or does the nicotine simply pass through the brain on its way out of the body?

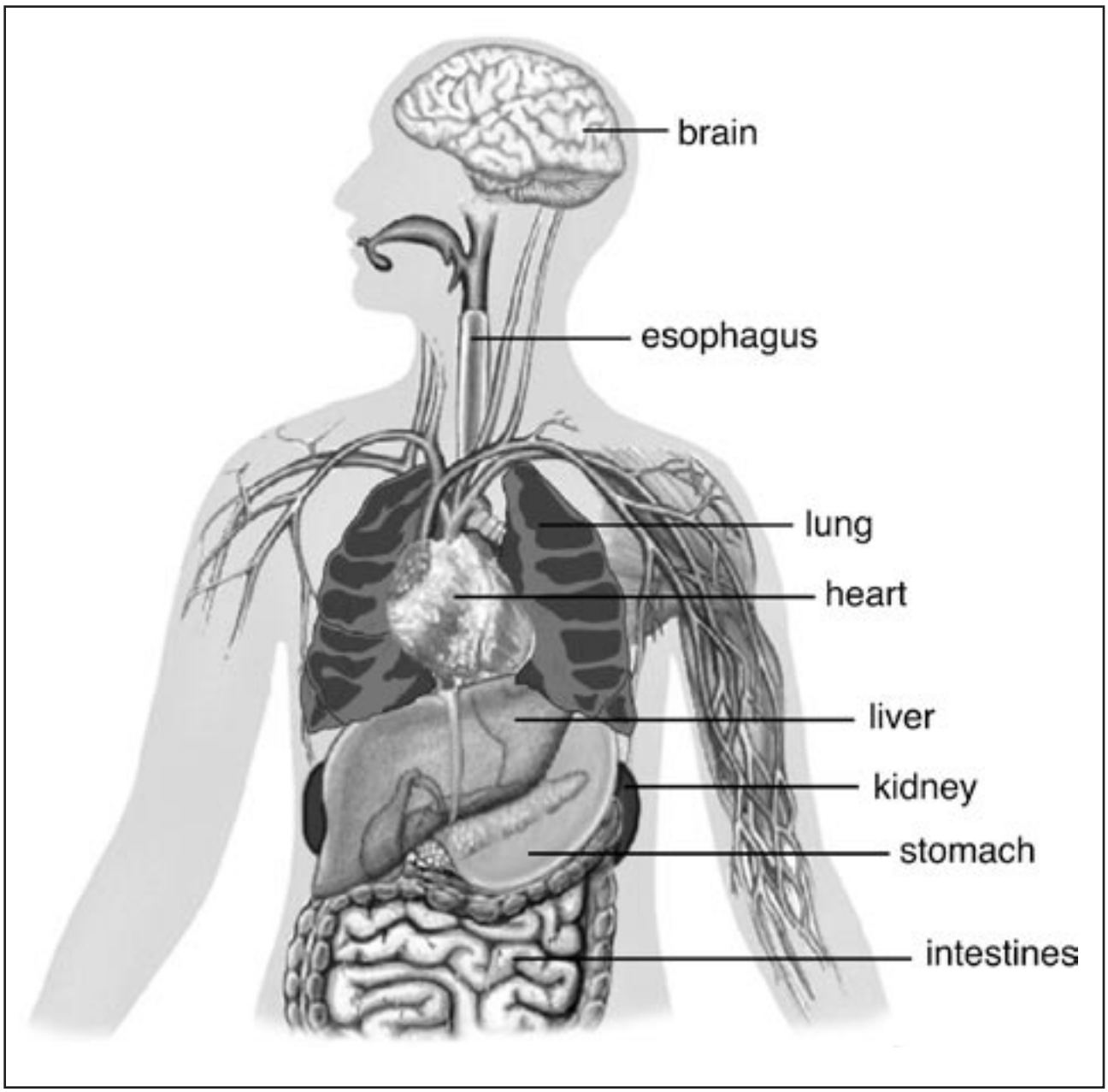


Figure 6. Tracing nicotine's path through the body

Elaborate

Reading: Nicotine, Neurotransmission, and the Reward Pathway

You have been introduced to nicotine as a drug of addiction. Now you will examine how nicotine affects the brain's ability to transmit messages from one nerve cell to another. You are going to watch a video that introduces nerve cells, shows how they communicate with each other, and relates this to a neural pathway in the brain that helps to reinforce important behaviors. Before watching the video, read the material below. As you watch the video, answer the questions at the end of this section.

How Do Drugs of Addiction Interact with Our Brains at the Molecular Level?
Drugs of addiction interfere with a natural neural pathway in the brain called the **reward pathway**. The reward pathway acts to reinforce certain behaviors that are important for survival such as eating and drinking. To understand how the reward pathway does this, we first need to examine the structure of nerve cells and learn about the process of **neurotransmission**.

Neurotransmission
There are billions of nerve cells, or **neurons**, in the brain. As shown in Figure 7, each neuron is made up of three parts: the central **cell body**, which directs the activity of the cell; **dendrites**, the short fibers extending from the cell body that receive messages from other neurons; and the **axon**, the single, long fiber that extends from the other side of the cell body and transmits messages to other neurons or other tissues, like muscle.

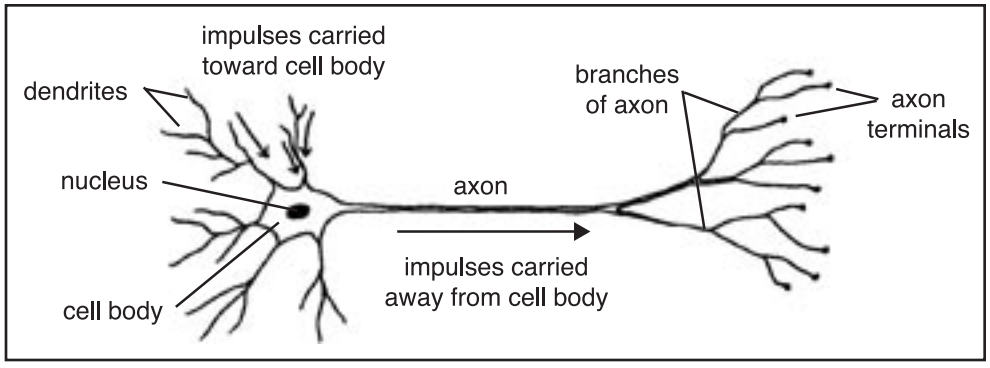


Figure 7. A typical neuron

Neurotransmission is the process of transferring a message from the axon of one nerve cell, across a small space or **synapse**, to the dendrites of a nearby neuron. This message is transmitted by a chemical substance called a **neurotransmitter**. A message from a nerve cell travels down its axon as an electrical impulse, triggering the release of a neurotransmitter at the end of the axon (see Figure 8). The neurotransmitter crosses the synapse and binds to its specific receptor on the dendrites of a nearby nerve cell. The binding of a neurotransmitter results

in either stimulation or inhibition of an electrical impulse in the receiving cell. After binding to its receptor, the neurotransmitter is quickly inactivated, either through enzymatic breakdown or by being reabsorbed by the axon that released it through special transporter molecules located on the cell membrane.

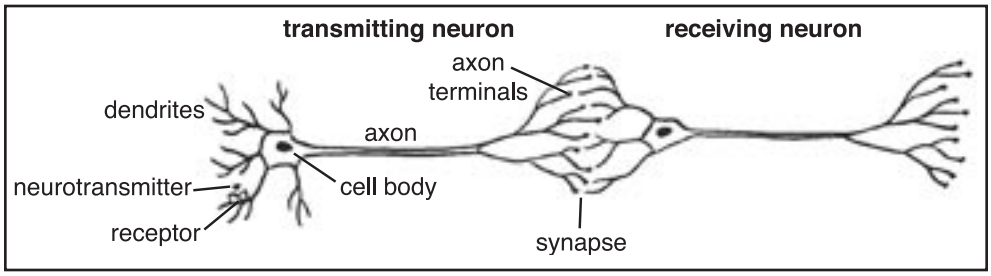


Figure 8a. A neural synapse

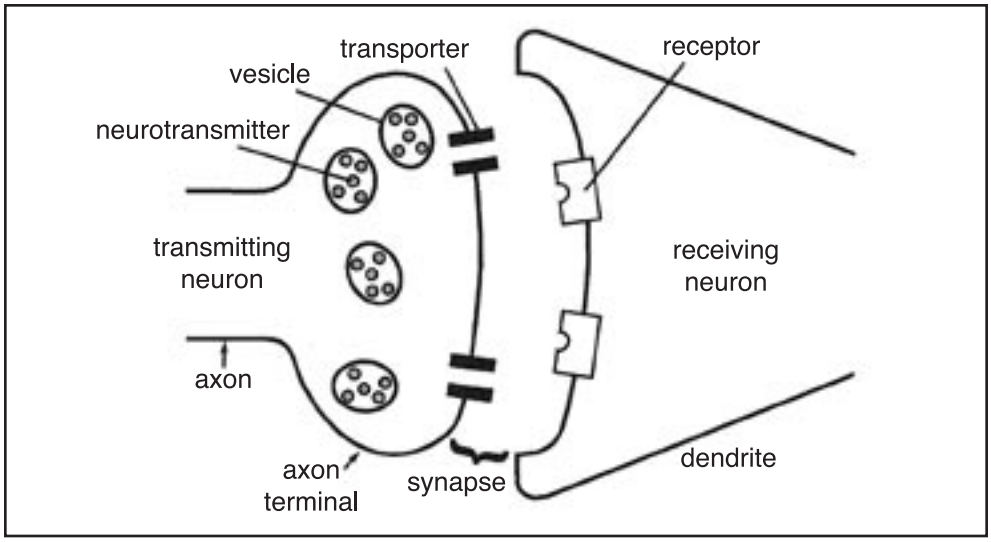


Figure 8b. Close-up of a neural synapse, showing an axon terminal of a transmitting neuron interacting with a dendrite from a receiving neuron

The Reward Pathway

Certain behaviors that are important for survival (e.g. eating, drinking, and sexual activity) are reinforced by the body through the reward pathway (see Figure 9). These behaviors stimulate a specialized set of neurons in the brain that create the sensation of pleasure. One part of the reward pathway consists of neurons in the **ventral tegmental area (VTA)** of the brain (located just above the brain stem). These neurons use the neurotransmitter **dopamine** to stimulate neurons in other parts of the brain. Once stimulated, a neuron in the VTA sends an electrical impulse down the nerve axon. At the end of the axon, vesicles containing dopamine fuse with the cell membrane, releasing dopamine into the synapse (see Figure 8b). Receiving neurons on the other side of the synapse include cells of the **nucleus accumbens**, a part of the emotional center of the brain (also called the limbic system), and neurons of the frontal region of the cerebral

cortex. After its release, dopamine is quickly reabsorbed by the cell that released it by a specialized pump called the **dopamine transporter**.

How Does Nicotine Interact with Neurotransmission and the Reward Pathway?

When cigarette smoke is inhaled, the nicotine in the smoke is absorbed into blood vessels in the lungs and reaches the brain within 10 seconds of the first puff. Once in the brain, nicotine binds to certain receptors located on the cell bodies of neurons in the ventral tegmental area, as well as the terminals of these neurons, which are situated in the nucleus accumbens. Normally, these receptors bind the neural transmitter acetylcholine. Nicotine is similar in structure to acetylcholine, so it is able to bind in place of acetylcholine on its receptor. To simplify our discussion, we will refer to the nicotine receptor, but you should remember that the normal function of this receptor is to bind acetylcholine.

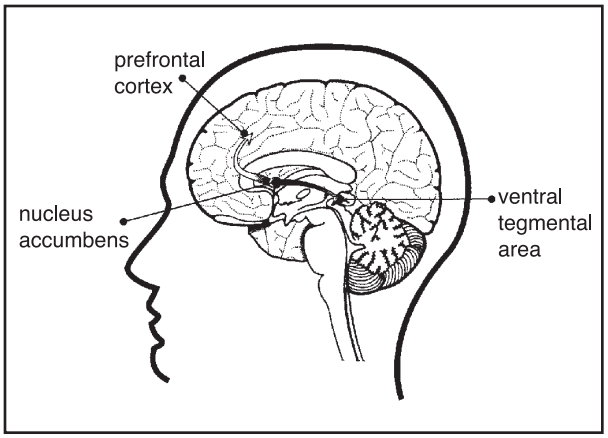


Figure 9. The reward pathway in the brain

The nicotine receptor is a transmembrane protein made up of five subunits (shown in Figure 10). The binding of either nicotine or acetylcholine to the receptor results in the brief opening of a pore in the receptor. This opening allows cations to move into the neuron. This influx of cations triggers an electrical impulse to move down the nerve axon. At the terminal of the neuron, dopamine is released and bound by receptors of nerve cells in the nucleus accumbens and the prefrontal cortex (Figure 11). This creates a sensation of pleasure through a process that is not well understood.

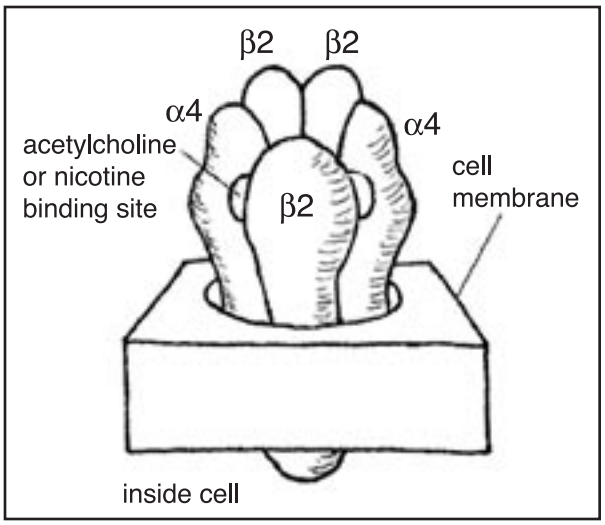


Figure 10. The acetylcholine/nicotine receptor

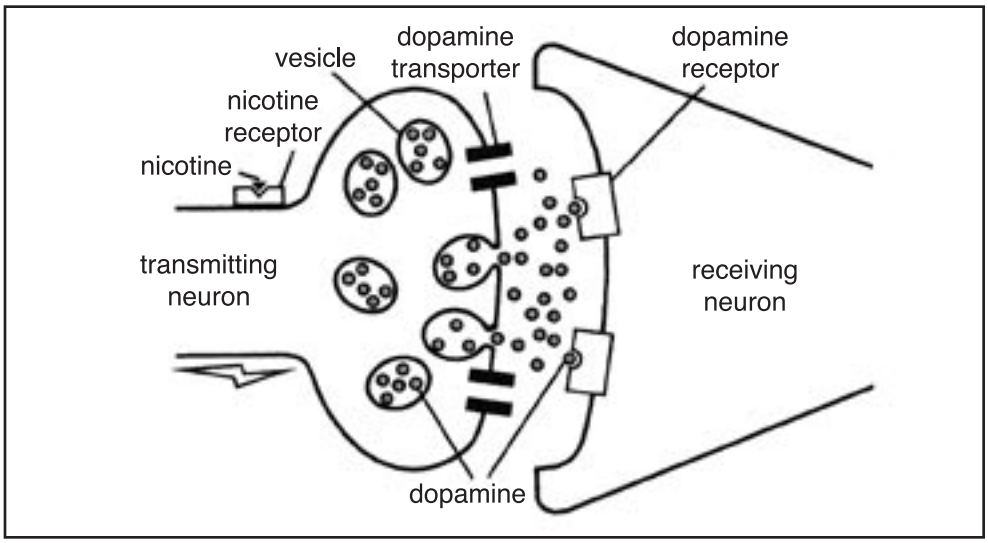


Figure 11. Neurotransmission in the reward pathway

Questions to Consider while Viewing the Video.

1. What kind of signal is used to transmit a message from one end of a neuron to its other end?

2. What kind of signal is used to transmit a message between two neurons?

3. What is the molecular target of each of the following drugs of abuse:

   a. Nicotine

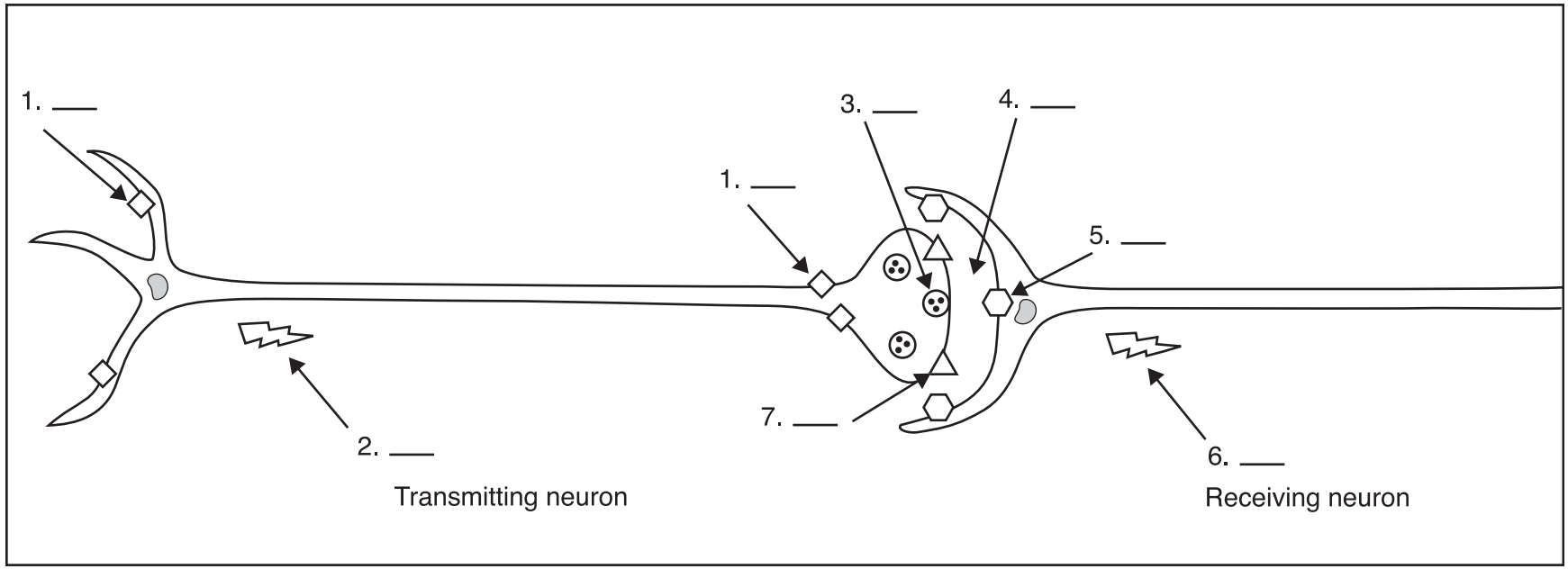
   b. Cocaine

   c. Marijuana

Opportunity for Assessment

Your teacher will give you a diagram that depicts the effect of nicotine on neurotransmission. Match steps 1 through 7 on the diagram with their corresponding descriptions (a through g).

Match steps 1 through 7 with their correct descriptions (a through g).



◇ = nicotine receptor

⚡ = electric impulse

⊙ = vesicle of dopamine

⬡ = dopamine receptor

△ = dopamine transporter

a. Electric impulse travels along axon away from cell body.
b. Dopamine transporter takes up dopamine from synapse.
c. Vesicles of dopamine fuse with cell membrane.
d. Nicotine binds to receptors on transmitting neuron.
e. Dopamine binds to receptors on receiving neuron.
f. Electric impulse traveling away from receiving neuron increases or decreases.
g. Dopamine released into synapse.

Figure 12. Effect of nicotine on neurotransmission

Elaborate

Reading: Metabolism of Nicotine

As you learned in the previous activity, nicotine mimics the neurotransmitter acetylcholine, binding to its receptors on neurons in the brain. However, unlike acetylcholine, nicotine is not broken down in the brain. As you will learn in this reading, nicotine is processed in the liver and excreted in the urine.

The liver carries out many important functions:

1. Production of bile, a liquid that is stored in the gallbladder and released in the small intestine. Among other things, bile contains bile salts, which aid in the digestion and absorption of fats.
2. Processing nutrients from the small intestine into substances that are used in other cells. The liver makes, breaks down, and stores many substances, including carbohydrates, fats, and blood plasma proteins.
3. Clearing the blood of drugs and other toxic substances. After breakdown in the liver, these substances enter the bile or bloodstream. Substances that enter the bile go into the gut and leave the body in the feces. Breakdown products that enter the blood are filtered out by the kidneys and are excreted in the urine.

The liver produces a class of enzymes called **cytochrome P450**s that are involved in the breakdown of toxins such as metabolic by-products, steroids, and drugs, including therapeutic drugs and drugs of abuse. Through the action of these enzymes, toxic substances are made more water soluble so that the body can excrete them in the urine (see Figure 13).

An important member of this enzyme family, called **CYP2A6**, converts nicotine to a substance called **cotinine**, which is modified by other enzymes and then excreted. About 70 percent of the nicotine in a person's bloodstream is converted to cotinine by CYP2A6. The rest of the nicotine is treated by other enzyme pathways before being excreted. What would happen if a person did not make the enzyme CYP2A6? The following activity, *The Case of the Missing Enzyme*, will help you to answer this question.

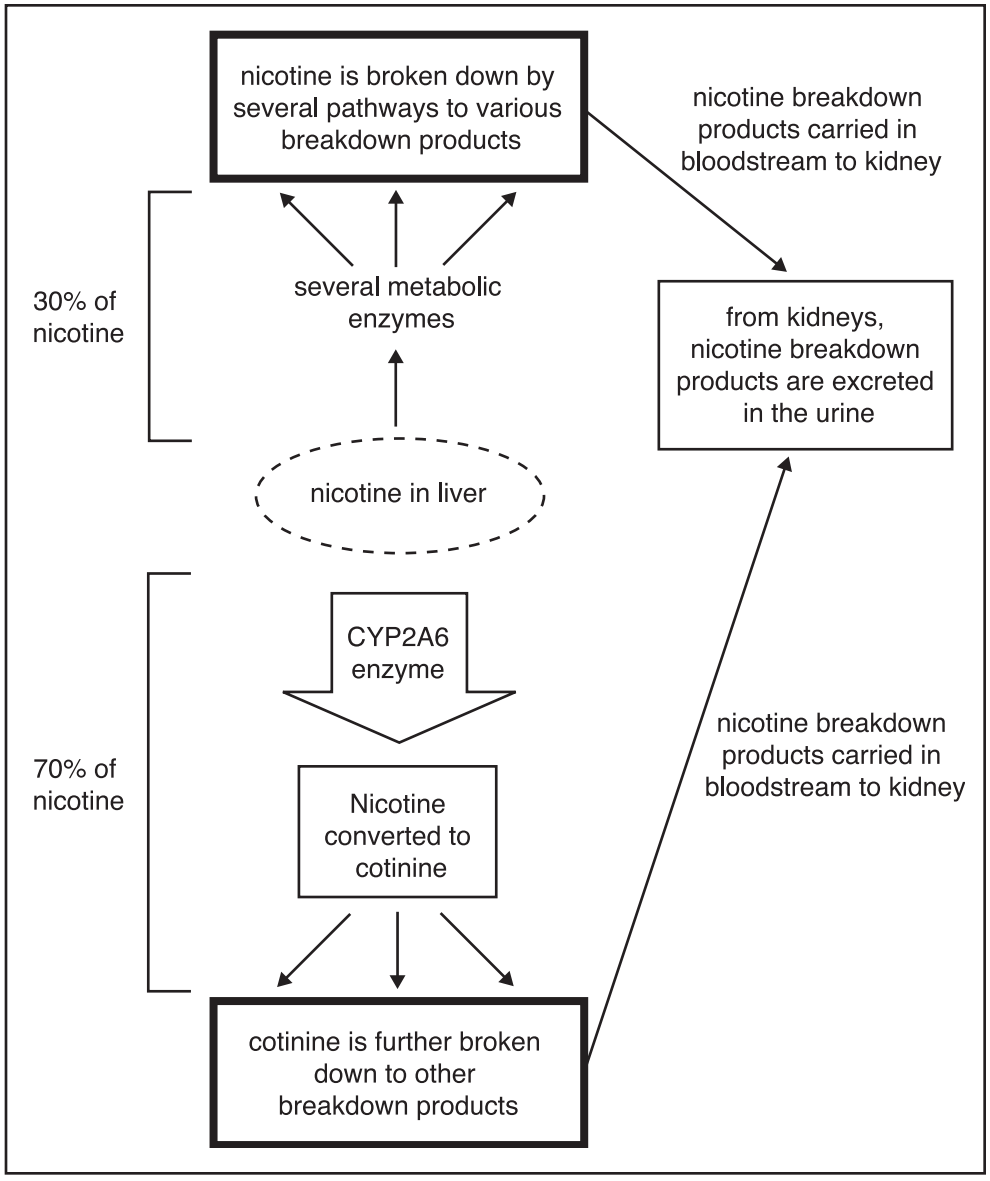


Figure 13. Nicotine is processed in the liver. In most individuals, about 70 percent of the nicotine is converted to cotinine by the CYP2A6 enzyme.

Elaborate

The Case of the Missing Enzyme

The previous reading introduced the enzyme CYP2A6 as important to the metabolism of nicotine in the body. In this activity, you will model how nicotine is processed in the liver of a person who has a normal amount of the enzyme CYP2A6 and compare it with a person who is missing this enzyme.

Materials

1 cup beans or beads (navy beans or kidney beans or beads of a similar size)
2 similar boxes, one with 2 circular holes in bottom (3 cm and 2 cm in diameter, see diagram in step 1)
1 timer or watch with second hand
1 250-mL graduated cylinder or a calibrated beaker
1 roll of cellophane tape

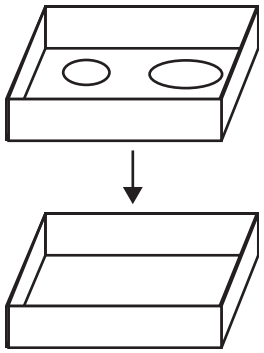
Procedure

In Part A, you will model the processing of nicotine in the liver of a person who has a normal amount of the CYP2A6 enzyme. In Part B, you will model the processing of nicotine in the liver of a person who is missing the CYP2A6 enzyme.

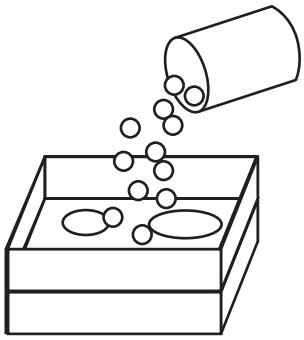
Part A:

Person with a Normal Amount of the CYP2A6 Enzyme

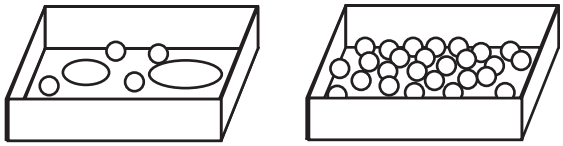
1. Place the box with holes on top of the other box.



2. Pour all of the beans (or beads) into the top box and shake for 5 seconds.



3. Remove the top box and set it on the table so any remaining beans don't spill out. Use a graduated cylinder to measure the volume of the beans in the lower box and record the volume in your lab notebook. (Copy a data table in your lab notebook as shown in Figure 14.)

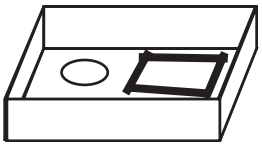


Number of holes in top box	Volume of beans (in mL) in the lower box after 5 seconds of shaking
2	
1	

Figure 14. Effect of open holes in box on number of beans passing to lower box

Part B:
Person Who Is Missing the CYP2A6 Enzyme

1. Cover the larger hole with a piece of cardboard. Tape the cardboard in place. Repeat steps 1–3 as described in Part A.



Questions for Understanding

Answer the following questions in your lab notebook.

1. How did covering one hole affect the number of beans (or beads) that passed through the top box into the lower box?

2. Relate your results with this simple model system to nicotine metabolism in the liver of an individual who has the CYP2A6 enzyme and one who lacks the enzyme.

3. How might a difference in CYP2A6 level affect a person's smoking behavior?

Elaborate

Variation in Proteins That Interact with Nicotine

Think back to your earlier discussion of the many ways that people vary in how they smoke tobacco. Scientists are interested in learning whether these behavioral differences are caused by variations in one or more of the proteins that interact with nicotine inside the body.

Procedure

1. Your teacher will ask you to revisit two diagrams: one that depicts the effect of nicotine on neurotransmission (Figure 12) and a second one that depicts the metabolism of nicotine (Figure 13). Circle areas on the diagrams that show where proteins are involved whose functions may contribute to variations in people's smoking patterns.

2. With your partner, discuss why you chose those proteins.

Consider This

Differences in proteins result from differences in the genes that code for them. Genetic variation accumulates over many generations through mutations. These include insertion or deletion of one or more nucleotides or a single nucleotide change. This last type of mutation is the most common cause of genetic variation. Single nucleotide differences among different individuals are called **single nucleotide polymorphisms** (**SNPs**).

If you were a scientist, how would you propose to find SNPs in genes related to nicotine addiction?

Section II:

Genetic Variation and the *CYP2A6* Gene

An Invitation from the StarNet Team

StarNet Project
Department of Genome Sciences
University of Washington
Seattle, Washington

Dear Students,

The StarNet Project invites you to participate in an inter-high school collaboration based on discovering genetic variation related to nicotine addiction. You and your classmates will be sequencing part of the *CYP2A6* gene, which codes for an enzyme involved in clearing nicotine from the body. Scientists have hypothesized that people who make less or none of this enzyme will have altered smoking behavior compared with people with a normal amount of this enzyme.

Several variations of the *CYP2A6* gene have been identified, and many of these encode inactive enzymes. Is there a correlation between which two variations (alleles) of this gene people have (and thus how much active CYP2A6 enzyme their bodies produce) and their smoking behavior? Several studies say yes. One research group demonstrated that people who make less CYP2A6 enzyme are less likely to be tobacco-dependent smokers. Smokers with reduced CYP2A6 enzyme smoke significantly fewer cigarettes per day than people with normal levels of this enzyme. A later study confirmed this result and showed that people who have a duplication of the gene in one of their alleles (and thus make more CYP2A6 enzyme) are heavier smokers than those who have two normal alleles. However, other studies do not support this interpretation.

The student sequencing project seeks to find new variations in the *CYP2A6* gene. Students are sequencing large portions of the *CYP2A6* gene from eight different people. These eight people are part of the Human Polymorphism Discovery Panel, assembled by scientists at the National Institutes of Health to aid studies on genetic variation. These individuals have given informed consent to use their DNA in such studies. Their identities are kept anonymous. The eight individuals in our study represent different ethnic groups found in the United States. Thus, it is quite likely that we'll identify some genetic differences. Within this population, we may find polymorphisms (SNPs) that have already been identified. However, we may also discover new ones. When the sequencing is completed, the student data will be submitted to GenBank, a DNA database that is used by scientists throughout the world.

We hope that your participation will deepen your understanding of DNA, genes, genetic variation, and the nature of science. Some of you may already know that you want to pursue a career in biological research. We hope that taking part in this project will open your eyes to the potential of scientific research to understand human genetics and the biology of drug addiction.

Sincerely,

The StarNet Team

Explore/Explain

DNA Structure, Synthesis, and Sequencing

A Review of DNA Structure

Now that you have accepted StarNet's invitation to help them sequence the *CYP2A6* gene, you will need to review what you know about the structure of DNA. The technique you will use to sequence DNA is based on the process that the cell uses to synthesize DNA. Therefore, following your review of DNA structure and its synthesis, you will apply your knowledge to two activities; the first models DNA synthesis in the cell and the second models DNA sequencing in the test tube.

Procedure

1. Your teacher will give you two different diagrams of DNA. Review them and check your understanding with a partner.

2. To assess your knowledge, define the following terms and label where they occur on Figures 15 and 16. Check your partner's work and discuss any areas of disagreement.

   a. Double helix

   b. Nucleotides (A, C, G, T)

   c. Complementary base pairs

   d. Antiparallel strands (5´ to 3´)

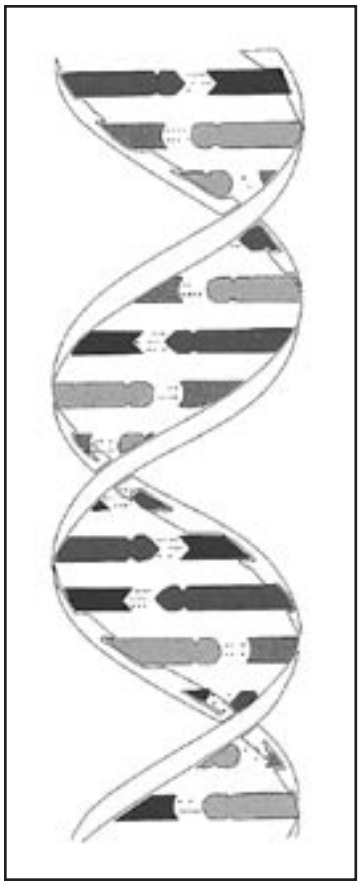


Figure 15. Structure of double-stranded DNA

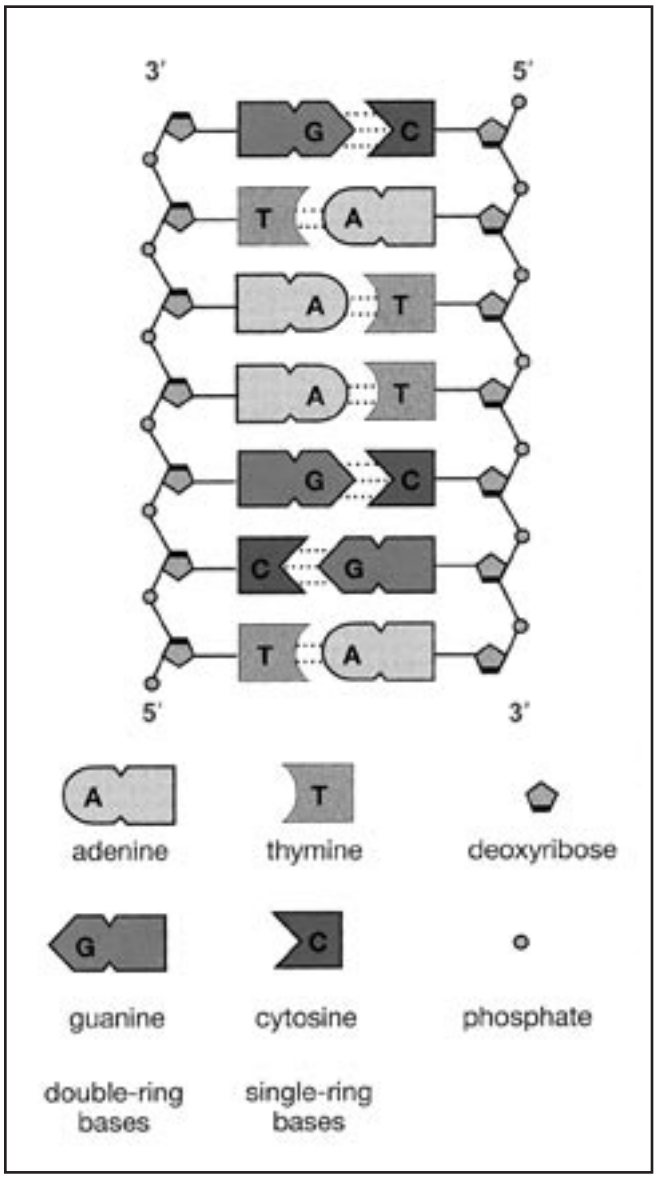


Figure 16. The DNA helix is unwound and flattened in this view.

Explore/Explain

A Review of DNA Synthesis

To appreciate how DNA is sequenced in a test tube, you first must understand how DNA is synthesized in a cell.

Procedure

1. Read the following two paragraphs and check your understanding with a partner.

   DNA Synthesis in the Cell
   Inside a cell, it takes over 20 different proteins to replicate DNA. The most important protein is **DNA polymerase**, the enzyme that assembles a new DNA strand by copying an existing DNA strand called the template strand. Inside a cell, DNA polymerase starts synthesis at a place where a short segment of RNA, called a **primer**, is bound through complementary base pairing to the DNA template. As shown in Figure 17, during DNA synthesis, a nucleotide is added to the 3′ end of the primer through the formation of a covalent bond between the 5′ end of the incoming nucleotide and the 3′ OH on the primer. DNA polymerase catalyzes the formation of this high energy bond. The incoming nucleotide is the complement of the next nucleotide on the template strand. The next nucleotide to be added is joined to the nucleotide that was just attached, and so on until the end of the DNA template is reached.

   Other proteins involved in DNA synthesis carry out functions such as unwinding the DNA and keeping it unwound, synthesizing the RNA primer, helping DNA polymerase to get started and to stay on the DNA, and finishing the job by removing RNA primers and filling in any gaps.

2. Read the following two paragraphs and check your understanding with a partner.

   Scientists can synthesize DNA in a test tube using just one protein—DNA polymerase. All of the ingredients needed for the process are added to a test tube containing a buffered solution that provides the cell-like environment that is needed for the DNA polymerase to function properly.

   The following ingredients are required:
   - DNA template strand
   - DNA primer (Scientists substitute a DNA primer for the RNA one used in the cell because it is more stable.)
   - **deoxynucleotides** (A, C, G, and T) (From now on we will refer to these as nucleotides, which is the widely used term for them.)
   - DNA polymerase
   - Buffered solution

3. To assess your knowledge, identify the ingredients needed for DNA synthesis on Figure 17 (excluding DNA polymerase and the buffered solution). Check your understanding with a partner and discuss any areas of disagreement.

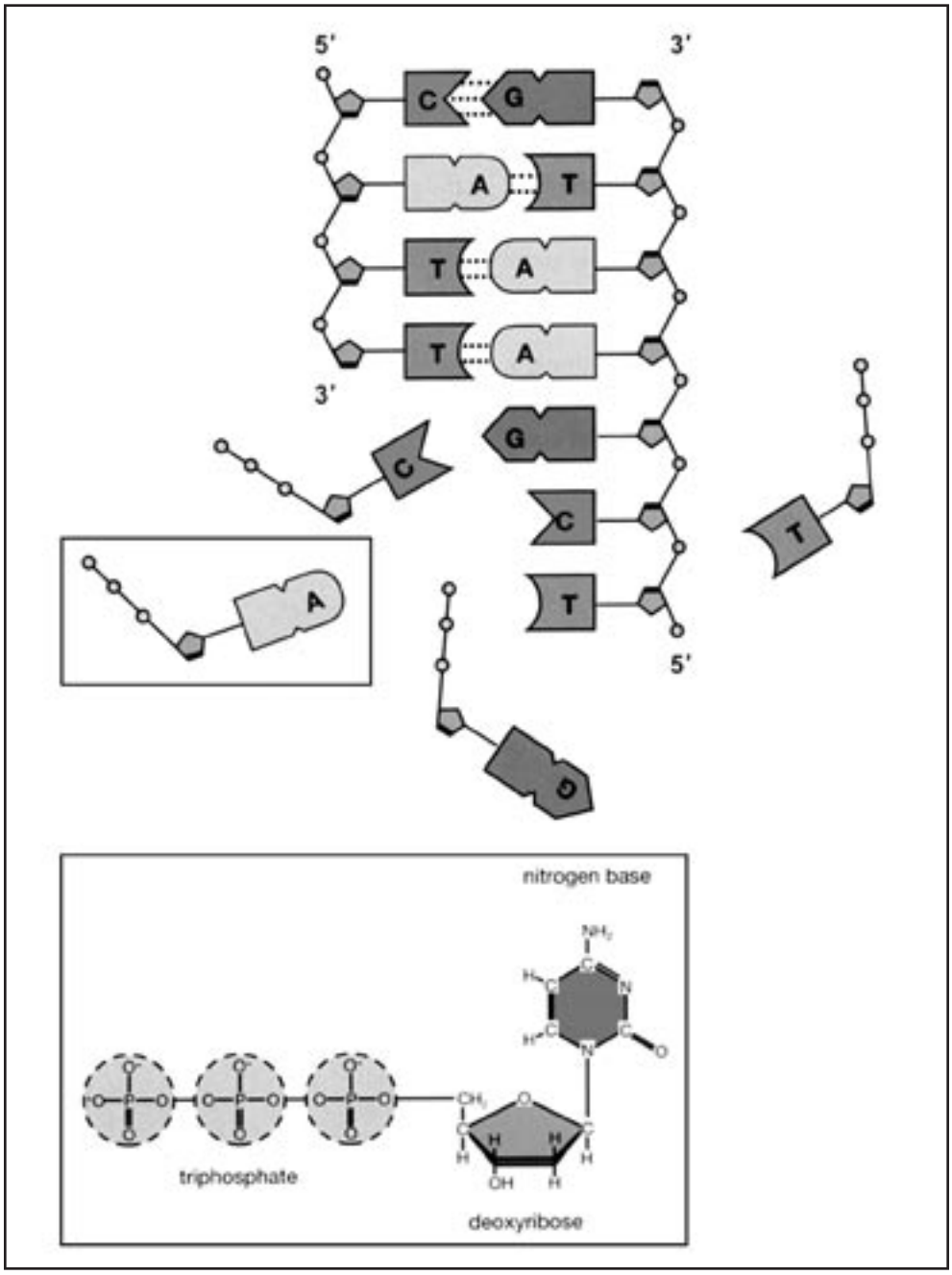


Figure 17. Addition of a nucleotide during DNA synthesis. A covalent bond is formed between the 3′ OH on the primer end and the first phosphate on the 5′ carbon of the incoming nucleotide. The other two phosphates are released into solution.

Explore

Modeling DNA Synthesis Using Pop-It Beads

To complete your review of DNA synthesis, you will model this process using pop-it beads. As you complete the activity, be sure you can relate the characteristics of the model to the structure of DNA.

Materials

Paper strip with DNA sequence = DNA template
4 joined beads = DNA primer
Individual pop-it beads of 4 colors = nucleotides (A is green, C is blue, G is yellow, and T is red)
Cellophane tape = hydrogen bonds
Masking tape and pen to label primer
Scissors

Procedure

1. Your teacher will provide you with a figure that contains a DNA template. Cut out the DNA template.

2. Use a piece of masking tape to label your bead primer with your group number or name.

3. Match your bead primer to the paper DNA template strand at the complementary sequence. The pop-it end of the bead should point to the right. Use the bead color code (listed under Materials section) to decide where the primer should bind. Use cellophane tape to hold the primer in place.

The DNA template is the piece of DNA being copied. A primer is a piece of DNA or RNA that binds to the DNA template through complementary base pairing. During DNA synthesis, the first nucleotide is added to the 3' end of the primer.

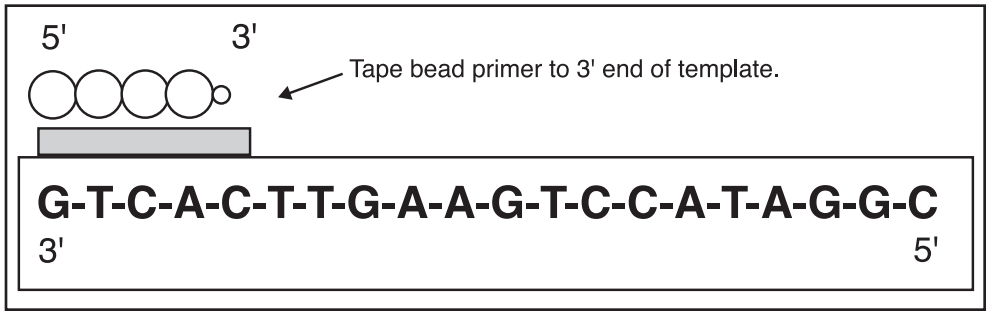


Figure 18. Primer binding to the DNA template strand

4. Now you will play the role of DNA polymerase. The beads are the nucleotides. Remember:
   - Green bead = A
   - Blue bead = C
   - Yellow bead = G
   - Red bead = T

   The hole in the bead represents the 5′ end of the nucleotide, and the pointy end represents the 3′ end (OH group) of the nucleotide.

   To start synthesizing the new DNA strand, you need to choose a bead of a color that is the *complement* of the nucleotide on the DNA template (as shown by the arrow in Figure 19). What color bead do you need? Remember that A pairs with T and C pairs with G. Add the bead to the pointy end of the primer. Tape the bead to the paper template to hold it in place. Now add the next bead, again making sure that it is complementary to the next nucleotide on the DNA template. Continue adding complementary beads until you reach the end of the template.

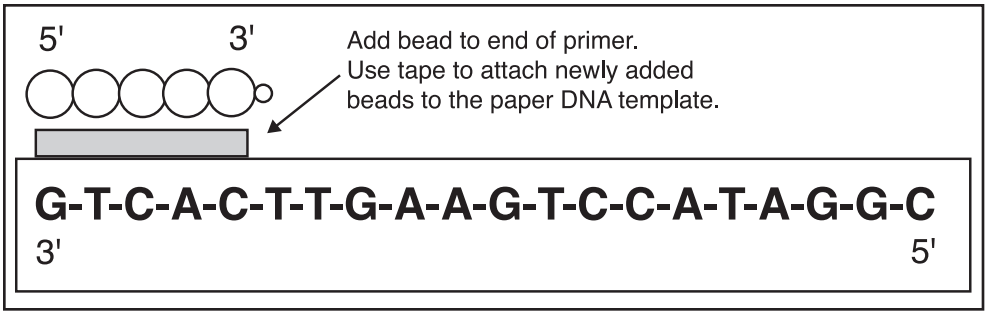


Figure 19. Incorporating nucleotides into the growing DNA chain

During DNA synthesis, DNA polymerase adds nucleotides one at a time to form the new DNA strand. Each nucleotide added to the chain is complementary to the nucleotide across from it on the template. The direction is always the same; the 5′ end of the incoming nucleotide is added to the 3′ end of the new DNA strand. DNA polymerase continues until it reaches the end of the DNA template.

5. Compare the sequence of your bead strand to the one made by your lab partner. Do they look the same? If not, decide whether one of you has made a mistake and correct it.

6. After the activity is completed, take the beads apart. Please leave the four beads of the primer attached to each other. Return the beads and primers to your teacher.

G-T-C-A-C-T-T-G-A-A-G-T-C-C-A-T-A-G-G-C

3' 5'

G-T-C-A-C-T-T-G-A-A-G-T-C-C-A-T-A-G-G-C

3' 5'

Template strands: cut along dotted lines.

Color Code
A = green bead
C = blue bead
G = yellow bead
T = red bead

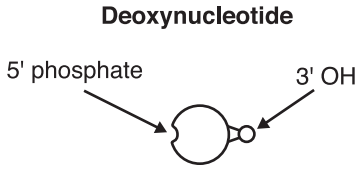


Figure 20. DNA templates

Explore/Explain

Chain Termination DNA Sequencing

Recall that the method used to sequence DNA is closely related to the process of DNA synthesis. In the 1970s, a British scientist named Fred Sanger developed a technique for sequencing DNA that is called the **chain termination** method. Recently, scientists used a modified and automated version of his method to obtain the complete sequence of the human genome.

Procedure

1. Read the following two paragraphs and check your understanding with a partner.

   Chain termination DNA sequencing is based on the DNA synthesis process. It requires a DNA template strand, a DNA primer complementary to the DNA template, nucleotides, and DNA polymerase. As you just learned, the nucleotides used during DNA synthesis are called deoxynucleotides (Figure 21). The prefix "deoxy" means "without oxygen," and this name reflects the fact that these nucleotides do not have an oxygen atom attached to the 2′ carbon atom (unlike the nucleotides of RNA that have an OH group at this position). In addition, DNA sequencing requires the use of **dideoxynucleotides**, modified nucleotides that lack an oxygen atom on the 3′ carbon as well as the 2′ carbon (Figure 21). These nucleotides can be added onto a growing DNA strand, but because they are missing the 3′ OH group, the next nucleotide cannot be attached. For this reason, dideoxynucleotides are also called chain terminators.

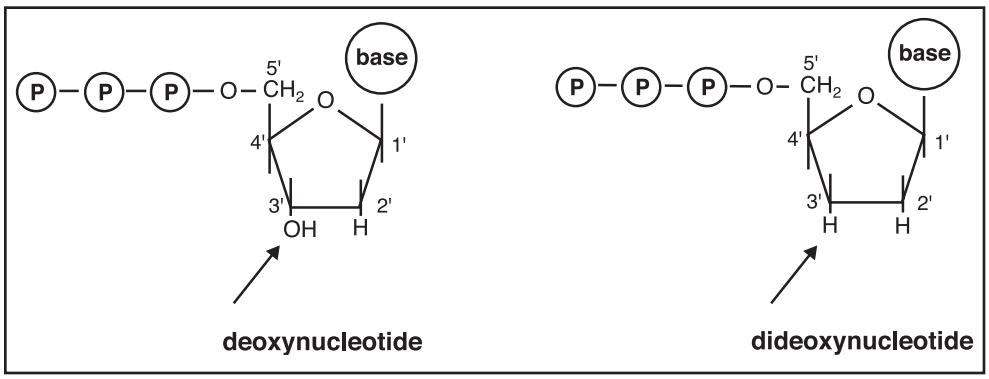


Figure 21. Structure of a deoxynucleotide and a dideoxynucleotide. The arrows indicate the single difference in the structures of these two molecules.

How do chain terminators help us sequence DNA? For each DNA template being sequenced, four different DNA synthesis reactions are needed: an A, C, G, and T reaction (Figure 22). Each reaction contains the DNA template, DNA primer, the four deoxynucleotides (A, C, G, and T), DNA polymerase, and a small amount of one of the four dideoxynucleotides (dideoxyA in the A reaction, dideoxyC in the C reaction, etc.). Refer to

the top portion of Figure 22. Starting at the primer, DNA polymerase adds nucleotides to each growing DNA strand until it incorporates a dideoxynucleotide, which causes chain termination. In the A reaction, all the new DNA fragments terminate in A, in the C reaction, the new fragments terminate in C, and so on. Millions of new DNA fragments are made in each reaction to ensure that there is a new fragment for every nucleotide in the DNA template. The length of each fragment is a *measurement* of the distance from the primer to one particular nucleotide position. The four reactions are loaded onto separate lanes in a sequencing gel, and electrophoresis is carried out to separate the fragments according to their length. Then the DNA is visualized by staining. The resulting sequence is read by starting at the lowest DNA band on the gel and identifying which lane it is in (A, C, G, or T). This tells what nucleotide this band corresponds to. The next highest band is the next nucleotide in the DNA sequence, and so on, up the length of the gel.

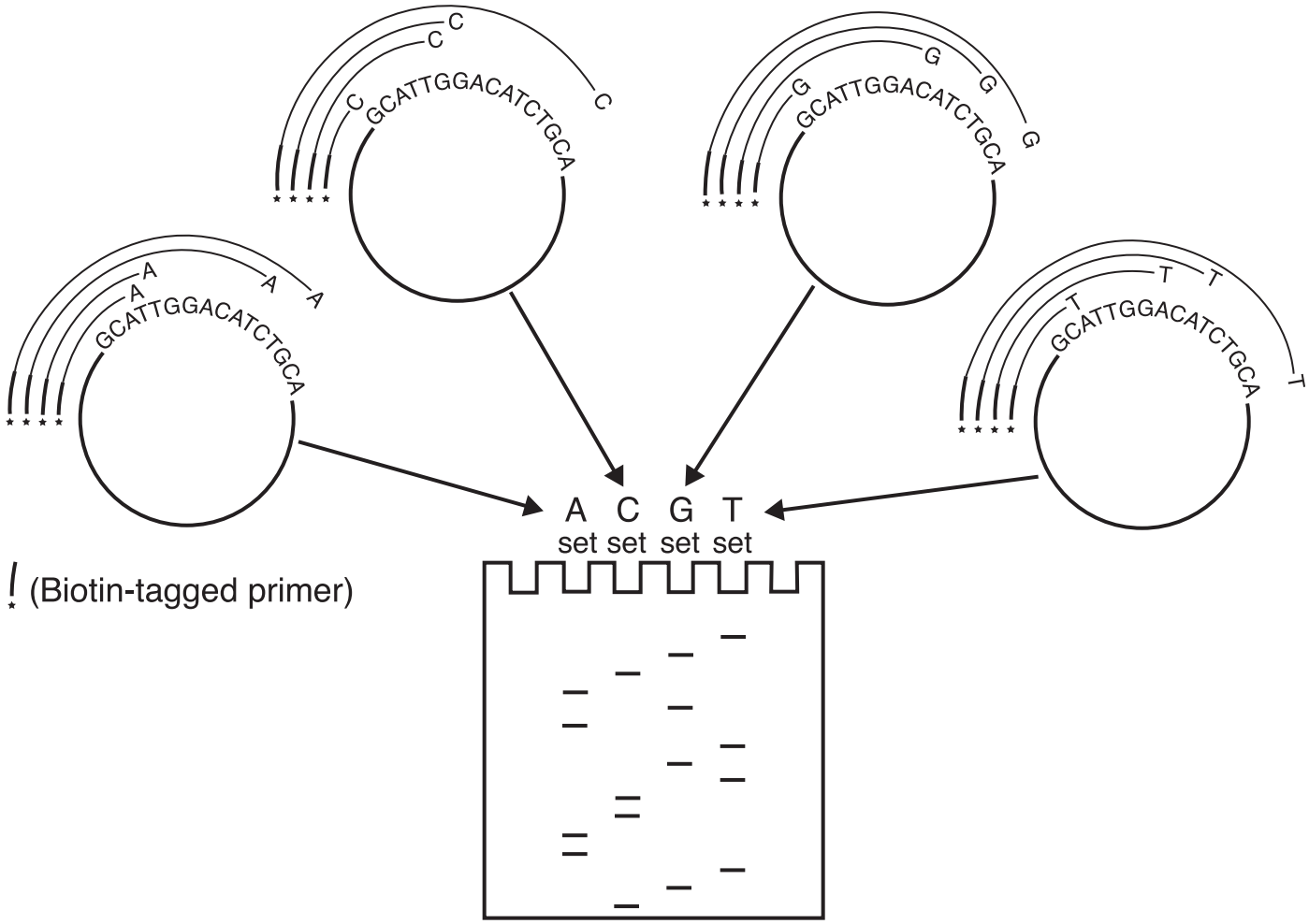


Figure 22. Chain termination DNA sequencing. Four sets of DNA fragments are produced from four separate reactions. In the A reaction, all the new fragments terminate in a dideoxyA, in the C reaction, they all end in a dideoxyC, and so on. The new fragments are separated according to length by gel electrophoresis and then stained. The DNA sequence is read from the bottom to the top, starting at the lowest band. The identity of this band is given by the lane it is in (in this example, C). The next nucleotide in the sequence is the next highest band on the gel (G).

Explore/Explain

Modeling DNA Sequencing Using Pop-It Beads

How is DNA sequencing different from DNA synthesis? In this activity, you will model the use of chain terminators to determine the sequence of bases in a short DNA strand.

Materials

Paper strip with DNA sequence = DNA template
4 joined beads = DNA primer
Pop-it beads of 4 colors = nucleotides (A is green, C is blue, G is yellow, and T is red)
A few pop-it beads of the same color that have the pop-it end removed = dideoxynucleotide = chain terminator
Cellophane tape = hydrogen bonds
Masking tape and pen to label primers
Scissors

Procedure

Part A: DNA Synthesis with Chain Terminators

1. Your teacher will provide you with a figure that contains a DNA template. Cut out the DNA template.

2. Use a piece of masking tape to label your bead primers with your group number or name.

3. Match your first bead primer to a paper template strand at the complementary sequence so the pop-it end points to the right. Use cellophane tape to hold the primer in place.

As in the modeling DNA synthesis activity, the strip of paper represents the DNA template, the four beads are the primer, and the cellophane tape stands for the weak hydrogen bonds that hold the two strands together.

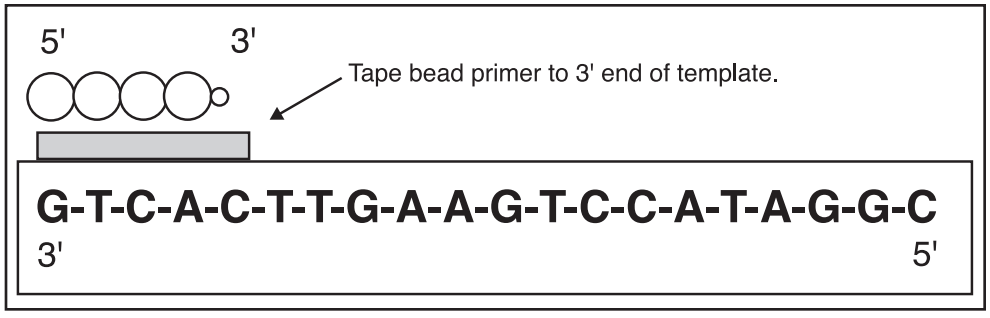


Figure 23. Primer binding to the DNA template strand

4. Now you will play the role of DNA polymerase. The beads are the nucleotides. Remember:
   - Green bead = A
   - Blue bead = C
   - Yellow bead = G
   - Red bead = T

Some of your beads have had the pop-it end cut off. These are the chain terminators. In this activity, there are two sets of each of the four reaction mixes. Your group will have one of four reaction mixes. If your group has the A reaction mix, then you will have a mixture of normal beads plus some modified green beads.

Remember that the hole in the bead represents the 5' side of the nucleotide and the pop-it end represents the 3' OH end. Beads that have their 3' end cut off can be added onto a growing chain, but the next bead cannot be added to it. This is what happens during DNA sequencing when a chain terminator is added to the growing DNA chain.

To start the chain, you need to choose the bead that is the *complement* of the nucleotide on the DNA template shown by the arrow in Figure 23, just as you did during the DNA synthesis activity. Add the bead to the pointy end of the primer. Tape the bead to the paper template to hold it in place. Now add the next bead, making sure that it is complementary to the next position on the DNA template, and popping it onto the first bead. Notice that you may randomly select a bead that lacks its pointy end, a chain terminator. Once you have added this bead you will not be able to add more beads to this chain. Continue adding beads until you add a chain terminator or reach the end of the template. Then repeat with the second primer and template.

Part B: Gel Electrophoresis

As you may recall, scientists use gels as sieves to separate DNA fragments. Since DNA has a negative charge, the samples are loaded at the negative electrode and, when the current is applied, they migrate toward the positive electrode. You and your classmates are going to arrange your DNA fragments on a "gel" in the pattern that you would expect to see if you had separated real DNA fragments on a sequencing gel.

1. Separate the DNA template from the new DNA strand by cutting the tape that holds them together.

   With real DNA we would heat the DNA to 95°C to break the weak hydrogen bonds that hold the two strands together.

2. Place the new DNA strands on the "gel" in the appropriate lane and position (see Figure 24). Remember that each of the four reaction mixes is represented. If you have the "A" reaction mix, put it in the A lane, if you have the C reaction mix, put it in the C lane, and so on. Determine how many nucleotides are in each DNA strand by counting how many beads

there are (including the four nucleotides in the primer). Then line up each strand with the same number at the side of the gel. If you have more than one fragment of a certain size, clump them together in a tight "band."

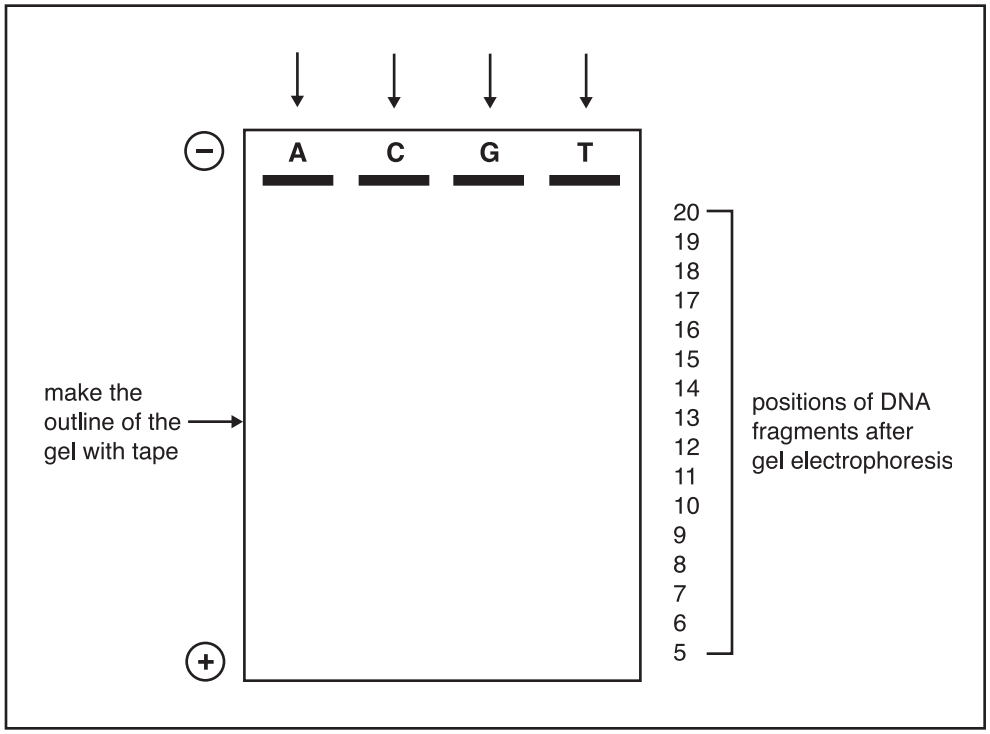


Figure 24. Place your DNA strands in the correct lane and line them up at the position that corresponds to their length.

Part C: Reading the DNA Sequence

Notice that the fragments are spread out with the smallest ones near the bottom and the largest ones at the top. Each fragment should have a primer on one end and a dideoxy bead at the other (except some of the 20-mers at the top of the gel). There should be at least one fragment at each position between 5 and 20, if enough *primers* have been extended. To read the gel:

1. Identify the smallest "band" (it should be at position 5). Which lane is it in: A, C, G, or T? If the lane is labeled "A," then this band corresponds to an A in the chain you are sequencing.

2. Now move up the band pattern to the next lowest band (it can be in any of the four lanes). What lane is it in? This is the next nucleotide in the sequence.

3. Continue reading the sequence until the top of the gel is reached. Why is this kind of gel often called a "sequencing ladder"?

Clean Up

Reclaim your group's DNA fragments from the tabletop gel by looking for your group number taped to the primer. Take the beads apart. Please leave the four beads of the primers attached to each other. Return the beads and primers to your teacher.

Opportunity for Assessment

Your school newspaper has asked you to write a short article describing DNA sequencing. The goal is to help other students understand the technique. In your article, be sure to include the key features of DNA synthesis, the use of chain terminators, and gel electrophoresis.

G-T-C-A-C-T-T-G-A-A-G-T-C-C-A-T-A-G-G-C

3' 5'

G-T-C-A-C-T-T-G-A-A-G-T-C-C-A-T-A-G-G-C

3' 5'

Template strands: cut along dotted lines.

Color Code
A = green bead
C = blue bead
G = yellow bead
T = red bead

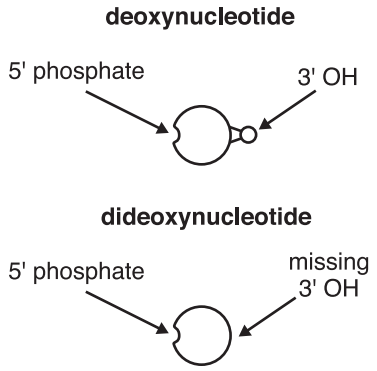


Figure 25. DNA templates

Explore/Explain

DNA Sequencing—Experimental Procedure

Day 1: DNA Sequencing by Cycle Sequencing

Background

Cycle sequencing is very similar to a technique called the **polymerase chain reaction** (**PCR**). In PCR, a piece of double-stranded DNA is amplified by completing multiple rounds of DNA synthesis. DNA primers are selected that are complementary to the 3′ end of each DNA strand. A special DNA polymerase that functions at high temperatures and tolerates near boiling temperatures is used. All necessary ingredients—DNA template, DNA primers, DNA polymerase, and nucleotides—are mixed together. The mixture is heated to 95ºC to break the hydrogen bonds in the DNA. Then it is cooled to about 45–65ºC (depending on the primers being used), a temperature that allows hydrogen bonds to re-form, and the primers bind to the DNA strands. The mix is then heated to 72ºC, the optimal temperature for the DNA polymerase, and the two DNA strands are copied. Then the sample is switched to 95ºC again to denature the DNA; back to 45–65ºC to allow primers to bind, and then to 72ºC for DNA synthesis. This cycle is repeated multiple times, and every time the DNA strands made in previous cycles are used for DNA templates. This results in exponential amplification of the DNA (2 strands become 4, then 8, then 16, 32, and so on).

During cycle sequencing, temperature is also used to break and re-form hydrogen bonds in the DNA, and the same heat-stable DNA polymerase is used. However, only one primer is used, so only one DNA strand is copied during each cycle (1 strand becomes 2, then 3, 4, 5, etc.)

Part A: Set Up Reaction Mixtures

All of the components needed for DNA synthesis are mixed in a tube: DNA template (the DNA molecule being sequenced), a DNA primer, the four deoxynucleotides (A, C, G and T), one dideoxynucleotide (chain terminator), buffer solution, and the heat-stable DNA polymerase. For each DNA template being sequenced, we need to prepare four reaction mixtures, each containing one of the four chain terminators (dideoxyA, dideoxyC, dideoxyG, or dideoxyT) plus all of the four deoxynucleotides.

Part B: Perform DNA Synthesis

DNA synthesis is carried out by incubating the four reaction mixtures at three different temperatures. Look at Part B in Figure 26. The three panels at the bottom show what is happening inside the tubes at each temperature. First, the reaction mixtures are heated to 95°C to separate the DNA strands by breaking apart the hydrogen bonds that hold them together. This is called the denaturation step. Next, the samples are cooled to 45–50°C to allow the DNA primers to bind to the DNA template strand. This is called the annealing step. Finally, the reactions are heated to 70–72°C, which is the optimum temperature for the heat-stable DNA polymerase to function. This is called the synthesis or extension step.

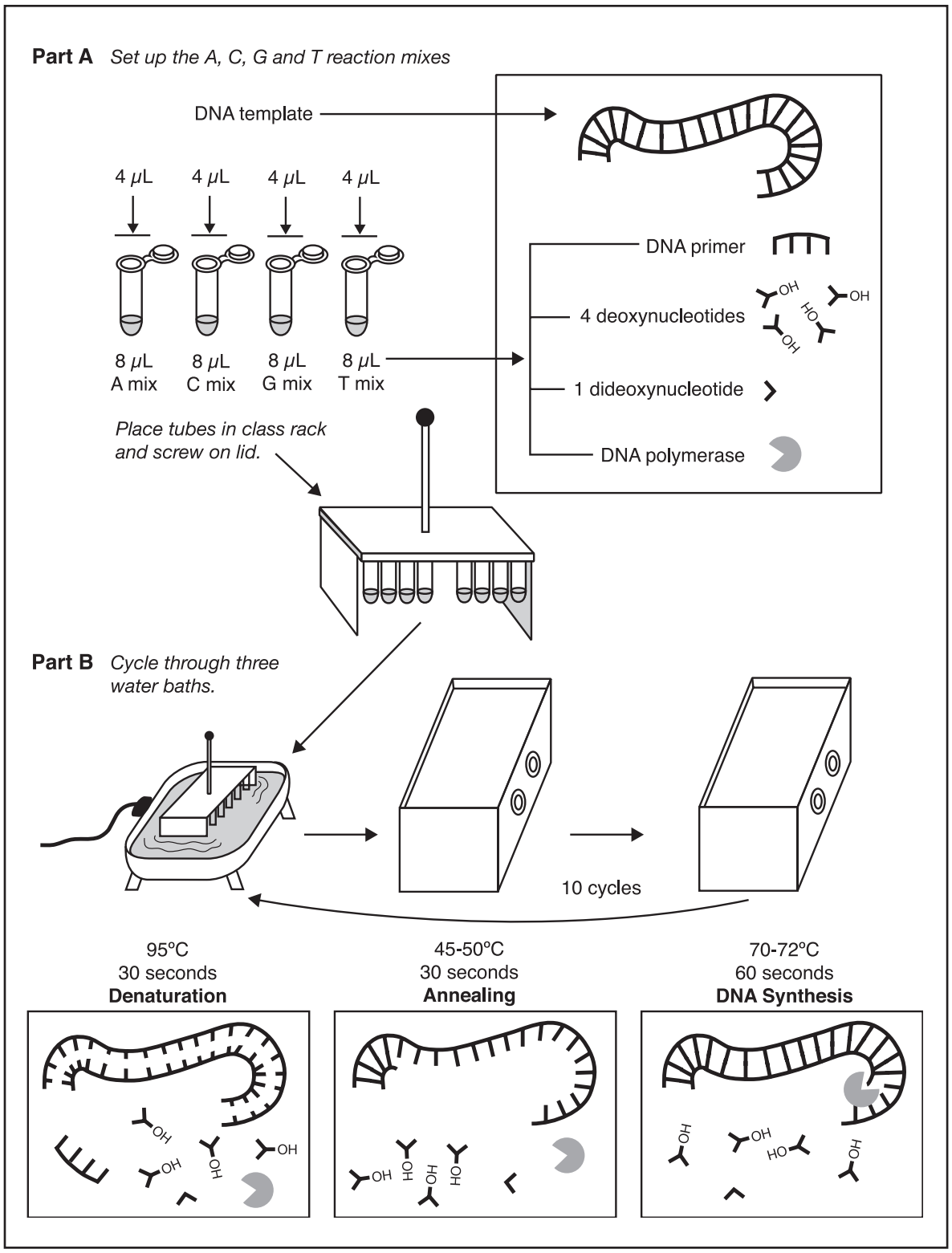


Figure 26. DNA sequencing by thermal cycling

The reaction mixtures are cycled through these three temperatures repeatedly to allow for several rounds of DNA synthesis to occur. The DNA templates are reused during each cycle because the 95°C incubation separates the template strands from the newly synthesized DNA strands, making them available to be copied again. Typically, 10–20 cycles of heating and cooling yields enough DNA fragments to determine the sequence.

Part C: Stop DNA Synthesis
At the end of the last cycle, the four reaction mixtures are cooled to 4°C to inactivate the DNA polymerase.

Procedure

Check off each step as you complete it.

Wear lab gloves for this day's activities.

Pre-experiment Setup.

____ Write in your lab notebook the name of the DNA template written on your DNA template tube; the type of primer (F or R) depending on if you are using the forward or reverse primer; and your group number. You will need this information later. The forward primer binds near the 3′ end of one strand of the DNA template, and is used to sequence that strand. The reverse primer binds near the 3′ end of the other strand and is used to obtain its sequence. Sequencing with a forward primer in one reaction and a reverse primer in another helps to ensure that the entire DNA template will be sequenced.

DNA Template Name: _____
Primer (F or R): _____
Group number: _____

Note, the forward primer is used to sequence one DNA strand, and the reverse primer is used to sequence its complementary strand.

____ Review the operation of the P-20 micropipette and practice using colored water if so instructed. Review when to change tips. Accurate micropipetting is essential to achieving good sequence data!

Part A: Set up reaction mixtures.
In this first part you will mix together all the ingredients needed for DNA sequencing.

___ 1. Your teacher will give you four reaction tubes containing 8 µL of the A, C, G, or T reaction mix. Each reaction mix contains a buffered solution plus the following: a DNA primer, the four deoxynucleotides, DNA polymerase, and one of the four dideoxynucleotides. Label each tube with your group number. The tubes are color-coded:
- The clear tube contains A reaction mix.
- The blue tube contains C reaction mix.
- The yellow tube contains G reaction mix.

- The pink tube contains T reaction mix.

___ 2. Use your micropipette to add 4 µL of the DNA template to each tube. You will end up with the four complete reaction mixtures (see Figure 27).

	A tube (clear)	C tube (blue)	G tube (yellow)	T tube (pink)
Reaction Mix (A, C, G, or T)	8 µL	8 µL	8 µL	8 µL
DNA template	4 µL	4 µL	4 µL	4 µL
Total	12 µL	12 µL	12 µL	12 µL

Figure 27. Setting up reaction mixtures

To make sample preparation easy, all you need to do is add your DNA template to each reaction mix tube. See Part A in Figure 26.

___ 3. Close the lids to your four tubes and spin in the microcentrifuge for a few seconds to ensure that all reaction ingredients are pooled at the bottom of the tube.

Part B: DNA synthesis by manual thermal cycling

In this part, you are going to carry out thermal cycling by manually moving your reaction mixtures between three water baths set at 95°C (to denature the DNA), 45–50°C (to allow the primers to bind), and 70–72°C (to carry out DNA synthesis).

___ 1. Check that the water level in each bath is deep enough that it covers the sides of the reaction tubes when they are sitting in the sample rack but does not go over the tops of the tubes.

___ 2. Make sure your four reaction tubes are clearly labeled and place them in the metal thermal cycling rack. Once everyone's samples have been loaded into the rack, screw on the metal lid.

___ 3. It takes three people to carry out the thermal cycling. Each group member should take a turn performing at least one cycle.
- *Person 1* is the cycler and moves the rack from one water bath to the next.
- *Person 2* is the timekeeper. Using the chart in Figure 28, the timekeeper tells the cycler when to move the racks and checks off each step on the chart as it is completed.
- *Person 3* watches the temperature and level of each water bath, adjusting the thermostat slightly and adding small amounts of water as needed. It is very important that the water baths be at the correct temperatures before starting and remain there throughout the cycling procedure.

Cycle #	95°C	Check	45–50°C	Check	70–72°C	Check
0	1 min		--		--	
1	30 sec		30 sec		1 min	
2	30 sec		30 sec		1 min	
3	30 sec		30 sec		1 min	
4	30 sec		30 sec		1 min	
5	30 sec		30 sec		1 min	
6	30 sec		30 sec		1 min	
7	30 sec		30 sec		1 min	
8	30 sec		30 sec		1 min	
9	30 sec		30 sec		1 min	
10	30 sec		30 sec		1 min	

Figure 28. Temperature cycling checklist

Part C: Stop DNA synthesis

In this part, you will stop the action of the DNA polymerase by lowering the temperature and adding loading mix, which contains EDTA, an inhibitor of DNA polymerase. It also contains formamide, which serves two purposes: it is a strong denaturant and helps to denature double-stranded DNA, and it is denser than water, so it helps the sample settle to the bottom of the gel well. Two colored dyes in the mix are used to track how far the gel has run during electrophoresis.

___ 1. Pick up your group's reaction tubes from the class rack.

___ 2. Spin the reaction tubes briefly in the microcentrifuge.

___ 3. Use your micropipettor to add 10 µL loading mix to each tube and mix.

___ 4. Rewrite the labels on any tubes if necessary.

___ 5. Place all your reaction tubes in the tube rack for your class. Place the rack in a –20°C freezer until you are ready to load the sequencing gel.

Dispose of used micropipette tips and gloves in the waste bag provided by your teacher.

Clean Up

Questions for Understanding

Answer the following questions in your lab notebook.

1. Draw a diagram that shows what is happening in the reaction tubes at each of the three temperatures used during thermal cycling.

   a. 95°C

   b. 45–50°C

   c. 70–72°C

2. Describe the set of DNA fragments synthesized in each of the four reaction mixes.

Explore/Explain

DNA Sequencing—Experimental Procedure

Day 2: Gel Electrophoresis

Background

Gel electrophoresis is a method of separating charged molecules, especially large biomolecules like proteins, RNA, and DNA. The process is carried out in a slab of gel that contains pores. The pores allow the gel to function as a molecular sieve. Electricity is used to move the samples through the gel. The gel is placed in a salt solution (**running buffer**) that conducts electricity. When the electric current is applied, samples move through the gel in a direction that depends on the sample's charge. The rate at which samples migrate through the gel is determined by their charge, size, and shape. The distance that samples migrate is affected by the conditions used for electrophoresis, such as the voltage applied, the ionic strength and pH of the running buffer, the temperature, the type of gel material used, and the total electrophoresis time.

Two types of gel materials are routinely used for studying DNA molecules: **agarose** (derived from seaweed) and **polyacrylamide** (a material similar to that used in making soft contact lenses). Agarose gels have the advantage of being easy to prepare. They are useful for analyzing DNA fragments that range from 300 to hundreds of thousands of base pairs in length. A disadvantage of agarose gels is that they cannot resolve DNA fragments that differ from each by a single base pair. Since this level of resolution is required for DNA sequencing, polyacrylamide is used to prepare gels intended for sequencing.

As the name implies, polyacrylamide is made up of long, linear chains of subunits called acrylamide. A catalyst is added to a liquid acrylamide solution that causes chemical bridges to form between the chains, forming an interwoven polyacrylamide mesh. While still liquid, the gel solution is poured between two glass plates held apart by plastic spacers along the sides. The spacers are usually between 0.2 and 1 mm in thickness, which produces a thin, even gel. A special spacer with flat teeth like a comb is inserted into the top of the gel. After a few minutes, the polyacrylamide solution forms a gel and the comb is removed, leaving wells along the top edge of the gel where the DNA samples are loaded. For our experiment, the polyacrylamide sequencing gel has been already been poured.

Sequencing gels are typically longer and thinner than agarose gels. This helps produce better separation. The apparatus used to run the gel is shown in Figure 29 (front and side views). The gel is between two glass plates that are held apart by plastic spacers along the sides. The back plate is about 2 cm shorter than the front plate, and it is attached to the upper buffer chamber. This chamber extends over the entire area of the back plate, so that when filled with buffer the top edge of the gel is covered. The upper buffer acts as a heat sink that maintains uniform heat distribution over the entire gel area during electrophoresis.

The plate assembly is placed in the base, which also contains buffer, and is held in place by a stabilizer bar. Both buffer chambers are protected by safety covers with attached electric cables. When the safety covers are in place, the cables contact the electrodes on the upper and lower buffer chambers. The other end of each electric cable is plugged into the power supply.

The phosphate groups of DNA give it a negative charge. Therefore, the DNA samples are loaded into their sample wells near the negative electrode. When the electric current is applied, the DNA fragments move out of the wells and down the gel toward the positive electrode.

It is important that the DNA fragments remain denatured during electrophoresis. Otherwise, the DNA can form areas of double helix that change its rate of migration through the gel. The DNA samples are kept denatured during electrophoresis by using 7 molar urea in the gel solution, including a denaturing agent called formamide in the loading mix, heating the samples prior to loading, and running the gel at high voltage to generate heat in the gel during electrophoresis (see Figure 30).

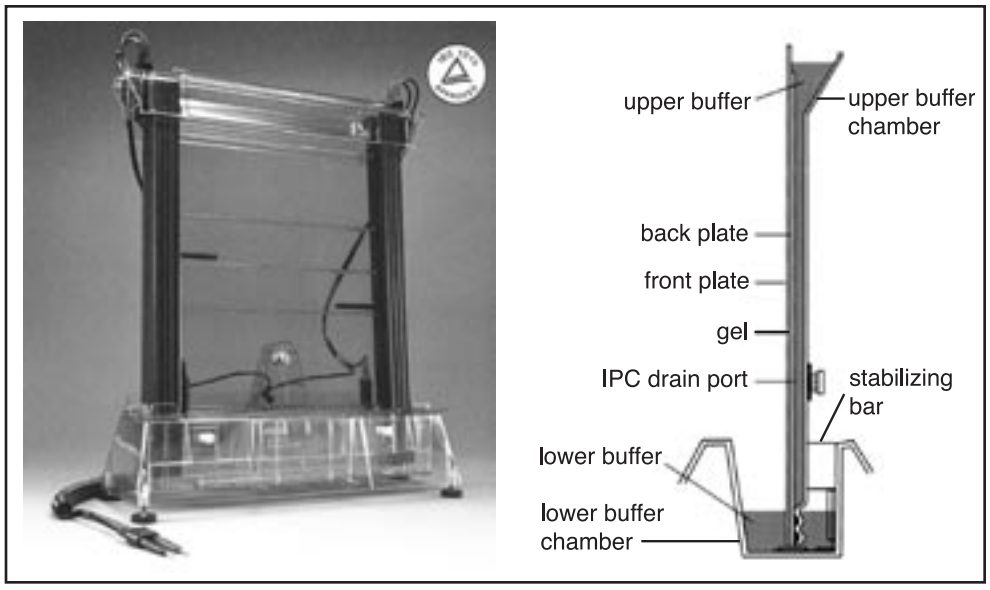


Figure 29. Front and side view of sequencing gel apparatus

After electrophoresis has finished, the apparatus is disassembled. The glass plates are separated, exposing the gel. Since it is difficult to visualize the DNA while it is embedded in the gel, a nylon membrane that resembles very delicate paper is laid on top of the gel. The membrane binds DNA very well, so the DNA fragments in the gel are adsorbed out of the gel and onto its surface.

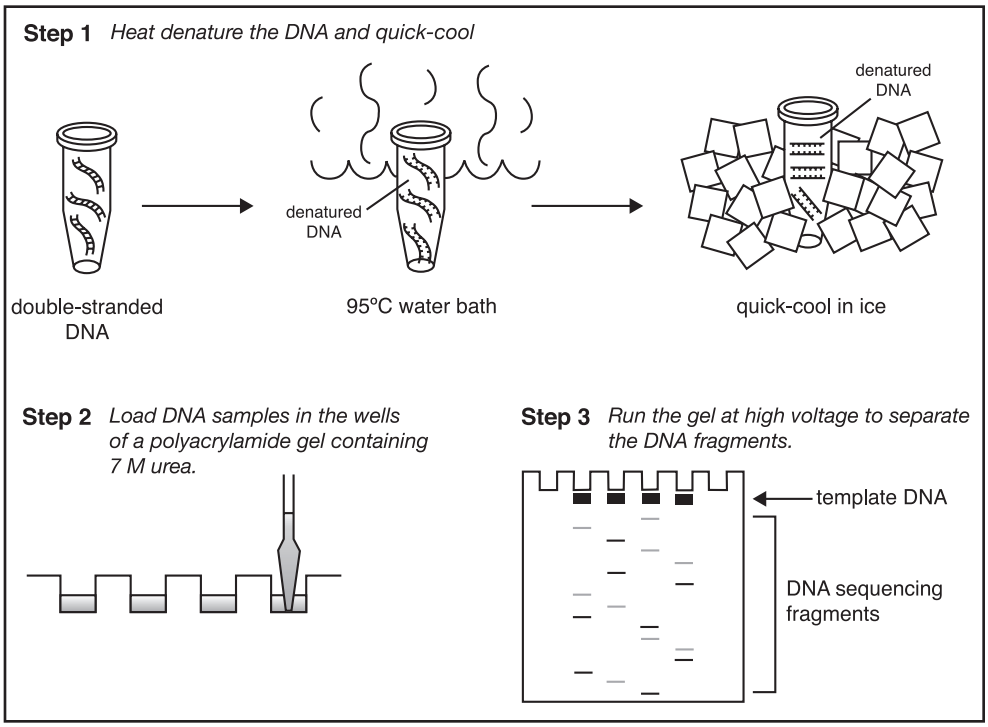


Figure 30. Denaturing gel electrophoresis

Wear lab gloves for this day's activities. Sequencing gels are run at 2,000 to 2,200 volts and so electrophoresis is potentially very dangerous. For this reason, gel loading will be supervised by your teacher or a visiting scientist. The supervisor also will be responsible for turning the power supply off and on at appropriate times.

Procedure

Check off each step as you complete it.

Pre-Experiment Set Up

The sequencing gel has been set up in the gel apparatus and prerun at 2,000 to 2,200 volts by your classroom teacher or the visiting scientist about one hour before class to bring the gel temperature to near 45°C. There is running buffer in the upper chamber (the space behind the glass plate) and in the lower chamber (the tray the gel sits in).

Part A: Preparing and Loading the Samples

All lab groups in the class will take turns loading their samples onto the gel, following Steps 1–5. Heat and cool your samples just before it is your turn to load. While group 1 is loading, group 2 should be heating and cooling its samples, and so on, until all groups have loaded their samples.

___ 1. Place your four reaction tubes in the metal thermal cycling rack and set it in the 95°C water bath for 3 minutes just before loading. Following the hot water incubation, place the samples in an ice-water bath to quick-cool the DNA. Leave the samples on ice until you load the gel.

The samples are heated to separate the new DNA strands from the DNA template strands. Then they are quickly cooled on ice to prevent the DNA strands from coming back together.

___ 2. Turn off the power supply and remove the upper safety cover.

___ 3. Prepare your lanes (sample wells) for loading. The first group to load should start in the second or third lane from the left side of the gel. Mark the positions of your four lanes on the outer glass plate with a felt-tip pen (label your lanes "A," "C," "G," and "T" and write your group number below). Use the transfer pipet to rinse your lanes just before loading your samples.

Rinsing flushes out undissolved urea, which would otherwise prevent the samples from settling to the bottom of the well.

___ 4. Using the P-20 micropipette and the flat-tipped microcapillary tips, load 5 µL of each sample in alphabetical order (A, C, G, T). Leave one lane between your samples and those of the previous group. To load, slowly suck your sample into the pipet tip, then lower the tip into the well by dragging the flat side of the tip down the back side of the front plate of the gel. Once the tip is in the well (just *above* the bottom of the well), slowly press down on the plunger of the pipet so that the sample forms a blue layer on the bottom of the well. Stop before you have loaded the entire sample (that is, leave about 1 µL in the tip) and remove the tip from the well. *Only at this time* should you remove your thumb from the plunger.

___ 5. When all four samples are loaded, the supervisor will replace the upper safety cover and turn on the power supply until both blue dyes have entered the gel (1–2 minutes).

Running your samples into the gel will ensure that they don't diffuse out of the sample wells while other groups are loading their samples.

Part B: Running the Gel

During electrophoresis, the DNA fragments are separated down the length of the gel according to size, with the smallest fragments migrating the fastest.

___ 1. The supervisor will turn on the power supply and adjust it to 2,200 volts. You can verify that current is flowing by looking for bubbles forming at the negative electrode.

During the gel run, your teacher will make sure that the temperature indicator strip on the outer gel plate reads around 45 to 50°C. As electrophoresis proceeds, the temperature will continue to rise and the voltage may need to be adjusted down to approximately 1,800 volts. The supervisor will also make sure that the level of buffer in the upper chamber is about 1 cm from the top edge of the gel plate. Occasionally during the run, your teacher will mark the left and right edges of the dyes corresponding to each set of samples. This will help you position the nylon membrane later in the experiment.

___ 2. Run the gel until the xylene cyanol dye (the slower-moving turquoise-colored dye) of the last sample loaded has migrated about three quarters of the way to the bottom of the gel (about to the drain port on the back plate). The gel run takes approximately 3 1/2 hours.

Part C: Disassembling the Gel

Your teacher will do the following steps on your behalf.

___ 1. Turn off the power supply, disconnect both safety covers, pull out the stabilizer bar, and remove the gel assembly.

___ 2. Carefully pour the upper buffer out of the upper buffer chamber into the lower buffer chamber or a nearby sink.

___ 3. Lay the gel assembly on the bench top with the outer (front) plate facing up. Using a ruler and pen, draw vertical lines along the pen marks that bracket the lanes for each lab group.

These lines show where each group's DNA samples are in the gel.

___ 4. Now, turn the gel assembly over so that the outer (front) plate faces down. Remove the side clamps from the gel. Remove the inner (back) glass plate with the attached buffer chamber by prying gently at the top of the outer plate using a metal spatula. The gel should stick only to the outer (front) plate, as the inner plate was treated to make it slippery. Remove the spacers along the sides of the gel.

Part D: Blotting Procedure

In this part, you will transfer the DNA in the gel to a nylon membrane, so you can visualize the DNA. Your teacher may complete this process for you.

___ 1. Each lab group is given a nylon membrane that is cut to fit its portion of the gel. Wearing gloves, handle the membrane only by its edges. Use a pencil to clearly label the edge of the membrane near one end with the data file name your teacher provides for you. See Figure 31 for a key to

2A6E1P1S1_F1SC_0504

E1 = exon 1 (exon = coding region and intron = noncoding part of a gene)
P1 = primer pair set 1
S1 = subject 1 (the individual whose DNA is being sequenced)
F = forward reaction (use R if you did the reverse reaction)
1 = lab group number (assigned by your teacher)
SC = school code
0504 = date (month and year)

Figure 31. The data file name. The DNA templates you receive are labeled with the information up to the first underscore. You will need to provide the additional information.

understanding what the file name means. Remember, you recorded your DNA template name, primer used, and group number on the day you began the DNA sequencing.

___ 2. Dip the membrane into a tray containing TBE (running) buffer. Hold the wet membrane at both ends with the pencil side down and toward the top of the gel, directly above your lanes on the gel (and not a neighboring lab group's).

Before lowering your membrane onto the gel, check its position. You cannot reposition it once you have started to lay it down! The bottom edge of the membrane should be about 1 cm (1/2 inch) above the bottom of the gel. Note that the membrane will not reach up to the top of the gel. Once the wet membrane appears to be properly positioned, lower it onto the gel surface, so that the middle touches first. Then slowly lower both hands so that the membrane flows smoothly onto the gel without trapping air bubbles (see Figure 32).

If the membrane is wider than the portion of the gel you want to cover then lay it down so that it overlaps the membrane covering the adjacent lanes, But not the neighboring lanes that are not yet covered in membrane.

___ 3. Once all of the membranes have been laid in place, cover them with two sheets of Whatman paper and a Plexiglas plate. Apply a weight of approximately 4 kg (8 textbooks).

___ 4. Allow 40 minutes for the DNA to be absorbed out of the gel and onto the nylon membrane.

___ 5. Remove the weights, Plexiglas plate, and the Whatman paper. Remove the membrane from the gel. Be careful not to pick up any of the gel. If some of the gel sticks to the membrane, squirt some running buffer on the membrane to loosen it and gently lift it off using a metal spatula or forceps.

___ 6. Place the membrane, DNA side up (pencil side up), on a fresh piece of Whatman paper that has been dipped in running buffer. Place the membrane-paper into the UV crosslinker and close the door. (Make sure not to cover the disk at the back of the crosslinker). Turn the power on, press "optimal crosslink," and then press "start." When the timer rings, remove the membrane and turn the power off.

The exposure to ultraviolet light permanently fixes the DNA onto the nylon membrane by inducing the formation of covalent bonds between the DNA and the membrane.

___ 7. Store the membranes inside a folded piece of paper, out of the light, until you are ready to visualize. (They will keep indefinitely.)

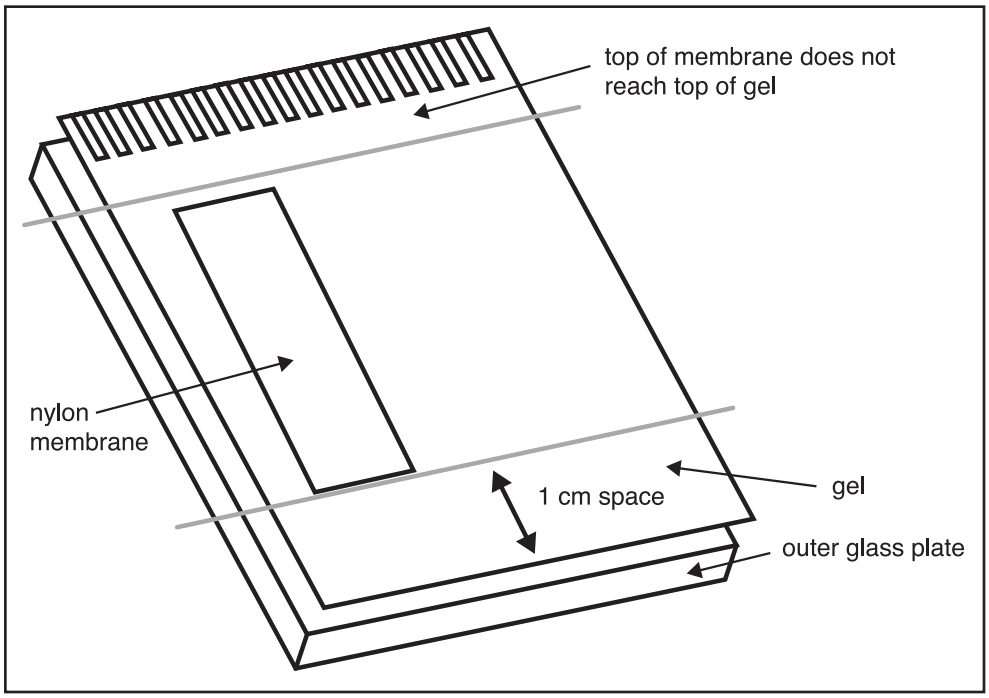


Figure 32. Positioning the nylon membrane over the sequencing gel

Dispose of used micropipette tips and gloves in the waste bag provided by your teacher.

Clean Up

Questions for Understanding

Answer the following questions in your lab notebook.

1. Name two gel substances that are commonly used for electrophoresis. What are the advantages of each?

2. During electrophoresis of the sequencing gel, the DNA is denatured.

   a. What does "denatured" mean?

   b. Why does the DNA need to be denatured when it is running on a sequencing gel?

   c. How is the DNA kept in the denatured form as the gel runs?

3. Why did we transfer DNA from the polyacrylamide gel to the nylon membrane?

Explore/Explain

DNA Sequencing—Experimental Procedure

Day 3: Visualizing the DNA

Background

So far, you have synthesized DNA, separated the different sized DNA fragments on a polyacrylamide gel, and transferred the DNA to a nylon membrane. Now you are going to wash the nylon membrane in a series of different solutions that will allow you to see the newly synthesized DNA as a series of dark purple bands. The steps in this procedure are outlined in Figure 33 and explained in detail below.

Procedure

1. The primer you used during day 1 of the experiment has a chemical tag on its 5′ end called **biotin** (a vitamin molecule). This visualization technique relies on the use of a protein that can specifically bind to biotin. To increase this specificity, the membrane is first washed in blocking solution to neutralize nonspecific charges on its surface.

2. The membrane is then washed in a solution containing a protein called **streptavidin** (a natural antibiotic made by the bacterium *Streptomyces avidinii*). Streptavidin specifically binds to the biotin molecules on the 5′ end of each DNA fragment. Each molecule of streptavidin has four sites for binding biotin. Usually, one streptavidin binds to each biotin molecule, leaving three empty biotin binding sites on each streptavidin molecule. These empty sites will be important in a later step. Two washes in a buffered solution remove the unbound streptavidin.

3. The membrane is next washed in a solution containing an enzyme called **alkaline phosphatase** (**AP**). The AP is covalently linked to a biotin label. These biotin-labeled AP molecules bind to the remaining sites on each streptavidin molecule already attached to the DNA. Following this step, the membrane is washed three times to remove unbound biotin-labeled AP.

4. The membrane is finally soaked in a substrate solution that is broken down by the AP enzyme. When AP removes a phosphate group from this substrate, a purple precipitate is formed. This results in the formation of a purple color at the position of each biotin-labeled DNA band.

Why do we go through all this trouble when it would be so much easier to simply label the DNA primer with a purple label in the first place? The answer is that no purple dye is intense enough to be seen at the low concentration of DNA primer that we use. However, one molecule of the alkaline phosphatase enzyme can react with thousands of substrate molecules, resulting in a great enhancement of the color signal. Since we are restricted in how much DNA we can load

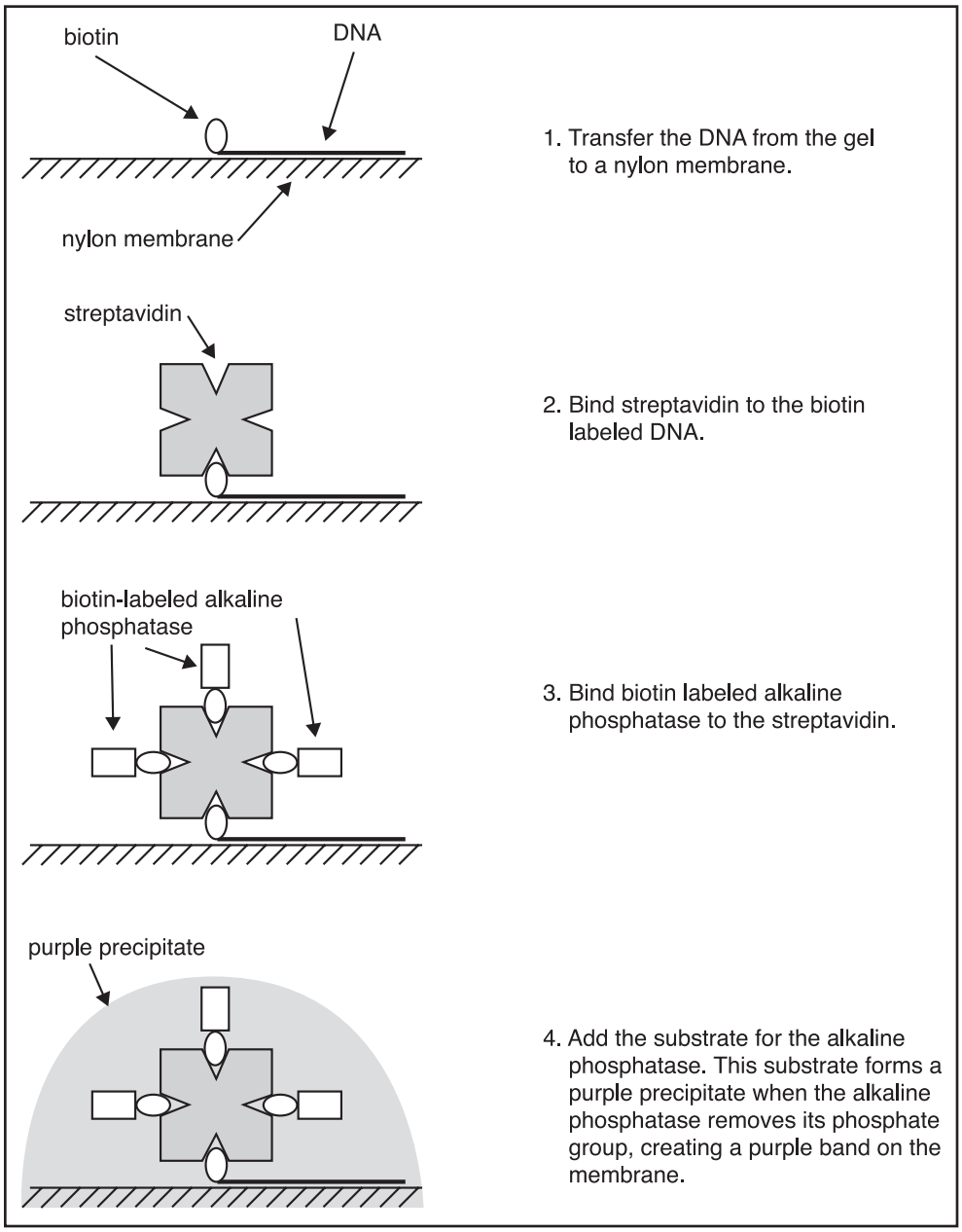


Figure 33. Steps in visualizing the DNA

Wear gloves for the entire blotting procedure, as some of the solutions may be harmful.

Wear gloves when handling the membrane and only touch the edges.

on a gel, we use this elaborate technique to create an amplified color signal.

Procedure

Check off each step as you complete it.

Pre-Experiment Set Up

In this lab period, you will wash your nylon membrane in a series of different solutions. Figure 34 lists each of these solutions and their compositions. Each

solution needs to be prepared just a few minutes before you use it. Assign two people in your lab group to be in charge of preparing the wash solutions and assign another two people to gently shake the membrane while it is being washed.

The First Wash

___ 1. Place the nylon membrane, DNA side up (that is, pencil mark up), in the tray. Pour 30 mL of blocking solution into the tray.

___ 2. Swirl the tray for the entire wash time listed in Figure 34, making sure that the solution covers the entire membrane.

___ 3. At the end of the wash time, pour the blocking solution into a sink, using the forceps or a metal spatula to hold the membrane in the bottom of the tray and touching the membrane only near the edge.

___ 4. Repeat steps 1–3 for each of the washes outlined in Figure 34, except the final wash in color substrate buffer. Check off each row in the table after the wash is complete.

Helpful Hints:

1. *Do not let the membrane dry out between washes.*
2. *It will not harm the experiment if you wash a few minutes longer than the time indicated.*
3. *Rinse the graduated cylinders*—not your membrane—*between uses with distilled water.*

___ 5. Once the membrane has been put in the color solution, cover the tray with saran wrap, then foil, and put in a dark, *level* drawer or cupboard. It is very important that the membrane be placed in a level position to ensure that the color solution covers the membrane evenly. Allow the color to develop for 12 to 16 hours. Do not leave the membrane in this solution beyond 16 hours, or the background will become too dark.

___ 6. When the DNA bands are visible, pour off the color solution and pour on 50 mL stop solution. Rinse for a few minutes, pour off, and then rinse twice with 100 mL distilled water.

Stop solution inhibits the alkaline phosphatase activity and prevents further color development.

___ 7. Place the membrane on a paper towel to dry. When dry, place the membrane in a plastic page protector. Store the membrane out of the light to minimize fading. Use a small piece of cellophane tape along one edge of the membrane to hold it in place on the black paper inside the protector.

___ 8. Additional copies of the dried membrane can be made on a photocopier. Experiment with the intensity setting on the photocopier to get the best image.

Figure 34. Visualization of biotin-labeled DNA

√	Solution	Composition	Wash Time
	Blocking solution *Neutralizes surface charges*	30 mL blocking solution	5 minutes with swirling
	Streptavidin solution *Binds to biotin tag on sequencing primer*	15 mL blocking solution + 15 µL streptavidin	5 minutes with swirling
	Wash solution I *Removes excess streptavidin*	150 mL wash solution I	5 minutes with swirling
	Wash solution I *Removes excess streptavidin*	150 mL wash solution I	5 minutes with swirling
	Biotin-labeled AP solution *Biotin binds to streptavidin on primer and links alkaline phosphatase (AP) to the DNA*	15 mL blocking solution + 15 µL biotin-labeled AP[1]	5 minutes with swirling
	Wash solution II *Removes excess biotin-tagged AP*	150 mL wash solution II	5 minutes with swirling
	Wash solution II *Removes excess biotin-tagged AP*	150 mL wash solution II	5 minutes with swirling
	Wash solution II *Removes excess biotin-tagged AP*	150 mL wash solution II	5 minutes with swirling
	Color solution *Do not rinse this solution off! *Contains a substrate for the AP. Substrate turns purple when it reacts with AP.*	50 mL color substrate buffer + 175 µL NBT + 175 µL X-Phosphate	12–16 hours in the dark, covered, without swirling

[1]The stock concentration of biotin-tagged AP is 0.5 mg/mL in this experiment. It is sometimes provided at 0.38 mg/mL, in which case 20 µL should be used.

Dispose of the used micropipette tips and gloves in the waste bag provided by your teacher.

Warning
Bands on the membranes will fade if exposed to light for more than an hour. When you are not using your membrane, store it in a dark drawer or closed notebook.

Clean Up

Questions for Understanding

Answer the following questions in your lab notebook.

1. What is the role of each of the following molecules in the DNA detection technique used in this experiment?

   a. Blocking solution

   b. Biotin

   c. Streptavidin

   d. Alkaline phosphatase

   e. Color solution

2. What is the advantage of using this visualization technique instead of attaching a colored dye directly to the DNA primer?

3. What result would you expect if you forgot to use the blocking solution in the first step?

4. What result would you expect if you switched the order that streptavidin and alkaline phosphatase are added?

Explain

Identifying and Analyzing Single Nucleotide Polymorphisms

Identifying Variants in the *CYP2A6* Gene

While you are in the process of obtaining your DNA sequence, it is important not to lose sight of the big picture. The main objective of our sequencing project is to identify genetic variation in the *CYP2A6* gene. Many different genetic variants (alleles) have already been identified, some of which differ by single nucleotide changes and others that contain deletions or duplications. DNA sequencing is useful for identifying single nucleotide changes. In this activity, you will discover how to use DNA sequence data to recognize a variant with a single nucleotide change.

The DNA template you sequenced was made by PCR amplifying small portions of the *CYP2A6* gene from **genomic DNA** isolated from a person's white blood cells. Since people have two copies of all their genes, our DNA templates are actually mixtures of the PCR products made from the two *CYP2A6* genes. If the person whose DNA we are sequencing were homozygous for a particular *CYP2A6* allele (that is, both copies of the *CYP2A6* gene were the same), then the DNA template made by copying his or her two genes would be the same. What would you expect if the subject were heterozygous for the *CYP2A6* gene (that is, the two copies of the *CYP2A6* gene were different)?

Procedure

1. Your teacher will provide you with figures showing sequences from two different *CYP2A6* alleles (*CYP2A6*1* and *CYP2A6*6*). In this activity, you will draw the sequencing ladder that you expect if the person is

   a. homozygous for *CYP2A6*1*,

   b. homozygous for *CYP2A6*6*, or

   c. heterozygous (*CYP2A6*1*/ *CYP2A6*6*).

2. Follow the example shown in Figure 35 for the person homozygous for *CYP2A6*1*. Then complete the analysis for the person homozygous for *CYP2A6*6* (Figure 36) and for the heterozygous person (Figure 37).

Imagine that you are sequencing the DNA of a person who has two copies of the *CYP2A6*1* gene. The DNA template is shown below, and you are using the primer 3'-AGGT-5'. The analysis has been completed for you as an example.

5'	C	A	G	C	T	C	C	G	G	C	G	C	T	T	C	T	C	C	A	3'
3'																A	G	G	T	5'

Step 1 *Draw the complementary sequence that would be made by replicating the entire DNA template. Complete the new strand to the end.*

5'	C	A	G	C	T	C	C	G	G	C	G	C	T	T	C	T	C	C	A	3'
3'	G	T	C	G	A	G	G	C	C	G	C	G	A	A	G	A	G	G	T	5'

Step 2 *Draw what the sequencing membrane would look like for the complementary sequence. (Why won't you see the primer on the ladder?)*

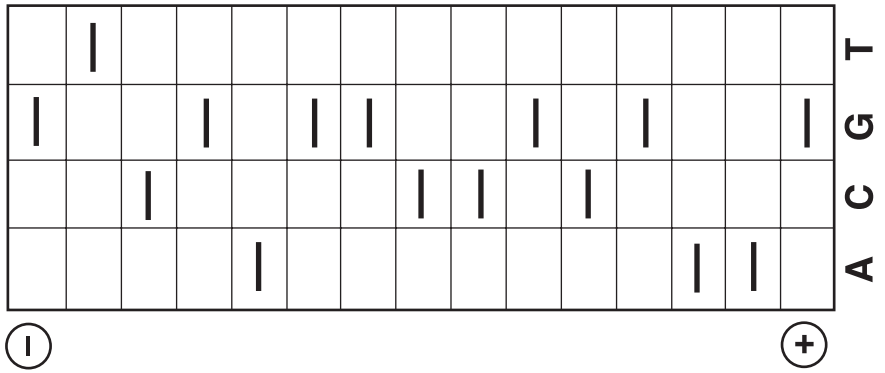


Figure 35. Person who is homozygous for *CYP2A6*1*

Draw the sequencing membrane for a person with 2 CYP2A6*6 *genes, following steps 1 and 2 shown in the previous example.*

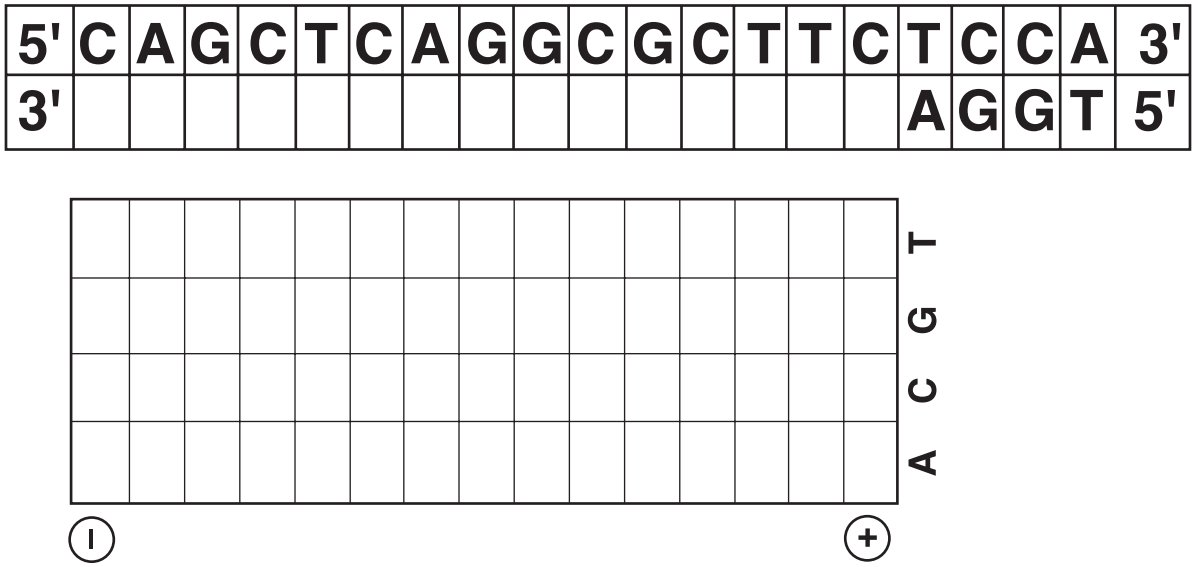


Figure 36. Person who is homozygous for *CYP2A6*6*

*Draw the sequencing membrane for a heterozygous person (*CYP2A6*1/CYP2A6*6*), following steps 1 and 2 shown in the previous example*

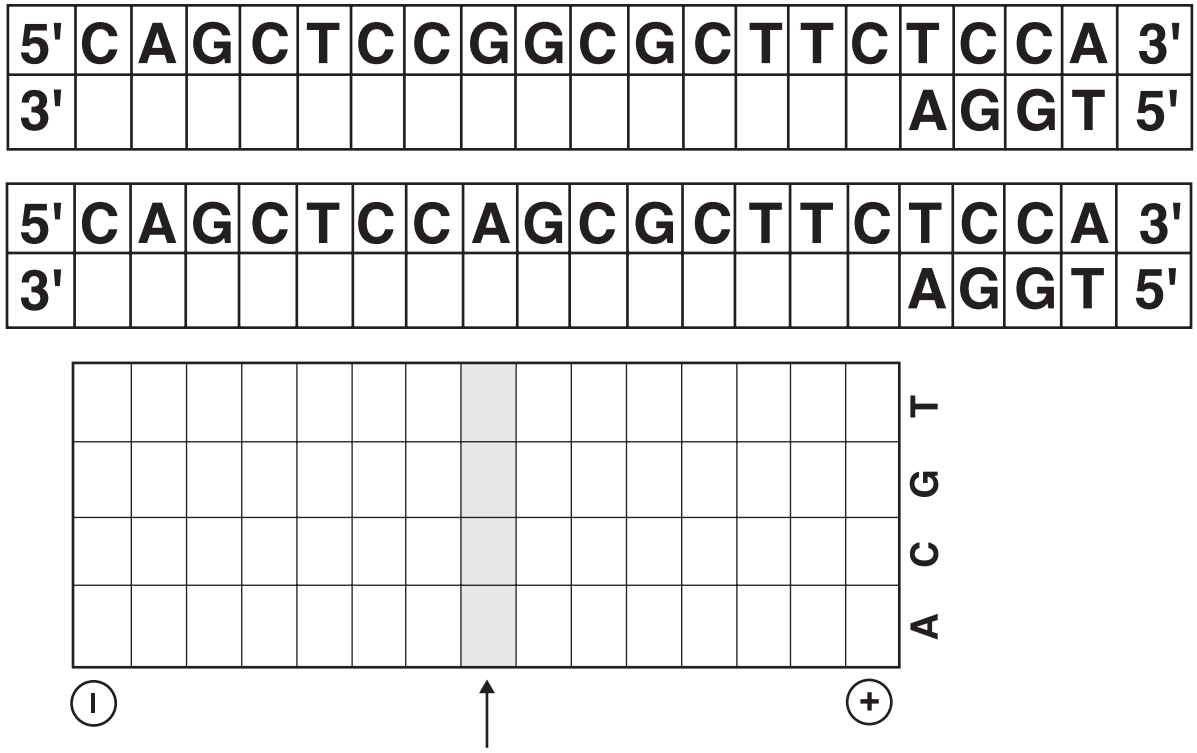


Figure 37. Person who is heterozygous for the *CYP2A6* gene (*CYP2A6*1/CYP2A6*6*)

Questions for Understanding

Answer the following questions in your lab notebook.

1. Describe what the membrane will look like if there is a SNP.

2. If you think you see a SNP on a DNA sequence using the forward primer, what do you predict you will see when you sequence the same DNA template using the reverse primer?

Explain

Identifying and Analyzing SNPs

Can a Single Nucleotide Change Affect the Enzyme (Protein) That Is Made?

Does the single nucleotide difference between the two alleles, *CYP2A6*1* and *CYP2A6*6*, result in a difference in the CYP2A6 enzymes encoded by them? This activity will answer that question.

During transcription, one strand of the DNA gene is copied to make mRNA. The mRNA is then transported to the cytoplasm, where its nucleotide sequence is translated by the ribosome to make an amino acid chain. Follow the steps below to determine the amino acid sequence of the two *CYP2A6* alleles.

Procedure

1. Your teacher will give you figures that show the *CYP2A6*1* and *CYP2A6*6* alleles in their double-stranded form (Figures 38 and 39). It is known that the bottom DNA strand is copied to make mRNA. Fill in the mRNA sequence that would be made by copying the bottom DNA strand of the *CYP2A6*1* and *CYP2A6*6* alleles.

2. Using the Genetic Code table (Figure 40), write the amino acid sequence that is encoded by each allele. Read the nucleotides in threes, starting with the first nucleotide on the left. Write the corresponding amino acids in the spaces provided.

Question for Understanding

Answer the following question in your lab notebook.

1. Do the two *CYP2A6* alleles code for the same amino acid sequence? If not, how are they different?

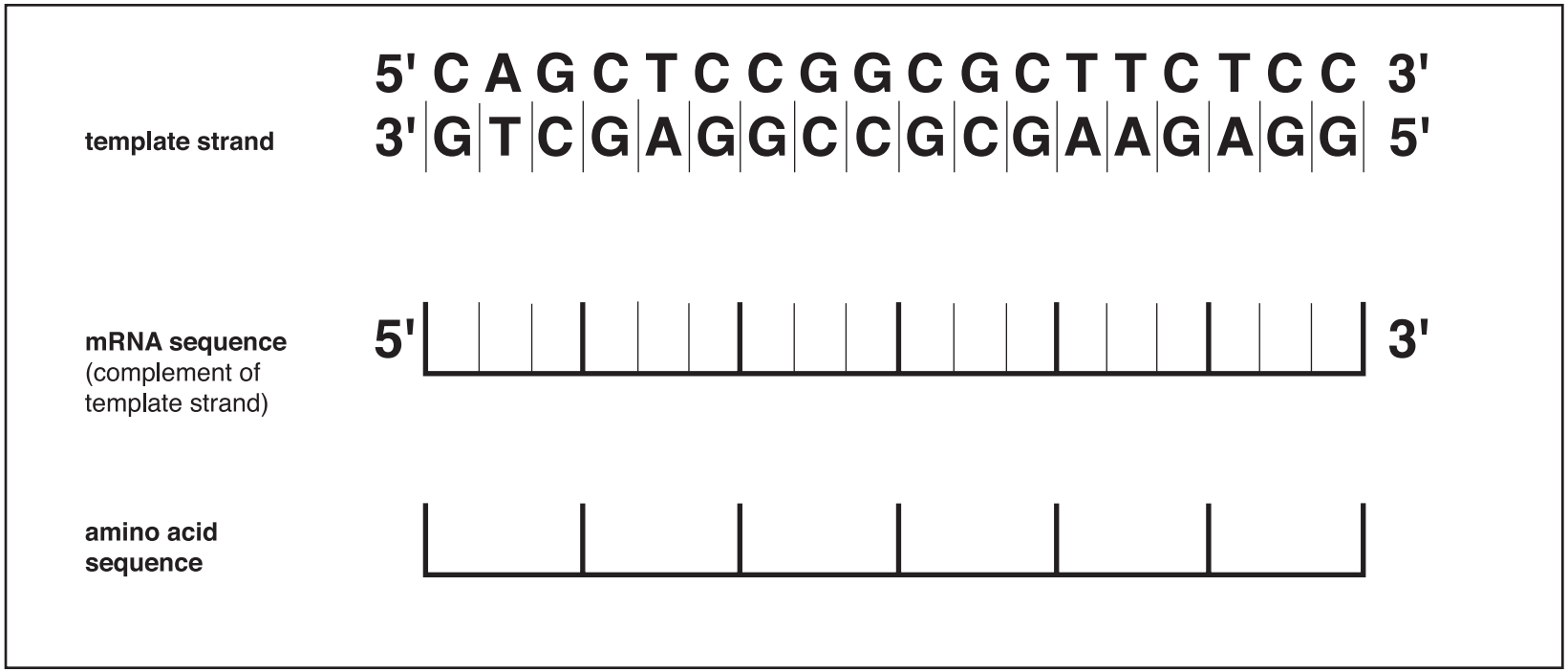


Figure 38. Transcription and translation of the *CYP2A6*1* allele

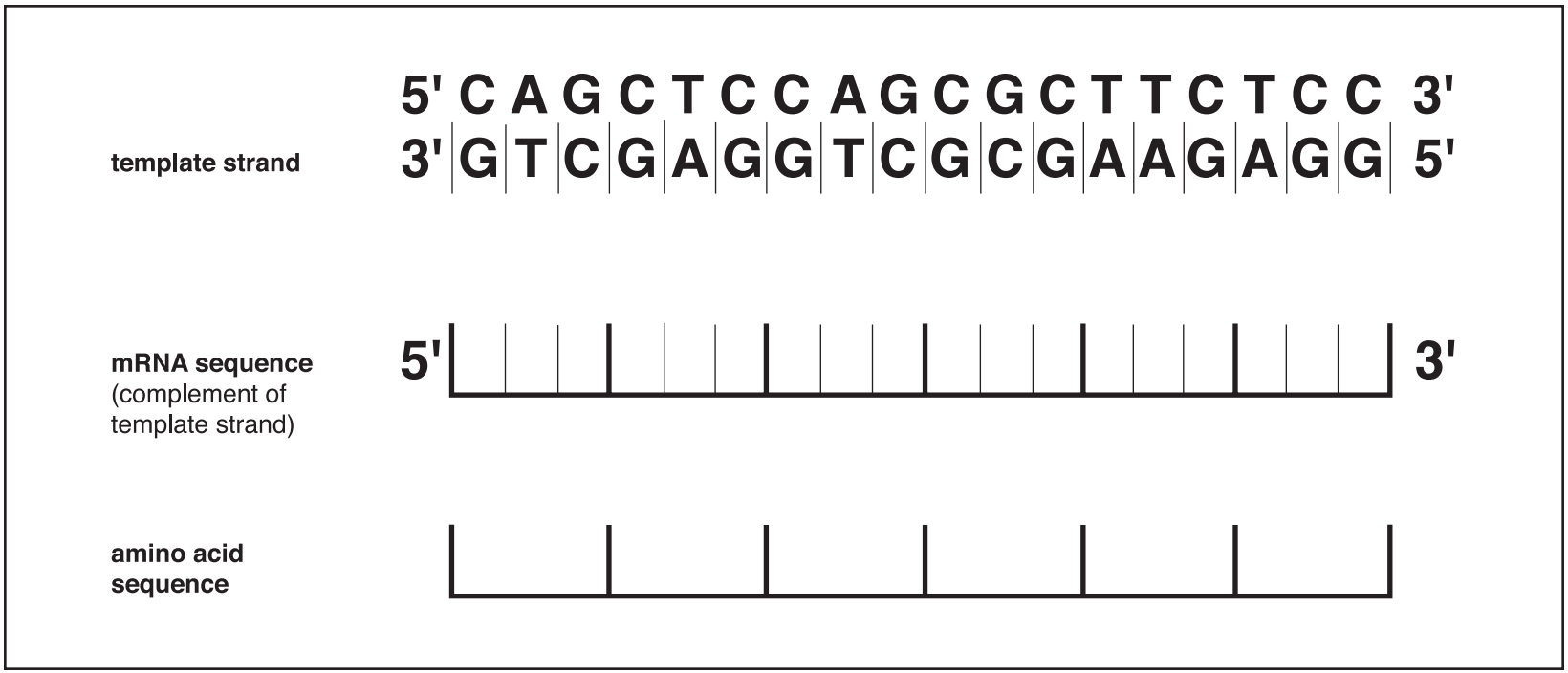


Figure 39. Transcription and translation of the *CYP2A6*6* allele

1st position (5′ end) ↓	2nd position U	2nd position C	2nd position A	2nd position G	3rd position (3′ end) ↓
U	Phe	Ser	Tyr	Cys	U
U	Phe	Ser	Tyr	Cys	C
U	Leu	Ser	STOP	STOP	A
U	Leu	Ser	STOP	Trp	G
C	Leu	Pro	His	Arg	U
C	Leu	Pro	His	Arg	C
C	Leu	Pro	Gln	Arg	A
C	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	U
A	Ile	Thr	Asn	Ser	C
A	Ile	Thr	Lys	Arg	A
A	Met	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	U
G	Val	Ala	Asp	Gly	C
G	Val	Ala	Glu	Gly	A
G	Val	Ala	Glu	Gly	G

Figure 40. Genetic Code table

Explore/Explain

DNA Sequencing—Experimental Procedure

Day 4: Data Analysis

You have now completed the technical portion of the experiment and are about to embark on the most important part, the analysis of your DNA sequencing data! As you examine the membrane, think about the many steps that went into completing the experiment. What was happening during DNA synthesis? What happened to these DNA fragments during gel electrophoresis? How did you visualize the fragments? Even if the membrane does not look the way you expect, there is a lot to learn from it.

One of our main objectives is to identify SNPs in the DNA from our eight individuals. Your template DNA was made by amplifying genomic DNA from one person, and as a result, the DNA template you used was a mixture of two copies of the *CYP2A6* gene. In most cases, the two copies are identical, so all the DNA templates in your sample will be the same. However, some individuals may have a SNP in one of their *CYP2A6* genes and you may be able to detect this from your results. Based on what you learned from the previous activity, what do you predict a SNP will look like on your sequencing ladder?

As you complete your analysis, remember that there is no "right answer in the back of the book." It's up to you and your group members to do the best job possible. *After all, this is research!*

Procedure

Check off each step as you complete it.

___ 1. Each lab group should retrieve its original, labeled nylon membrane (inside the plastic page protector), plus a photocopy of it for each lab member. The name of your data file should appear on every photocopy.

___ 2. Write the appropriate nucleotide letters at the bottom of the gel lanes. Label the membrane's page protector or a photocopy.

Note: Although the lanes are usually loaded in the order A, C, G, and T, the order can vary because of a mix-up during loading or because the gel sticks to the back plate instead of the front plate during the transfer to the nylon membrane.

___ 3. Your teacher will give you a data record sheet (Figure 42) on which to record a draft version of your DNA sequence. Work in pairs, with one person reading the DNA sequence and the other person writing it down on the data record sheet. To read the DNA sequence, start near the bottom of the membrane or photocopy. Sequencing ladders are read from the bottom up so that the DNA sequence is written in the 5′ to 3′

direction, which is the standard way DNA sequences are expressed. Mark your starting position on the photocopy or the membrane's page protector, using a washable felt-tip pen to write on the plastic.

___ 4. Every 10 nucleotides, number the nucleotide position on the membrane's page protector or photocopy so that you can easily find a position at a later time.

___ 5. If you cannot read the nucleotide at a certain position, either because there is no band visible or there are bands of equal intensity in all four lanes, then record this position as N (for nucleotide). If two bands appear at the same position but one band is darker, record the darker nucleotide. If you see *two bands of equal intensity at the same position*, use the alternate letter code given in the table below.

Nucleotides observed at same position	Letter used
A and G	R
C and T	Y
G and T	K
A and C	M
G and C	S
A and T	W

Figure 41. Alternate letter code table

___ 6. Read the sequence as far up the membrane or photocopy as possible. The bands become increasingly more compressed as you proceed from bottom to top, and at some point it will become too difficult to read the pattern.

___ 7. Once the first person has read the sequence, switch roles and have the second person read the sequence from the membrane while the first person compares to what is recorded on the data record sheet. Resolve any differences between the two sequence readings.

___ 8. Compare your sequence with those of your other lab partners. Do they agree? If not, go back to the membrane and examine your differences.

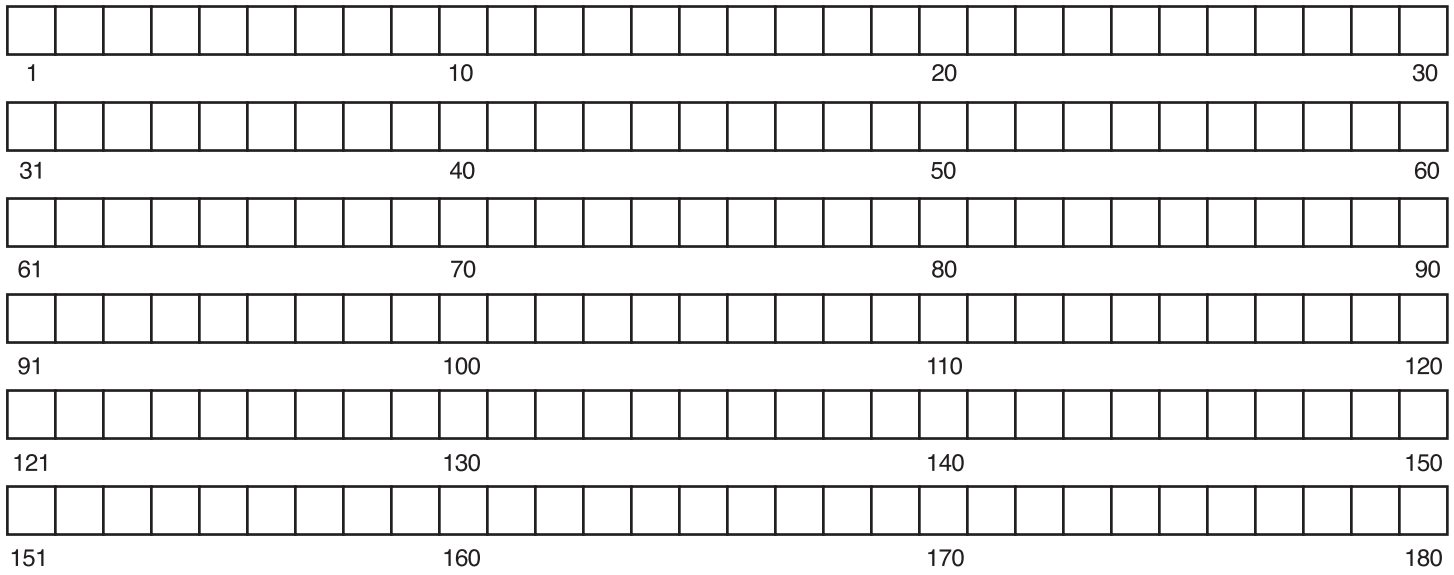
___ 9. Now compare the sequence from your lab group with the sequences obtained by other lab groups that used the same DNA template. Resolve any differences by checking the original membranes.

___ 10. Once you and your lab partners have agreed on the correct DNA sequence, your teacher will give you another data record sheet on which to record your final agreed upon DNA sequence. The final version of the Data Record Sheet is copied onto pink paper. Label the sheet with the date, your school, class section, lab group, DNA template number, and the names of all the people in your lab group and their signatures.

Student Name: _____ **Date:** _____

School: _____ **DNA Template:** _____

Class Section: _____ **Lab Group:** _____



Comments:

Figure 42. High School human genome program data record sheet—draft version

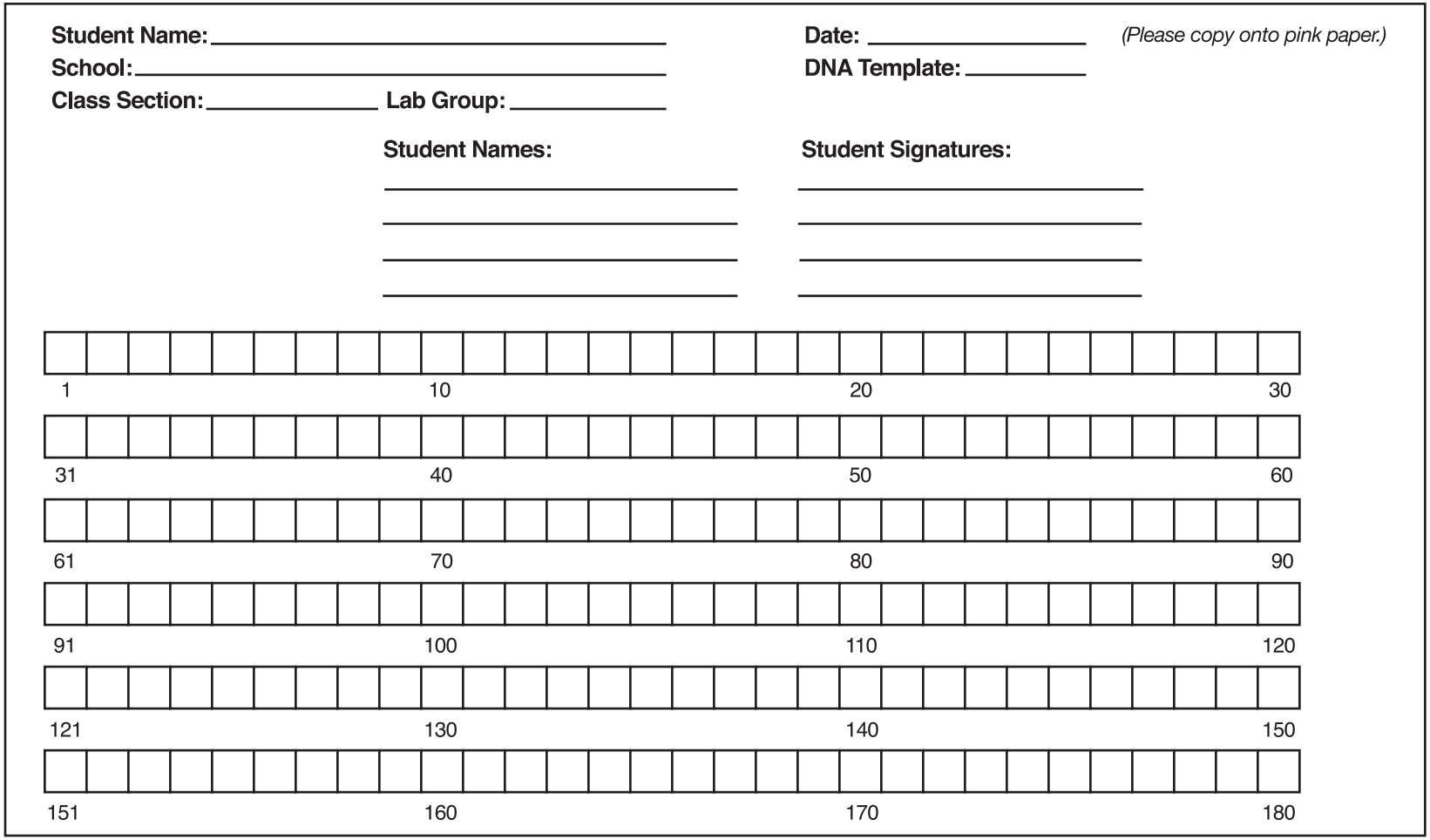


Figure 43. High School human genome program data record sheet—final version

Questions for Understanding

Answer the following questions in your lab notebook.

1. What are the bands on a sequencing gel?

2. Focus on the T lane as an example. Why do you get a band for each T in the DNA sequence?

   Think back to what occurred in your reaction tubes on the first day of the experiment.

3. Think about how we read a DNA sequence from the bottom to the top of a membrane.

   a. Are we reading the newly synthesized DNA or the template strand?

   b. What direction are we reading on the DNA strand, 5′ to 3′ or 3′ to 5′?

   c. Why is it important to always read and record the sequence in the same direction?

4. Provide explanations for the following experimental results:

   a. No bands are visible anywhere on the membrane.

   b. No bands are visible in one lane of the membrane.

   c. Two bands of equal intensity are visible at the same position on the membrane.

   d. Bands are visible at the same position across all four lanes of the membrane.

   e. Severe blotching on the membrane makes it hard to read.

5. Discuss how a membrane would be affected by each of the following scenarios:

   a. A lab group forgets to heat and ice its DNA samples prior to loading the gel.

   b. Dideoxynucleotide mixtures are mistakenly used that contain dideoxynucleotides at a much higher ratio (1:1), for instance, instead of the correct 1:9 ratio. What about a lower ratio like 1:999?

   c. A lab group accidentally loads both its A and C tube mixtures into the same lane on the sequencing gel.

6. A commonly asked question is, How do we know that every possible nucleotide will be represented in our sequencing ladder? Convince yourself that there will be DNA bands at every possible position on the gel by doing the following set of calculations:

   a. Calculate the maximum number of DNA molecules that can be synthesized in each reaction.

   Hint: How many DNA template molecules are in each reaction? The DNA concentration is 0.11 pmol/µL and the volume used is 4 µL (1 pmol = 10^{-12} mol). How many primer molecules are in each reaction? The primer concentration is 0.13 pmol/µL and volume is 8 µL. Since we do a total of 10 cycles, what is the theoretical maximum number of new DNA molecules in each reaction tube? (Think about which reagent will be limiting, the primer or the template.) Avogadro's number = 6.022 x 10^{23} molecules/mole.

   b. How many of these newly synthesized DNA molecules were loaded in each lane of the gel?

   Hint: What proportion of your total reaction did you load on the gel?

   c. How many DNA molecules are there in each band on the gel? (Assume that there is the same number in each band and in any one reaction. The DNA template is 300 nucleotides long. DNA synthesis could terminate at approximately one fourth of these in each reaction to make 75 fragments.)

Explain

Modeling the Assembly of DNA Fragments

Background

You have just sequenced a portion of the *CYP2A6* gene from one individual. Other students involved in this project are sequencing other pieces of the same gene from the same and other individuals. Eventually, we will get enough sequences to completely sequence this gene in all eight individuals. How will we compare the hundreds or thousands of different DNA sequences this project is generating?

Fortunately, we have computer technology to help with this awesome job. You will be using a software program called **Sequencher** to carry out what is called an **assembly**. The assembly program puts all the small segments of DNA together by finding overlaps in the nucleotide sequences of the different fragments to build one long, contiguous strand, or **contig** (see Figure 44). Students involved in this project will build a contig for each of our eight individuals.

Figure 44. Contig assembly

This activity models what the assembly program does. You will be the computer and will look for overlaps in the five DNA fragments on Figure 45.

Materials

Scissors
Tape
Paper DNA fragments

Procedure

1. Your teacher will give you a figure showing five DNA fragments.

2. Compare the sequences of the different fragments and find places where the sequences are the same.

3. Tape the overlapping fragments in place.

4. Continue until all five fragments have been assembled into one long contig.

Consider This

In this activity, you assembled five short DNA sequences, all of which are the same 5′ to 3′ strand and have perfectly matching overlaps. In reality, DNA assemblies are carried out on thousands of DNA fragments, and both strands of the original sequence are usually present. In addition, the matching sequences on different fragments may vary because of sequencing errors. How does this help you appreciate the power of computers in a project the scale of the Human Genome Project, in which 3 billion nucleotides were sequenced?

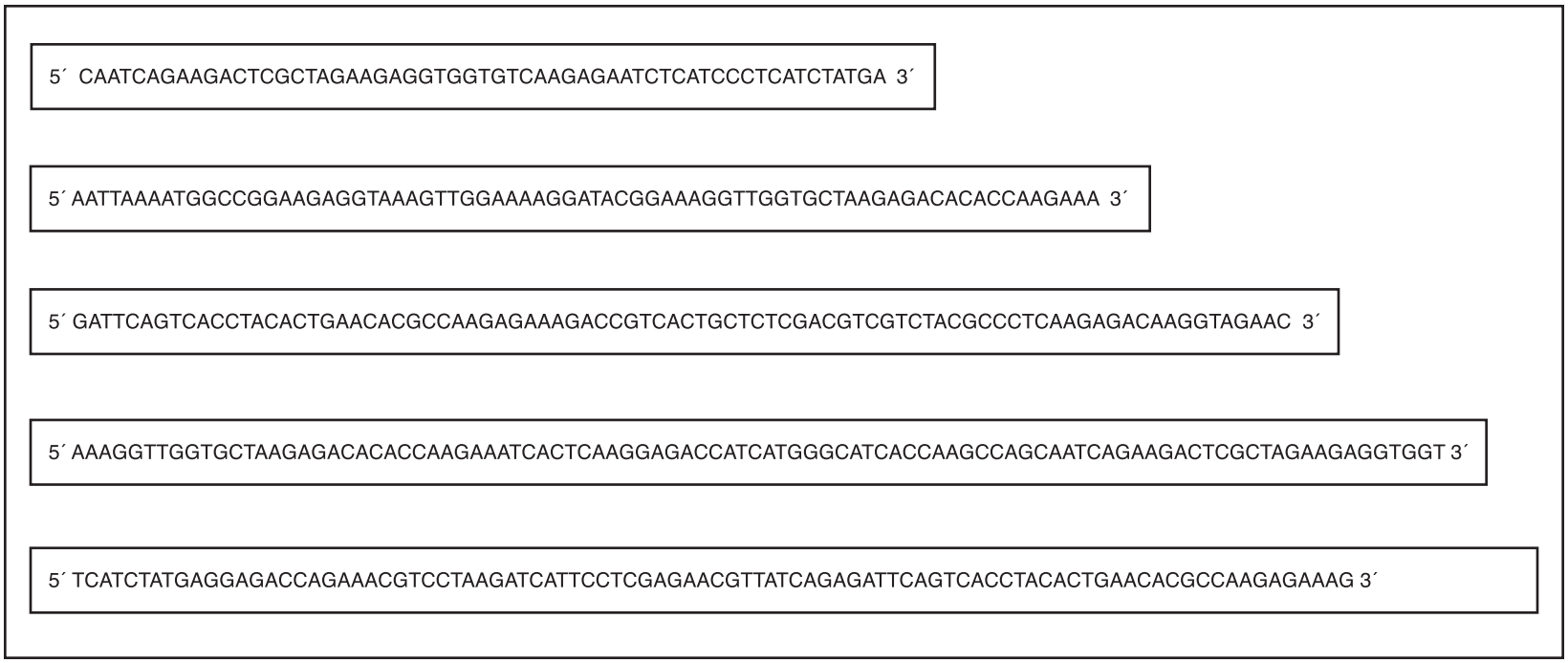


Figure 45. Fragments for contig assembly

Elaborate

DNA Sequencing—Experimental Procedure

Day 5: Building the Big Picture: Using the BLAST Search, Assembling Contigs, Amino Acid Translation, and Analysis of SNPs

Now that you have sequenced a piece of DNA, what are you going to do with your data? Short sequences of DNA can be analyzed by simple visual inspection. Typically, however, scientists are working with thousands of base pairs of sequence data and they may need to compare them with millions of base pairs of other sequences. This situation makes computer analysis of sequence data essential. In fact, this need has given rise to a new sub-discipline of biology known as bioinformatics.

In the first step of data analysis, you will use a program called BLAST to confirm that you sequenced a portion of the *CYP2A6* gene and to compare your sequence with those obtained by other scientists. Then you will use a computer program called Sequencher, which looks for overlaps in the sequences of DNA fragments and assembles them into a long contiguous piece called a **contig**. After you have assembled your classroom data, you will use Sequencher to translate the nucleotide sequence of your contig into the amino acid sequence it codes for. Finally, if you think you have found a SNP, you will test whether your potential SNP changes the amino acid sequence of the CYP2A6 enzyme.

Preparation for Building the Big Picture: Making a Folder of Your Class's DNA Data Files

Your teacher has created a new folder on your class computer labeled with your class name and the date. The DNA sequences from each of the lab groups in your class will be entered into separate data files in this one folder.

___ 1. Open a new Word document.

___ 2. Type in your nucleotide sequence. Do not put in any spaces, capital letters, or returns.

___ 3. Save the data as a text file, and use the DNA data file name as the name of the file. Save the file to the class folder.

___ 4. Each lab group will perform steps 1–3 for their sequencing data. Before proceeding, your teacher will load the class data onto all the computers you'll be using.

Elaborate

BLAST Search

Background

BLAST (Basic Local Alignment Search Tool) is a software tool that is used to determine whether a particular DNA or protein sequence matches with any other DNA or protein sequence in one of several large public databases. It can be accessed through the Internet at the National Center for Biotechnology Information (NCBI) Web site at http://www.ncbi.nlm.nih.gov/BLAST. In this activity, you will use BLAST to discover more about the fragment of DNA you have sequenced. This part of the experiment is best done in the school computer lab with two students working at each computer.

What can you learn about your DNA sequence from doing a BLAST search? First, you can check to make sure that you really have sequenced the DNA you intended to sequence. Second, you can see how well your sequence matches the sequences that other scientists have found for the same gene. Third, you can see how your sequence is similar to other genes, either different genes in the same organism, the same gene in different organisms, or both.

Follow the steps outlined below and then answer the questions that follow.

Procedure

Check off each step as you complete it.

___ 1. Go to the BLAST Web site at http://www.ncbi.nlm.nih.gov/BLAST.

___ 2. Select Nucleotide-nucleotide BLAST (blastn) under Nucleotide. The screen you see will look similar to Figure 46. The National Center for Biotechnology Information frequently changes the appearance of the submission page, but the basic functionality of the site will remain the same.

Figure 46. BLAST search window

___ 3. Enter your sequencing data into the box provided by copying from your text file in Word and pasting into the box. Then go to the beginning of your sequence and hit return to create an empty line. You are going to put the name of your template on this line. Type > followed by the name of your template. This is called **FASTA format**, and it should look like this:
>template name
aagtggctccgtagctactagctaatcgtacctag

___ 4. Click on the BLAST! button to submit your data. The screen will give you an estimate of how long your search will take, as well as a request ID number. Click on Format Results to view your results. This should take 1–3 minutes.

___ 5. Figures 47–49 show an example of a BLAST search result. The results are presented in three different ways:

a. A colored chart (Figure 47) shows the query (input) sequence as a line across the top and then uses color-coded bars to show significant regions of overlap with other sequences.

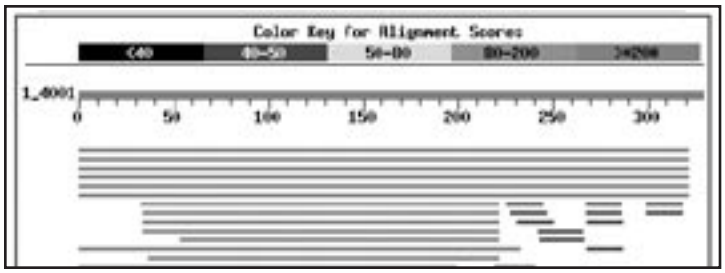


Figure 47. Chart depicting alignments obtained from a BLAST search

b. Under the colored chart is a list of sequences producing significant overlaps (Figure 48). Each line represents a different sequence, starting with the best match. The information on each line is divided into three segments. The first segment gives information about the entry of the sequence into a database, and it is often divided into three columns. The first column indicates in which database the match was found. Possible databases are GenBank (gb, located in the United States), EMBL (emb, located in Europe), DDBJ (dbj, located in Japan), and PDB (pdb, located in the United States). The second and third columns show numbers that can be used to access the sequence and associated information in the databases. If the third column looks word-like, your sequence has probably been found to be associated with a particular gene location. Note that the three columns are hotlinked. If you click on them, you can see the actual database record for the matching sequence. The second segment is a verbal description of the entry. This is followed by the score and E value, which are estimates of how strong the match is. The better the match, the higher the score and the lower the E value (for more information, check the

tutorial on BLAST at http://www.ncbi.nlm.nih.gov/Education/BLASTinfo/information3.html).

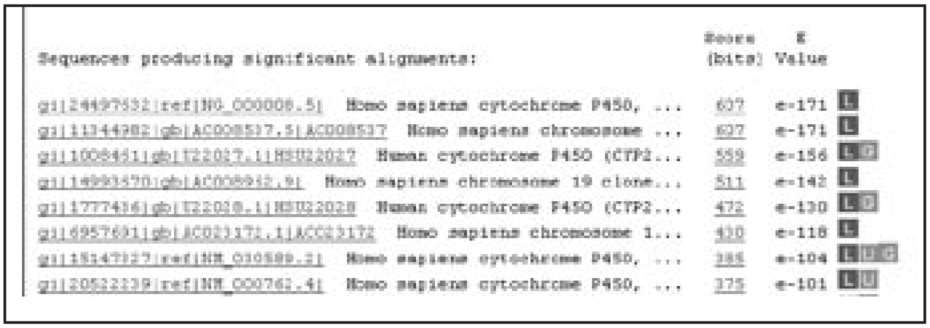


Figure 48. List of sequences that produce significant overlap with the query sequence

c. The third section (Figure 49) shows the regions of the query sequence that align with each of the sequences listed in Figure 48. For any sequence in the list, you can go directly to the sequence alignment by clicking on its score to the right. In this section, the query sequence is aligned with the subject sequence in the region or regions of overlap. A line is drawn between the perfectly matched nucleotides.

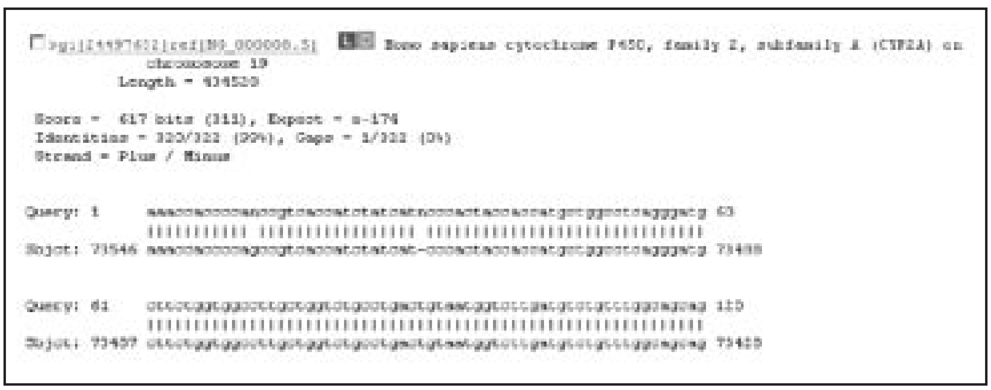


Figure 49. Alignment of your sequence with a sequence in the database

Questions for Understanding

Answer the following questions in your lab notebook.

___ 1. Does your BLAST search confirm that you sequenced part of the *CYP2A6* gene? How do you know?

___ 2. Compare your DNA sequence with the best match from the BLAST search. How similar are they? Do you observe any differences? If so, what do you think causes these differences?

___ 3. As you look down the list of sequences that matched your sequence, what patterns do you observe in the types of sequences that match yours? Discuss the patterns you observe.

Elaborate

Assembling Contigs, Amino Acid Translation, and Analysis of SNPs

Procedure

Check off each step as you complete it.

Part A: Assembling the Data

In this part, you will use the Sequencher program to assemble the short DNA sequences obtained by your class, and other classrooms involved in this project, into one long contig.

Preanalysis Set Up

You will need the following items on your computer:

- The Sequencher program
- Your classroom folder of sequence data (loaded onto your computer by your teacher)
- The subject data file for your subject (downloaded from the program Web site by your teacher)

___ 1. Open the Sequencher program by double-clicking on its icon. This is a demonstration version of the software, so you will not be able to save or print your work. All assemblies must be viewed on-screen.

___ 2. When Sequencher starts up, the Assembly window appears on your screen. You are now ready to import the DNA data files into the Assembly window. You can import data one file at a time or one folder at a time. You need to import two folders of sequences, your class folder and the subject data file for the individual whose DNA you sequenced. To import student data, go to the toolbar and chose File > Import & Export > Import Folder of Sequences. Select the folder containing your class data and choose Open. The file names contained within that folder will appear in the window.

___ 3. Sequencher will then prompt you to make sure you wish to import the entire folder. Choose Import All Files in Folder. Each data file will be imported into the Sequencher window. Repeat this step for the folder containing all the data for your subject.

___ 4. Select all the files in the Sequencher Assembly window by holding down the Apple key and the A key if you are using a Macintosh or the Ctrl + A keys if you are using a PC (see Figure 50).

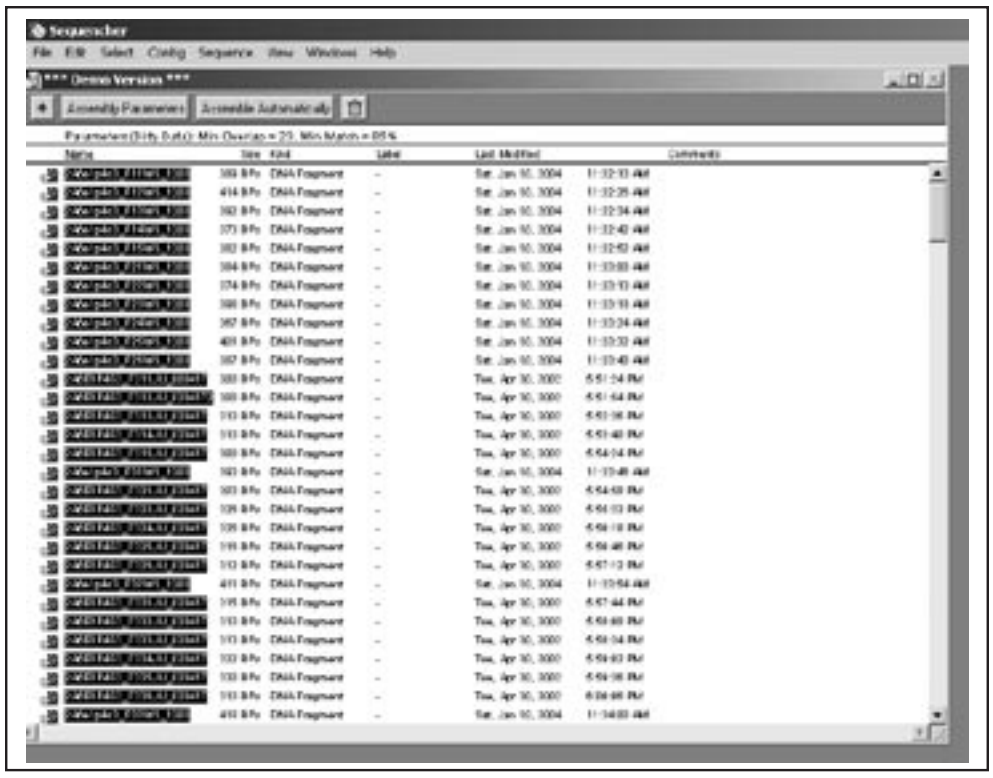


Figure 50. Sequencher assembly window

___ 5. To start the assembly process, click on Assemble Automatically at the top of the window. It may take some time for your computer to complete the analysis. Once it is finished, you will see files in the analysis window with names starting with "Contig." These are the assembled DNA fragments. All icons not labeled "Contig" denote sequences that Sequencher was not able to include in the assembly.

___ 6. To view any contig, double-click on its icon. This will open the Overview window, as shown in Figure 51. In this view, each horizontal line represents one automated or manual DNA sequence. The overview gives you a general idea of how the different DNA fragments overlap with each other to cover the contig. Important features of the sequence are marked by red and green symbols. A legend appears at the bottom of the window to assist you in interpreting your data.

The largest contig contains the reference sequence and everything that matches it. It should also provide the most complete coverage of the gene. Smaller contigs contain sequences that don't match well with the reference sequence, and generally, these will be sequences with more errors in them.

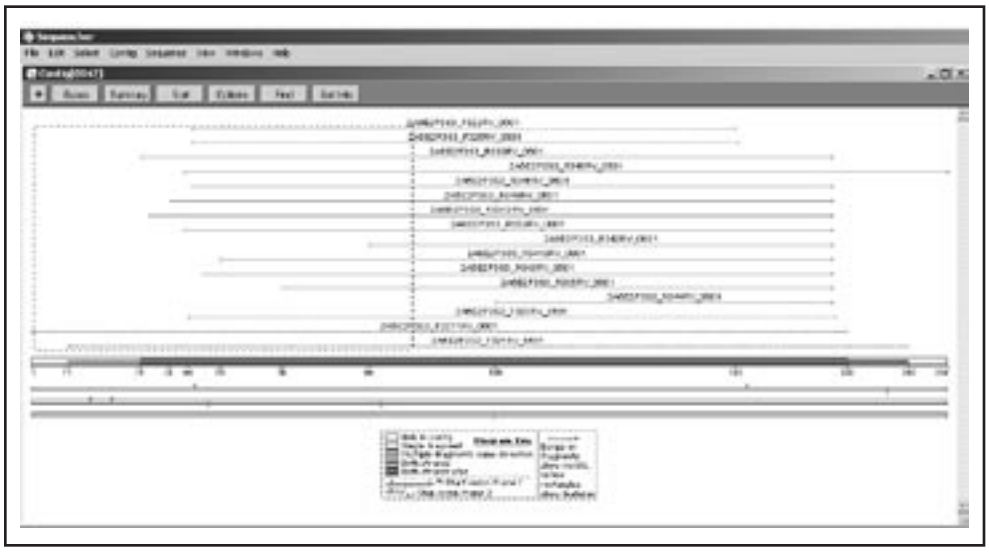


Figure 51. Overview window

___ 7. Click on the Bases button in the upper left-hand corner of the window. A different view will appear that shows the raw sequences of each fragment making up the contig. Below the raw data is the **consensus sequence**, which is the sequence that best agrees with all of the raw sequences in the contig.

___ 8. How well does your sequence agree with the consensus sequence? First, you need to find your sequence. Click on the first contig to open it. Scroll down the sequence names on the left-hand side until you find your sequence. Highlight your sequence name by clicking on it once (it will be highlighted in blue). Now use the scroll keys at the bottom and right-hand side of the screen to find your sequence (it will line up with its name). If you don't find your sequence in the first contig, check the others. If your sequence doesn't fit in any of the contigs, it will appear as a single file at the bottom of the list of contigs.

___ 9. To check the quality of your data, scroll across your fragment, noting whether there are any disagreements with the consensus. Symbols under the consensus sequence show where a disagreement has occurred:
- A large dot (•) under a nucleotide indicates that one or more of the sequences do not agree with the consensus at this nucleotide.
- A plus sign (+) means that one or more sequences contain an *N* at this position.
- A colon (:) is inserted in the consensus sequence or one of the raw sequences in place of a missing nucleotide in order to keep the sequences in register.

If there is a disagreement in your data, reexamine the membrane to make sure you read the sequence correctly. Did you make an error reading the sequence? If so, you can change your sequence on the assembly. The

changed nucleotide will be bright pink to show that it has been changed by the editor. This will not change the sequence in the original data file. To change the original file, you need to open that file and change it directly.

___ 10. Now scroll through each of the contigs to see how well the data agree. What do you think about the quality of the data we have collected so far? Are there any areas where you think we need to do more sequencing? Why?

Part B: Amino Acid Translation and Analysis of Potential SNPs

In addition to assembling DNA sequences, Sequencher can be used to learn the amino acid sequence encoded by a string of nucleotides. In this section you will use Sequencher to translate your nucleotide sequence into its amino acid sequence. If you think you have found a SNP, you will test whether it would result in a change in the amino acid sequence of the CYP2A6 enzyme.

Before getting started, you need to review how the information of DNA is used to code for a protein in human cells. To make a protein, the cell first makes an RNA copy of its gene through a process called **transcription**. The RNA copy is called **messenger RNA** or **mRNA**. After transcription, the mRNA is processed in several ways. Pieces of the mRNA called **introns** are cut out of the mRNA, and the remaining pieces, called **exons**, are joined end to end (shown in Figure 52). The mRNA is moved to the cytoplasm, where its information is used by ribosomes to make an amino acid chain. The ribosome reads the sequence of

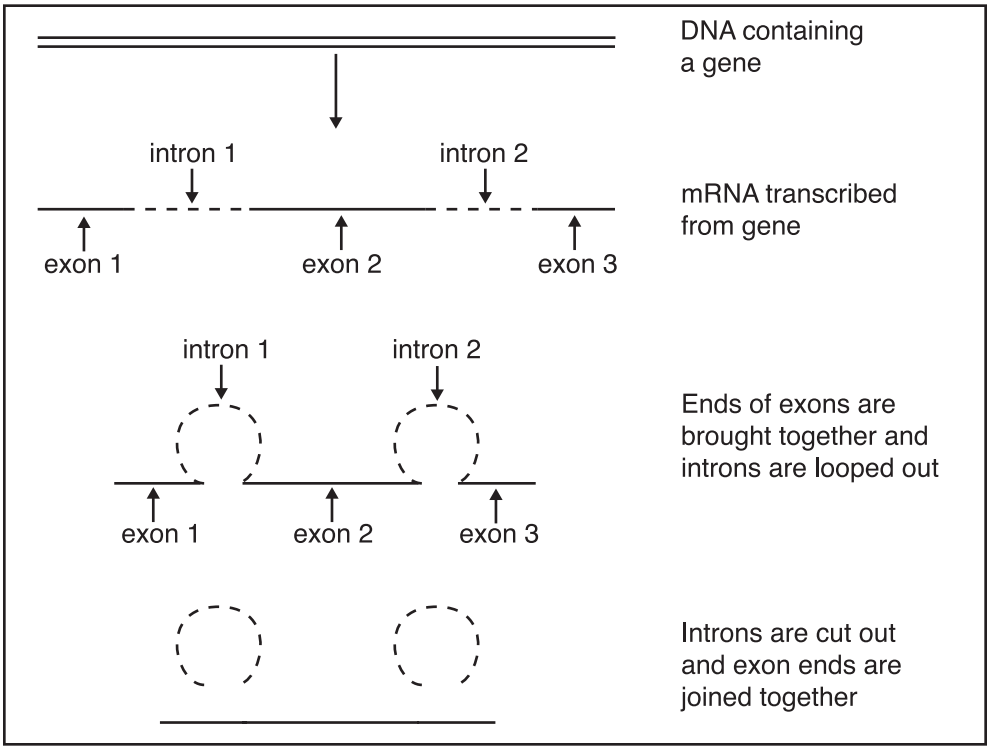


Figure 52. Removal of introns from mRNA

the mRNA in groups of three nucleotides, called codons. Each codon specifies a specific amino acid, and the ribosome joins amino acids in a chain in the order specified by the mRNA. The ribosome continues to add amino acids to the chain until it reaches a stop codon, which signals the end of synthesis of that chain.

The DNA template you sequenced may contain regions of one or more introns as well as an exon. Before you can translate your DNA into its amino acid sequence, you need to get rid of the introns. You'll do this using Sequencher to a) assemble your classroom sequencing data together with the known DNA sequence of the exon you have sequenced, b) trim away the introns, and c) translate the remaining exon into its amino acid sequence.

In this section, you will also use Sequencher to analyze possible SNPs in your DNA sequence. The most interesting SNPs are the ones that result in a change in the amino acid sequence of the CYP2A6 enzyme. You will use Sequencher to test whether any SNPs you identified result in a change in the amino acid sequence.

Preanalysis Set Up

You will need the following items on your computer:

- The Sequencher program
- Your classroom folder of sequence data
- The file of the exon for the DNA fragment you sequenced (for example, if your DNA template name contained E1, then you need the sequence for exon 1). These are available on the program Web site under Teacher Resources in the Software/Data Download Area.

In addition to the computer, you need a copy of the CYP2A6 amino acid sequence and a good-quality photocopy of each of your classroom sequencing membranes.

___ 1. Open the Sequencher program by double-clicking on its icon. This is a demonstration version of the software, so you will not be able to save or print your work. All assemblies must be viewed on-screen.

___ 2. When Sequencher starts up, the Assembly window appears on your screen. You are now ready to import the DNA data files into the Assembly window. As you learned in *Part B*, Sequencher allows you to import one file at a time or one folder at a time. You need to import your class folder. Go to the toolbar and chose File > Import & Export > Import Folder of Sequences. Select the folder containing your class data and choose Open. The file names contained within that folder will appear in the window.

___ 3. Sequencher will then prompt you to make sure you wish to import the entire folder. Choose Import All Files in Folder. Each data file will be imported into the Sequencher window.

___ 4. Load all the forward and reverse DNA sequences from your class that have the same template number as your group. Do not include DNA sequences from different templates.

___ 5. Find the exon file that corresponds to your DNA template and import this sequence into the Sequencher Assembly window as well. Exon files can be found on the program Web site under Teacher Resources in the Software/ Data Download Area.

___ 6. Select all the files in the assembly window and click on the Assemble Automatically button.

___ 7. Once your data has assembled, open the contig by double-clicking on its icon. In the Overview view, you should see an arrow for each class sequence and the exon (either a green arrow pointing left to right or a red arrow pointing right to left). In order to complete the following analysis, it is imperative that the exon has a *red* arrow that points from *right to left*. If the arrow for the exon is pointing left to right instead, go to the View menu and select Reverse & Comp. This will prompt Sequencher to show the complementary sequence for every sequence in the contig.

___ 8. Does some of your data extend beyond the exon? These sequences are portions of the introns (the regions of the gene that are removed when the mRNA is made) on both sides of the exon (the region that is translated into protein) you're sequencing and are not useful in this analysis. To remove them,

a. click on the Bases button at the top of the window.
b. using the nucleotide selection box, highlight any bases on the consensus sequence that extend to the *left* of the known exon (shown in Figure 53). This will also highlight the corresponding data in all of your data files. Stop highlighting at the base immediately to the left of the first base of the exon and hit Backspace (PC) or Delete (Mac). If Sequencher asks if you'd like to fill bases from the left or the right, click the button that says Fill Void from Left.

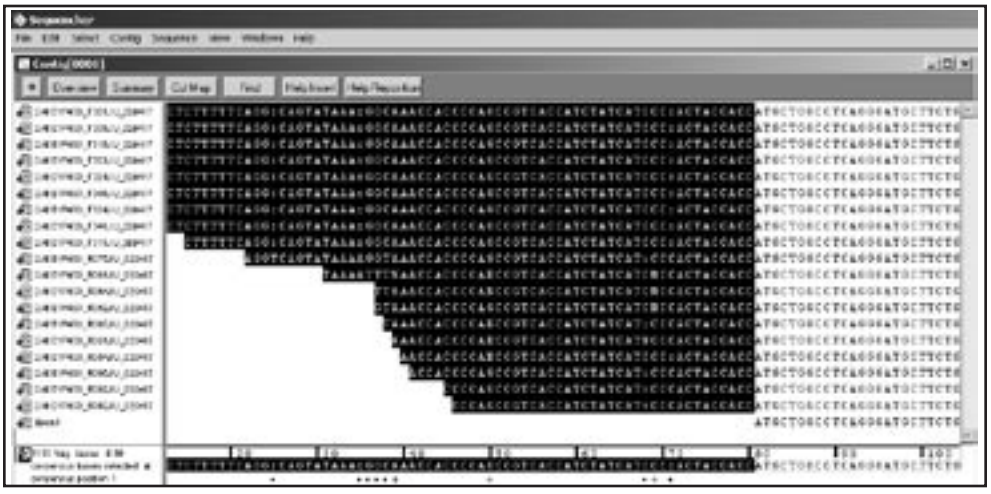


Figure 53. Highlight the bases to the left of the exon

___ 9. Next, highlight the bases to the right of the exon if there are any and hit Backspace or Delete. This time you want to fill the void from the right.

___ 10. When you've finished removing the unneeded data, click the Overview button. Your contig should look like Figure 54 in the Overview window. Notice that no arrows extend beyond the length of the exon arrow.

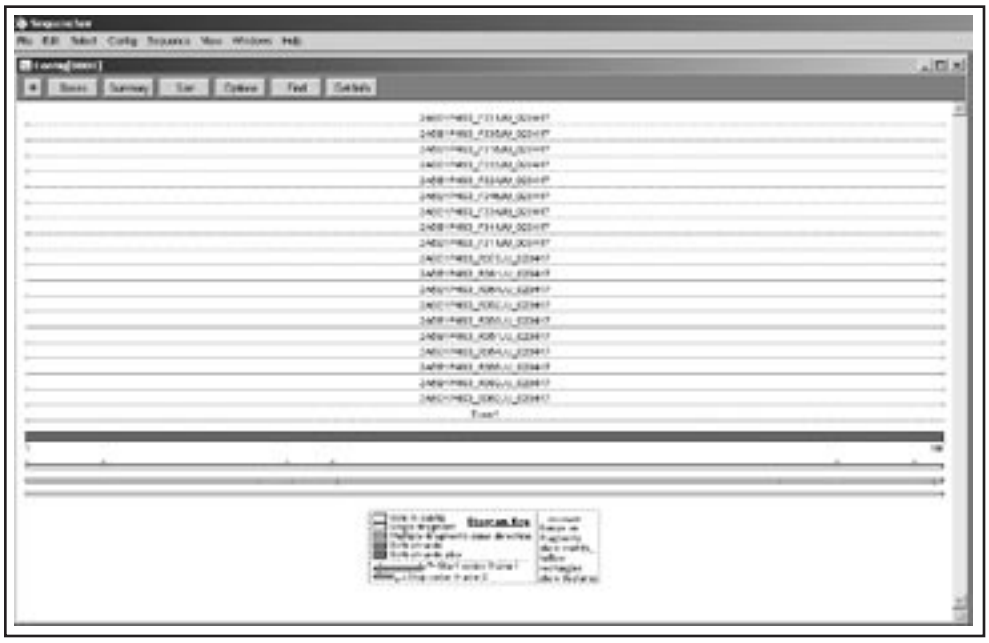


Figure 54. Overview window

___ 11. Now go back to the Bases view so you can compare the sequences of the classroom data and the exon.

___ 12. Your first task is to remove incorrectly inserted nucleotides. Below the raw data is a consensus sequence. If one or more sequences have an extra nucleotide at one position, Sequencher puts a colon (:) in the consensus sequence to keep all the sequences lined up. To remove this position, highlight the colon in the consensus sequence, as shown in Figure 55, and hit Backspace or Delete.

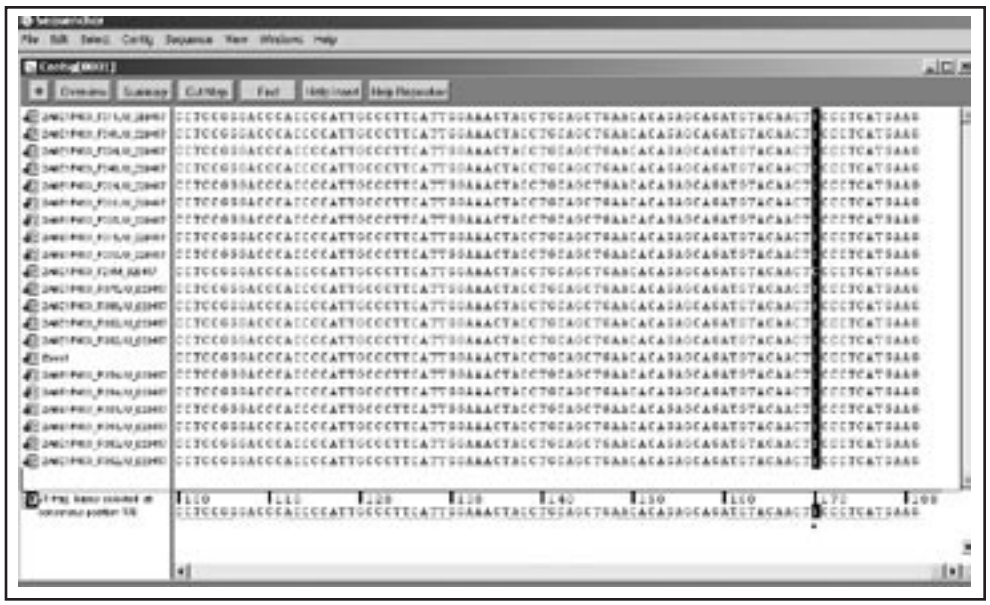


Figure 55. Highlight the colon in the consensus sequence

Now you need to edit any remaining differences in the assembly. As mentioned before, a large dot (•) indicates that one or more of the sequences do not agree with the consensus at this nucleotide, and a plus sign (+) means that one or more sequences contains an *N*. There are several possible reasons for each disagreement: (1) an error was made in reading one or more of the sequences, (2) an error was made in entering data, (3) the sequencing data cannot be read because there are more than two bands at this position or this part of the membrane is hard to read, or (4) this is a potential SNP.

___ 13. You will need to decide whether each sequence difference is a SNP or an error. In the Bases view, begin at the left side of the consensus sequence until you reach the first + or •. Check the membrane of all of the sequences that do not agree with the consensus. Was this nucleotide entered correctly? If not, change the nucleotide in each disagreeing sequence to what you think is the correct one. If you see two equal bands at this position and think that this is a potential SNP, use the alternate letter code in the table in Figure 56. (One of the alternate letters may have already been used at this position, but confirm that you agree with this assignment by checking the membrane.) Additional directions for editing data are given on the tool bar above the Bases view of the assembly (e.g. "Help Insert" and "Help Reposition").

Nucleotides observed at same position	Alternate letter code
A and G	R
C and T	Y
G and T	K
A and C	M
G and C	S
A and T	W

Figure 56. Alternate letter code table

___ 14. Continue as in step 13 for all of the disagreements in the consensus sequence.

___ 15. Now you are ready to translate the consensus sequence. Here is what you need to do:

a. Make sure that the exon appears as a red arrow in the Overview window. If it is a green arrow, then go to the View menu and select Reverse & Comp.

b. Now go to the Bases view, and from the toolbar select View, Translation, and Protein 1st Frame.

Your amino acid sequence should be identical or very similar to the one given in Figure 57 for the exon you sequenced. If it is completely different, it probably means that you have inserted or deleted one or more nucleotides into your sequence, so that the codons are out of phase. If this happens, check your DNA assembly again for possible errors and then repeat the translation.

___ 16. Now comes the fun part—analyzing any potential SNPs to see if the nucleotide difference confers a change in the amino acid sequence of CYP2A6. Examine any position where you find the alternate nucleotide code (R, Y, M, K, S, and W). Remember, we use these letters to indicate that two nucleotides appear at this position. In the consensus sequence, change the base at this position to reflect one of the two possible nucleotides. What amino acid is encoded by this sequence? Now change the sequence to the other nucleotide. Did the amino acid change? Your teacher will let the StarNet staff know if you find a potential SNP.

Figure 57. Complete amino acid sequence of the CYP2A6 enzyme:

Exon 1
{MetLeuAlaSerGlyMetLeuLeuValAlaLeuLeuValCysLeuThrValMetValLeuMetSerVal
TrpGln GlnArgLysSerLysGlyLysLeuProProGlyProThrProLeuProPheIleGlyAsnTyrLeuGln
LeuAsn ThrGluGlnMetTyrAsnSerLeuMetLys}

Exon 2
{IleSerGluArgTyrGlyProValPheThrIleHisLeuGlyProArgArgValValValLeuCysGlyHisAsp
AlaValArgGluAlaLeuValAspGlnAlaGlnGlnPheSerGlyArgGlyGluGlnAlaThrPheAsp
TrpValPheLysGlyTyrGly}

Exon 3
{ValValPheSerAsnGlyGluArgAlaLysGlnLeuArgArgPheSerIleAlaThrLeuArgAspPheGly
Val GlyLysArgGlyIleGluGluArgIleGlnGluGluAlaGlyPheLeuIleAspAlaLeuArgGlyThrGly}

Exon 4
{GlyAlaAsnIleAspProThrPhePheLeuSerArgThrValSerAsnValIleSerSerIleValPheGlyAsp
ArgPheAspTyrLysAspLysGluPheLeuSerLeuLeuArgMetMetLeuGlyIlePheGlnPheThrSer
ThrSerThrGlyGln}

Exon 5
{LeuTyrGluMetPheSerSerValMetLysHisLeuProGlyProGlnGlnGlnAlaPheGlnLeuLeu
GlnGlyLeuGluAspPheIleAlaLysLysValGluHisAsnGlnArgThrLeuAspProAsnSerProArg
AspPheIleAspSerPheLeuIleArgMetGlnGlu}

Exon 6
{GluGluLysAsnProAsnThrGluPheTyrLeuLysAsnLeuValMetThrThrlLeuAsnLeuPheIle
GlyGly ThrGluThrValSerThrThrLeuArgTyrGlyPheLeuLeuLeuMetLysHisProGluValGlu}

Exon 7
{AlaLysValHisGluGluIleAspArgValIleGlyLysAsnArgGlnProLysPheGluAspArgAlaLys
MetProTyrMetGluAlaValIleHisGluIleGlnArgPheGlyAspValIleProMetSerLeuAlaArg
ArgValLysLysAspThrKysPheArgAspPhePheLeuProLys}

Exon 8
{GlyThrGluValPheProMetLeuGlySerValLeuArgAspProSerPhePheSerAsnProGlnAsp
PheAsnProGlnHisPheLeuAsnGluLysGlyGlnPheLysLysSerAspAlaPheValProPheSerIle}

Exon 9
{GlyLysArgAsnCysPheGlyGluGlylLysAlaArgMetGluLeuPheLeuPhePheThrThrValMet
GlnAsnPheArgLeuLysSerSerGlnSerProLysAspIleAspValSerProLysHisValGlyPheAlaThr
IlePro ArgAsnTyrThrMetSerPheLeuProArg}

Questions for Understanding

Answer the following questions in your lab notebook.

1. Figure 58 shows a part of an assembly file for some CYP2A6 data. The column at the left lists the names of the DNA sequence files. To the right, the corresponding nucleotide sequences are lined up. The consensus sequence appears across the bottom of the screen.

   a. What do the symbols • and + below the consensus sequence mean?

   b. Do you think that we have good-quality data in this region? Explain.

   c. Suggest two things that we could do to improve the quality of data in this region.

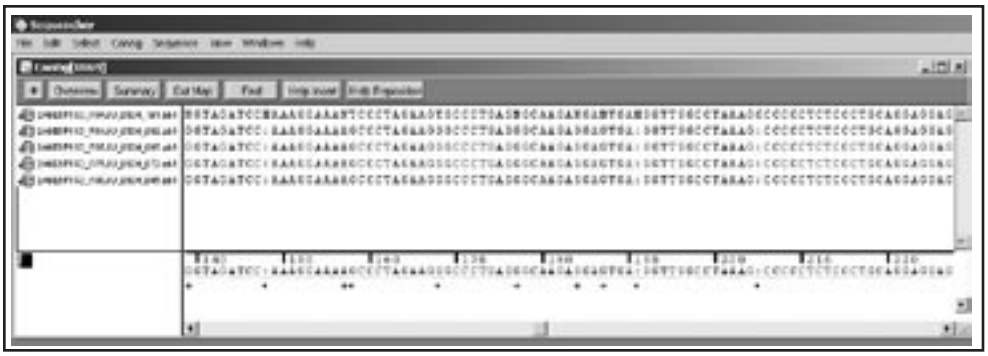


Figure 58. Partial assembly file

2. In *Part C*, why did you remove the intron sequences from your consensus sequence before carrying out the amino acid translation?

3. How would the following changes in the nucleotide sequence of an exon affect amino acid sequence it encodes?

   a. Insertion of one nucleotide

   b. Deletion of two nucleotides

   c. Insertion of three nucleotides

4. Use the Genetic Code table in Figure 40 and the alternate letter code table (Figure 56) to answer this question.

   After assembling their classroom data and trimming the intron sequences, a class obtained the following consensus sequence from a segment of exon 3.

   GTG GT**R** TTC AGC A**W**C GGG GAG

Their data suggested that the person they sequenced might be heterozygous at position 6 (A and G) and at position 14 (A and T). The students used the translation option of Sequencher to determine the amino acid sequence encoded by this nucleotide sequence and tested whether the two potential SNPs results in a change in amino acid sequence.

Complete the amino acid translation for trial 1 and 2 below, and then answer the questions that follow.

In trial 1 they tested A in Position 6 and A in Position 14

GTG	GTA	TTC	AGC	AAC	GGG	GAG
Val						

In Trial 2 they tested G in Position 6 and T in Position 14

GTG	GTG	TTC	AGC	ATC	GGG	GAG
Val						

a. Did the SNPs at positions 6 and 14 of exon 3 result in a change in the amino acid sequence of CYP2A6? Explain your answer.

b. Discuss in general terms why some SNPs result in changes in the encoded protein and others do not. Refer to the Genetic Code table to develop and support your arguments. Are certain positions in a codon (positions 1, 2, or 3) more or less likely to result in a change? Give examples to support your answer.

Elaborate

Reading:
The Role of the StarNet Project in Drug Addiction Research

What will happen to your sequencing data? After you have analyzed your DNA sequencing membrane, your teacher will submit your data to the StarNet Project staff, who will check the sequences and post them on the project Web site. If you think you have found a SNP, the StarNet scientists will first check your sequencing ladders to make sure they agree with your analysis. If the data appear conclusive, then they will send the same DNA template to additional classrooms for resequencing. Once confirmed, the SNP will be submitted to a SNP database at the National Center for Biotechnology Information called dbSNP.

How will the sequencing data from the StarNet Project contribute to scientific research, especially the understanding of drug addiction? The SNPs discovered by students will be added to the growing list of known mutations in the *CYP2A6* gene. So far, nearly 40 different genetic variants of the *CYP2A6* gene have been identified. These include deletions, duplications, and SNPs. The amount of CYP2A6 enzyme a person makes depends on which two forms of the CYP2A6 gene the person has inherited.

Why is there so much interest in genetic variation in the *CYP2A6* gene? One reason is that this enzyme is involved in the metabolism of a number of therapeutic drugs, either converting an inactive form of a drug to the active form or making the active form inactive. Understanding how our genes contribute to the way our bodies metabolize drugs is a new field called pharmacogenetics. In the future, physicians may check what forms of certain drug-metabolizing genes a patient has before prescribing a drug that best fits his or her needs. CYP2A6 is also involved in the metabolism of environmental chemicals. It activates several pre-carcinogens to the carcinogenic form, including aflatoxin B1, a compound produced by a fungus that commonly infects peanuts, and several compounds found in tobacco smoke. And finally, different variants in the *CYP2A6* gene either increase or decrease nicotine metabolism, which may have an effect on people's smoking behavior.

Research and our own observations show that smokers adjust their smoking to keep nicotine at a certain level. For example, when switched to low nicotine cigarettes, smokers tend to smoke more cigarettes. Also, when smokers use a nicotine patch or are given intravenous nicotine, they smoke less. These observations led researchers to hypothesize that variation in the amount of CYP2A6 enzyme made by an individual would affect smoking behavior. Several studies have shown that people who make lower amounts of CYP2A6 are either non-smokers or they smoke less than people who make normal amounts of the CYP2A6 enzyme.

However, not all research studies support this hypothesis. What are some reasons for this conflict? In some of the earlier experiments, researchers overestimated the number of certain *CYP2A6* alleles because of technical difficulties.

These experiments are being repeated using new procedures. A second reason is that different labs use different methods to measure how much people smoke. Some studies rely on self-reporting of number of cigarettes smoked, but this does not take into account how intensely the person smokes or how much of the cigarette is consumed. Other studies use additional measurements that are independent of self-reporting, including the level of carbon monoxide in the breath, and levels of nicotine and cotinine in the urine and plasma. These techniques give a more accurate measurement of how much the person has smoked. Another consideration is that other genes may also be involved in determining how much a person smokes (e.g. dopamine receptor, dopamine synthesizing and metabolizing enzymes, nicotinic acetylcholine receptor). This means that the contribution from CYP2A6 may be only part of the effect. Finally, certain alleles are more common in some ethnic groups than others, so results from different labs may be skewed depending on which populations were studied.

Understanding which genes predispose individuals to become smokers or continue to smoke may help in developing treatments and cessation programs. The drug, methoxsalen, inhibits CYP2A6 in vitro. When this drug was given to smokers, they smoked less. These findings are exciting because they suggest that it might be possible to help smokers quit smoking by inhibiting their CYP2A6 enzyme. More extensive studies are needed to determine whether this approach could be widely effective.

Once we know that a gene is associated with a genetic condition, it is possible to develop a DNA-based test for it. However, it is unlikely that genetic testing for the *CYP2A6* gene would be offered widely, because of the possible harm such a test could cause. For example, some people might choose to smoke based on their genotype, even though they would still be at risk for the many other diseases associated with smoking. Other people might face genetic discrimination if their genotype suggests that they are more likely to smoke or develop cancer based on their genetic profile. Additional studies are still needed to determine whether there is a correlation between the amount of CYP2A6 enzyme that a person makes, smoking behavior, and susceptibility to lung cancer.

Evaluate

Pulling It All Together

This final activity provides an opportunity for you to pull together everything you learned during this unit. Complete one of the two activities below (Case Study 1 or Case Study 2).

Case Study 1: Family Dilemma

Hi, I'm Chris. Your teacher said that I could come in and ask you for help. I heard that you've been studying how nicotine addiction can affect families. I'm so angry at my twin brother, Kyle. Last week I caught him smoking on the way home from school. When I confronted him about it, he turned up his nose and said, "You smoke, why shouldn't I?" But that's the whole point—I do smoke, but I don't want to anymore and I just can't quit. It has already messed up my life. I was kicked off the track team last week because my coach caught me smoking.

I started smoking in 9th grade. It was so easy—Mom and Dad both smoke, so there were always cigarettes around the house. In fact, almost all of our relatives smoke. My dad coughs all the time and grandma, his mom, died of lung cancer when she was pretty young. I remember visiting her before she died—she was still smoking through her tracheotomy. I know people who have been able to quit smoking without problems, but no one in my family has been able to stop.

Here's where I need your help. I need to convince Kyle that he shouldn't even start smoking. He's annoying, but he's pretty smart. I can usually convince him if I back up what I say with facts and evidence. Can you come up with a presentation that I can show to my brother that helps him understand why he shouldn't start smoking?

1. In order to help Chris convince Kyle that he shouldn't start smoking, first choose a format for your presentation, such as a PowerPoint presentation, poster, brochure, or letter to Kyle.

2. Your presentation should include the following information:
   - Discuss how drugs of addiction interfere with the normal process of neurotransmission.
   - Discuss the factors that determine a multifactorial trait like nicotine addiction. Be sure to include the following:
     a. How DNA codes for proteins and proteins can result in a physical trait.
     b. Why sequence variation in a gene may or may not affect the protein it encodes.
     c. How genetic variation and the environment both influence addictive behavior.
   - Relate the process you used to sequence DNA to your understanding of DNA structure and synthesis (chemical and physical perspective of DNA) and specialized molecular biology reagents. Be sure to include the

use of chain terminators, gel electrophoresis, and thermal cycling.
- Describe the use of bioinformatics tools to analyze DNA sequence data
- Suggest further research needed to establish a causal relationship between genetic variation and smoking behavior.
- Discuss the application of the results generated by this type of research

3. Develop your presentation using what you have learned throughout the unit. Include artwork, diagrams, and data where appropriate.

4. Share your findings with your class.

Case Study 2: Laboratory Group Meeting

You are a university professor, and the focus of your research is the genetics of drug addiction. In particular, your lab studies genetic variation in genes related to nicotine addiction, including the *CYP2A6* gene. The School of Medicine at your university sponsors a lecture series called "Science and Medicine," which is intended for faculty, students and staff from the campus. You have been invited to give a lecture in this prestigious series.

Your challenge is to prepare a talk explaining the significance of your research to a diverse audience from a wide range of disciplines. You plan to include a general background, procedure, analysis of your results and future directions for your research.

1. Choose a format for your presentation to the faculty, such as a PowerPoint presentation, poster, or paper.

2. Your presentation should include the following information:
   - Discuss how drugs of addiction interfere with the normal process of neurotransmission.
   - Discuss the factors that determine a multifactorial trait like nicotine addiction. Be sure to include the following:
     a. How DNA codes for proteins and proteins can result in a physical trait.
     b. Why sequence variation in a gene may or may not affect the protein it encodes.
     c. How genetic variation and the environment both influence addictive behavior.
   - Relate the process you used to sequence DNA to your understanding of DNA structure and synthesis (chemical and physical perspective of DNA) and specialized molecular biology reagents. Be sure to include the use of chain terminators, gel electrophoresis, and thermal cycling.
   - Describe the use of bioinformatics tools to analyze DNA sequence data
   - Suggest further research needed to establish a causal relationship between genetic variation and smoking behavior.
   - Discuss the application of the results generated by this type of research

3. Develop your presentation using what you have learned throughout the unit. Include artwork, diagrams, and data where appropriate.

4. Share your findings with your class.

Glossary

acetylcholine: One of several chemicals used as a neurotransmitter in our bodies.

agarose: A polysaccharide derived from seaweed that is used to prepare gels for electrophoresis.

alkaline phosphatase (AP): An enzyme that removes a phosphate group from its substrate. It can be used as part of a DNA visualization process. Removal of the phosphate group produces a colored precipitate that indicates the presence of a DNA fragment on a nylon membrane.

assembly: Putting sequenced fragments of DNA into their correct order along the chromosome.

axon: Extensions of nerve cells that carry the nerve impulses away from the nerve cell body.

biotin: A vitamin molecule. In the context of DNA visualization, biotin specifically binds to the protein streptavidin, helping attach the enzyme alkaline phosphatase to the DNA fragments being visualized.

BLAST (Basic Local Alignment Search Tool): A computer program that searches for sequence similarities. It can be used to identify homologous genes and proteins in different organisms.

cell body: The portion of a neuron that contains the nucleus and most of the cytoplasm; distinct from the axon and dendrites.

chain termination: Refers to a type of DNA sequencing developed by Fred Sanger. It involves the incorporation of modified nucleotides that stop growing DNA chains at each nucleotide position.

codon: Three nucleotide sequence that codes for an amino acid during protein synthesis. There are 61 codons that code for the 20 different amino acids and three codons, called stop codons, which result in termination of a protein chain.

consensus sequence: A DNA sequence that is characteristic of a particular gene element. Consensus sequences for a particular gene element share most but not all nucleotides.

contig: 1. A contiguous sequence of DNA created by assembling shorter, over-

lapping sequenced fragments of a chromosome. 2. A list or diagram showing an ordered arrangement of cloned, overlapping fragments that collectively contain the sequence of an originally continuous piece of DNA.

cotinine: A breakdown product of nicotine, which is modified and eliminated from the body.

cycle sequencing: A DNA sequencing technique that combines the chain termination method with aspects of the polymerase chain reaction.

CYP2A6: A protein member of the cytochrome P450 family. This enzyme converts nicotine to cotinine, which is then modified and eliminated from the body.

cytochrome P450: A family of enzymes produced in the liver that are involved with the detoxification of fat-soluble molecules including those associated with drug metabolism.

dendrite: A cluster of small fibers on the cell body of a neuron that receives chemical messages from neighboring neurons and transmits them to the cell body.

deoxynucleotide: A building block of DNA. A deoxynucleotide consists of a deoxyribose sugar, a triphosphate group, and one of four nitrogen bases (A, C, G, or T).

dideoxynucleotide: Synthetic nucleotides lacking both 2′ and 3′ hydroxyl groups. They act as chain terminators during DNA sequencing reactions.

DNA polymerase: An enzyme that adds nucleotides to a replicating DNA strand.

dopamine: A chemical messenger (neurotransmitter) that regulates brain processes such as those that control movements, emotions, pleasure, and pain.

dopamine transporter: A specialized pump located in the cell membrane of a neuron. It reabsorbs dopamine molecules after they have been released during neurotransmission.

drug: Any chemical substance other than food or substances needed for normal life that causes a change in the structure or function of the body through its chemical actions. In a medical context, this includes any substance used in the diagnosis, prevention, treatment, or cure of a disease. In an abuse context, this includes substances that alters consciousness and may be habit forming.

drug abuse: The continued use of a drug for any reason other than a medical one. This includes taking a psychoactive drug to get high or taking steroids to increase athletic ability.

drug addiction: Uncontrollable, compulsive drug seeking and use, even in the face of negative health and social consequences. Addiction results in a physical

or psychological dependency on a drug.

eukaryotic cell: A cell that has a membrane-bound nucleus, as well as other membrane-bound structures. Organisms that have this cell type include animals, plants, protists, and fungi, and they are referred to as eukaryotes.

exon: A segment of mRNA in a eukaryotic cell that is translated into protein. During mRNA formation, segments of the mRNA are cut out, and the remaining pieces, called exons, are joined end to end (see **intron**).

FASTA format: A method of entering DNA or amino acid sequence data that is compatible with popular sequence comparison programs such as BLAST and FASTA.

genomic DNA: DNA isolated from the nucleus of a cell.

genotype: The genetic makeup of an individual. The expression of genotype as visible traits is called the phenotype.

intron: A segment of RNA that is removed from an mRNA before the mRNA is translated into protein (see **exon**).

multifactorial trait: A trait whose phenotype is influenced by the combined action of many genes and the environment.

neuron: A principal class of cells in the nervous system, composed of three parts: the cell body, dendrites, and axons. Neurons receive and conduct electrical impulses.

neurotransmission: The process by which neurons transmit messages to other neurons, muscle cells, or gland cells.

neurotransmitter: A chemical substance that transmits a nerve impulse across a synapse.

nicotine: An addictive chemical found in tobacco leaves.

nucleus accumbens: The part of the brain related to the limbic system that controls emotions.

phenotype: The externally or internally detectable characteristics of an organism that represent the influence of environment and genetic information (genotype).

polyacrylamide: A synthetic polymer, similar to that found in a soft contact lens, which is used to prepare gels for electrophoresis.

polymerase chain reaction (PCR): A laboratory technique that uses a heat-stable DNA polymerase to amplify a short DNA sequence.

primer: A short sequence of RNA or DNA that binds to a single-stranded region of DNA and serves as a binding site for DNA polymerase.

reward pathway: A specialized network of neurons in the brain that produce and regulate pleasure associated with eating, drinking, and sex. These neurons use dopamine as a neurotransmitter.

ribosome: A structure made of protein and RNA that is located in the cytoplasm of a cell and is the site of protein synthesis.

running buffer: A salt-containing solution that is used during electrophoresis to conduct electricity.

Sequencher: A computer program that is used to analyze DNA sequences.

single nucleotide polymorphism (SNP): A common single-base-pair variation in a DNA sequence.

streptavidin: A small bacterial protein. In the context of DNA visualization, streptavidin specifically binds to biotin, helping attach the enzyme alkaline phosphatase to the DNA fragments being visualized.

synapse: The tiny space between two nerve cells or between a nerve cell and a muscle or gland cell.

transcription: The process of making an RNA copy of a gene.

translation: The process of making a protein by joining amino acids together in the order specified by an mRNA.

ventral tegmental area (VTA): A region of the brain involved in the reward pathway; located near the top of the brain stem.

Appendix I:

Automated DNA Sequencing Protocol and Reagents

Background

Please review the section Chain Termination DNA Sequencing.

In principle, manual and automated sequencing are very similar and follow the same basic steps:

1. Synthesis of DNA fragments that are partial copies of the DNA piece being sequenced
2. Separation of DNA fragments according to size by gel electrophoresis
3. Detection of DNA fragments

As you'll see below, automated sequencing uses a fluorescent label on the DNA so that it can be detected automatically, thus enabling high throughput sequencing of many long fragments at one time. Each step is outlined below.

Part A: DNA Sequencing by Thermal Cycling

Cycle sequencing is very similar to the polymerase chain reaction. Like PCR, cycle sequencing uses a heat-stable DNA polymerase that functions at a high temperature and is resistant to near boiling temperatures for many hours. This process has two main advantages: it requires much less DNA template than other techniques, and the high temperatures help to "melt out" secondary structure in the DNA template (like hairpins) that could otherwise inhibit the DNA polymerase. It is also much easier!

First, all of the following components needed for DNA synthesis are mixed in a tube:

- DNA template (the DNA molecule being sequenced)
- A DNA primer
- The four deoxynucleotides (A, C, G, and T)
- The four dideoxynucleotides (dideoxyA, dideoxyC, dideoxyG, and dideoxyT)
- Reaction buffer
- Heat-stable DNA polymerase

For each DNA template being sequenced, we need to prepare just one reaction. Each of the dideoxynucleotides has a different colored fluorescent label (A is green, C is blue, G is yellow, and T is red). This means that the DNA fragments synthesized during the sequencing reactions are color coded.

DNA synthesis is carried out by incubating the reaction at three different temperatures. DNA samples are placed into an automated heating block called a thermal cycler. The thermal cycler is programmed to quickly reach and hold

the samples at the desired temperatures for as many cycles as needed. First, the reactions are heated to 95°C to break apart hydrogen bonds (that is, the base pairing is disrupted). Then the DNA is cooled to 45–55°C to allow the primer to base pair to the DNA template. Finally, the reaction tube is raised to 70–72°C, the optimum temperature for the DNA polymerase to function. After 1 minute at 70–72°C, the reaction is cycled through the same three temperatures again, allowing another round of DNA synthesis to take place. Every time the cycle is repeated, a new DNA strand is made. The DNA templates are re-used during each cycle because the 95°C incubation releases the template strands from the newly synthesized DNA strands, making them available to be copied again. Typically, about 25 cycles are needed to ensure that enough DNA is synthesized to be analyzed. Following the last cycle, the reaction is cooled to 4°C and then stored at –20°C. Before being sequenced, the reaction is purified using ethanol precipitation which removes the unused nucleotides and polymerase molecules.

To simplify the sample preparation, the four deoxynucleotides and dideoxynucleotides are premixed with DNA polymerase and reaction buffer by the manufacturer. All you need to do is put some of the reaction mix in a color-coded tube and add the DNA template, a primer, and sterile water. A low concentration of DNA template can be used because it is recycled each time. The primer is often 50 times more concentrated than the template.

Parts B and C: DNA Separation and Detection

As in manual sequencing, the DNA fragments are separated by gel electrophoresis in a denaturing polymer. The main difference in automated sequencing is that the gels are placed inside glass capillary tubes. The arrangement of glass tubes (typically 96 are used) is called a capillary array. One sample is loaded into each glass tube. The capillary array is placed inside a light-proof container with a laser detector positioned at one end of the capillary gels. The detector is on a slider so that it can move back and forth across the width of the gel. The entire system is attached to a computer, which coordinates the loading of the reactions, running of the fragments, movement of the detector, and storing and processing of the data collected by the detector.

The polymer that separates the DNA fragments is automatically loaded into the glass tubes. Once the computer is programmed with the appropriate sample names, a loading arm takes some of each reaction and loads it into the corresponding capillary gel. After all the samples are loaded, the electrophoresis run is started.

The detection of the DNA sequence also differs from the method used during manual sequencing. In automated sequencing, the colored labels on the DNA fragments are detected by the laser detector at the end of the capillary gels. As each DNA fragment migrates past the laser beam, the detector determines the color of the fluorescent label on the fragment and sends this information to the computer. The detector samples each lane in rapid succession, sending information from all 96 capillaries simultaneously to the computer. This information is integrated and processed by a sequencing software program. The sequence data

for each DNA sample is presented as a four-color chromatograph.

Procedure

Check off each step as you complete it.

Pre-experiment setup

___ Write down your full DNA template name in your lab notebook.

2A6E1P1S1_F1SC_0504

E1 = exon 1
P1 = primer pair set 1
S1 = subject 1 (the individual whose DNA is being sequenced)
F = forward reaction (use R if you did the reverse reaction)
1 = lab group number (assigned by your teacher)
SC = school code
0504 = date (month and year)

Figure A1. The data file name. For two-letter school designations, see DNA Sequencing—Experimental Procedure, Day 4: Data Analysis on page 61. The DNA templates you receive are labeled with the information up to the first underscore. You will need to provide the additional information.

Part A

___ 1. Label the top of a purple or orange 0.5 mL tube with your DNA subject number and group number. Label orange tubes "F"—you will be using the forward sequencing primer. Label purple tubes with "R" and use the reverse sequencing primer.

___ 2. Put the following in the appropriate colored tube in the order listed:

Sterile water	4 µL
Sequencing Primer (Forward or Reverse)	1 µL
BDT Reaction Mix (buffer, deoxynucleotides, dideoxynucleotides, DNA polymerase)	4 µL
DNA Template	1 µL
TOTAL	**10 µL**

Figure A2. Setting Up Sequencing Reactions

___ 3. Close the lids and spin in the centrifuge for a few seconds.

___4. Place your reaction in the thermal cycler. Note the position of your tube in case the label wipes off. When everyone's sample is in place, close the lid.

Cycle sequence, use the following program:

Hold at 95°C	3 minutes	
95°C	10 seconds	
45°C	10 seconds	25 cycles
70°C	1 minute	
Hold at 4°C	indefinitely	

Setting Up and Using the Thermal Cycler
(MJ Research Model PTC-100/60)

A. Plug the cord from the Hot Bonnet to the back of the thermal cycler (cylindrical plug). Plug the thermal cycler into an electrical outlet and turn on power (switch on back).

B. On the display you will see SELF TEST for a few seconds and then

	RUN PROGRAM	ENTER PROGRAM

C. Select RUN PROGRAM and press PROCEED.

D. The screen will say:

	RUN HS1?	Press PROCEED	to indicate that this is the program you want.

E. The screen will ask whether you want to use the Hot Bonnet by asking

	ENABLE	DISABLE	HEATED LID?
Select	ENABLE	and press PROCEED.	

F. The program will now run. At the end of the run, or if you need to stop the run at any time, press STOP or CANCEL and then PROCEED.

Consult the manual if you have any other questions.

___ 5. Collect the sample with a quick spin. Check to make sure that each tube is still clearly labeled.

___ 6. Place samples from the class in a rack and put in a –20°C freezer until your visit to the University of Washington Genome Center or the sequencing center at your partner site.

___ 7. Prior to loading your samples, the StarNet staff will precipitate them to remove the remaining nucleotides and polymerase molecules.

___ 8. Place plate in sequencer and program computer.