

Chapter 3:

Molecular Biology Problems

Molecular Biology Problems

If you were a molecular biologist, you would focus on biological molecules like DNA, RNA, and proteins. Although generally true, your work would overlap with other areas like genetics and biochemistry. In this chapter, we have given you problems that will allow you to explore the structure and function of DNA and RNA and how proteins are elaborated in the cell.

The first problems examine some of the most important experiments that led to the conclusion that DNA is the genetic material. They are good examples of the history of science as well as opportunities to analyze real data. Before attempting the question in this section, refer to your textbook or lecture material.

The first diagnostic question for this chapter is found in section 2. Work through the diagnostic question on your own and then look at our approach to solving it. If any of the terms are unfamiliar, consult the appropriate chapter in your textbook.

(1) PROBLEMS EXPLORING CLASSIC EXPERIMENTS

Key words: "Griffith," "*Streptococcus pneumoniae*," "Transforming Principle," "Avery," "Pneumococcus," "Hershey," and "Chase."

Note that this section does not have a diagnostic problem; you should consult your textbook for further information if you have difficulty working these problems.

(1.1) Frederick Griffith showed that a component of dead bacterial cells could confer new properties to live bacteria of the same species. The property in question was the presence of a surface polysaccharide capsule of the bacterium *Streptococcus pneumoniae*, the bacterium that causes pneumonia.

Two types of these bacteria are found, based on two main types of surface polysaccharide:

- 1) R (rough) cells have no polysaccharide capsule. These bacteria do not cause disease. They are nonvirulent.
- 2) S (smooth) cells have a polysaccharide capsule. These bacteria cause disease and are virulent.

The S (smooth) cells fall into several different subtypes based upon the polysaccharide capsule. These are designated S_I through S_{XXIII}.

You can isolate mutants of all these S strains that no longer make a capsule and are no longer virulent. The lack of a polysaccharide capsule makes one R strain indistinguishable from another. However, R strains are also designated R_(I) through R_(XXIII), depending upon their origin. An R strain derived from S_{III} is designated R_(III).

The central experiment was:

Experiment: Inject a mixture of $R_{(II)}$ and heat-killed S_{III} into a mouse.

Results: The mouse dies of pneumonia, and virulent S_{III} bacteria can be isolated from the mouse.

Griffith et al. also did some control experiments to support experiment 1.

Negative Control: Inject $R_{(II)}$ or heat-killed S_{III} alone into a mouse.

Results: The mouse lives.

Positive Control: Inject live S_{III} alone into a mouse.

Results: The mouse dies of pneumonia, and only virulent S_{III} bacteria can be isolated from the mouse.

a) For each of the control experiments listed, explain:

- What alternative models does the result of this control experiment rule out? That is, complete the sentence, "Without this control result, you could argue that the mice died in experiment 1 because...."
- What would it have meant for their model if the results of this control experiment were the reverse of those expected (the mouse dies instead of living and vice versa), assuming that the results of experiment 1 were the same as above?

b) Another potential problem with their experiments was the possibility that, at a low frequency, $R_{(II)}$ can mutate back to the S_{II} from which it was derived. (Note: $R_{(II)}$ cannot revert to any other subtype of S .) Based on this, why was it essential that they use a mixture of $R_{(II)}$ and heat-killed S_{III} instead of a mixture of $R_{(II)}$ and heat-killed S_{II} ?

(1.2) The team of Avery, McCarty, and MacLeod was attempting to purify a substance from smooth (S) pneumococci which was capable of transforming rough (R) pneumococci into smooth; they called this substance the “transforming substance.”

a) They presumed that the transforming substance was genetic material (a.k.a. genes). Explain why they believed that this was so.

The authors were trying to distinguish between two models for the “transforming substance” (genes):

- (1) Genes are made of protein.
- (2) Genes are made of DNA.

It was their belief that model (2) was correct. They wanted to purify the transforming substance away from other cellular material and then determine whether it was pure DNA or a mixture of protein and DNA. They hoped to show that protein was either not present or not essential for the transforming substance to be able to transform R to S.

b) If they had found traces of protein in their preparations of the transforming substance (even if it was >99% DNA), this would have made it impossible to rule out model (1). Explain why this is so.

c) In their paper, they presented an analysis of the elemental composition of several of their purified preparations. In particular, they compared the ratio of nitrogen to phosphorus (N/P ratio) of their preparations, with the N/P ratio predicted based on the structure of DNA.

i) Using the structure of DNA from your text, the atomic weight of N = 15 and the atomic weight of P = 31, show that the ratio for the DNA:

$$\frac{\text{total mass of N}}{\text{total mass of P}} = 1.69.$$

ii) If their preparations were contaminated with protein, would you expect the ratio of the preparations to be higher or lower than 1.69? Explain your reasoning.

These are their actual data (all of these preparations could transform R to S):

<u>Preparation #</u>	<u>N/P ratio</u>
37	1.66
38B	1.75
42	1.69
44	1.58

iii) What conclusions would you draw from these data, and what would be your reservations about these conclusions?

d) In another series of experiments, preparations were treated with various enzymes of known function. They wanted to determine whether these enzymes were capable of destroying the transforming ability of the preparations. They treated their preparations with the following enzymes alone or in combination:

- Trypsin: an enzyme that breaks proteins down into amino acids.
- Chymotrypsin: another enzyme that breaks proteins down into amino acids.

None of these treatments had any effect on the transforming ability of their preparations.

i) Why do these data support model (2)?

ii) Why are these data, on their own, not completely conclusive? If you believed in model (1), how would you argue that these data are consistent with model (1)?

(1.3) Hershey and Chase provided strong evidence that DNA, and not protein, is the genetic coding material of the cell through their experiments involving differential partitioning of ^{32}P -labeled DNA and ^{35}S -labeled protein of bacteriophage T2.

a) In their studies, there were two crucial experiments.

Experiment 1: Bacteriophage were labeled with ^{32}P , which is incorporated only into DNA.

Experiment 2: Bacteriophage were labeled with ^{35}S , which is incorporated only into protein.

The bacteriophage were allowed to attach to the bacterial cells and transmit the genetic material. The bacteria and the bacteriophage were then separated, and the amount of radioisotope in each was measured. For each scenario below, describe how much radioactivity you would expect to find in the phage as opposed to the bacteria (for example: a lot, some of it, little, or none). Explain your reasoning briefly.

i) If the phage contained both DNA and protein but the genetic material injected was protein and the DNA remains in the phage head.

ii) If the genetic material were a mixture of mostly DNA and a little protein.

iii) If the genetic material were protein that is carried on the DNA and the DNA is only a scaffold for carrying the protein genetic material.

b) Actually, the data were not as unambiguous as described in most textbooks. (For an interesting discussion of how this experiment has changed in the telling and retelling, see "How history has blended," *Nature* **249**:803-805, 1974.) They performed the experiment several times and got the following results:

(1) Between 15% and 35% of the ^{32}P was found in the supernatant.

(2) Between 18% and 25% of the ^{35}S was found in the pellet.

Assume that their model is correct (the phage injects DNA and not protein).

i) What could have caused the presence of the ^{32}P in the supernatant? How crucial for their model is it that this number be 0? Why?

ii) What could have caused the presence of some of the ^{35}S in the pellet? How crucial for their model is it that this number be 0%? Why?

c) Based on the above data, which of the models (i, ii, or iii) in part (a) can you rule out? Explain your reasoning.

(1.4) On a mission to a new solar system, you discover an alien virus that contains nucleic acids, proteins, and lipids. You also find that this virus can infect *E. coli* cells, making it easy to study in the laboratory.

a) You grow this virus with one of the following radioisotopes: ^{32}P , ^3H , or ^{35}S .

- Which of the viral macromolecules will be labeled with ^{32}P ?
- Which of the viral macromolecules will be labeled with ^3H ?
- Which of the viral macromolecules will be labeled with ^{35}S ?

b) You analyze the nucleic acid and find the following:

Percentage of each base:				
A	G	T	C	U
27.6	23.2	28.1	24.0	0.8

What nucleic acid is the virus carrying? How do you know this?

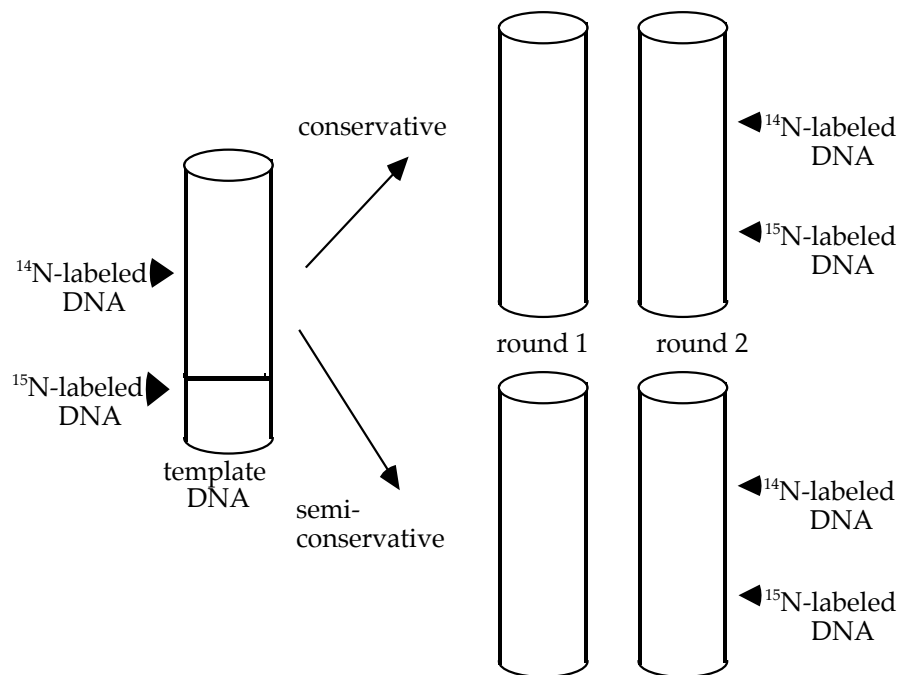
c) Because this is an alien virus, you want to determine which of the macromolecules (nucleic acids, proteins, and lipids) is the hereditary material. Explain how you would do this.

d) You examine DNA replication to determine whether it is similar to DNA replication on Earth. You begin by constructing an in vitro system for DNA replication.

i) Assuming that the process of alien DNA replication is similar to that seen on Earth, what four components would you include in your system?

ii) Prior to replication, all the template DNA is labeled with ^{15}N . Where is nitrogen found in DNA?

iii) You repeat the Meselson-Stahl experiments. On the diagram below, draw the results expected at each round for both conservative and semiconservative replication.



(2) PROBLEMS EXPLORING THE STRUCTURE OF DNA AND RNA

Although biology is often a science of exceptions, there are several important concepts that apply to the nucleic acids, DNA and RNA. These statements hold in the vast majority of cases.

Directionality – All strands have a direction. This is specified in terms of the 5' and 3' carbons in the sugar backbone.

Antiparallel – In order to form proper base pairs, the two strands must run 5' to 3' in opposite directions.

Base pairing – In DNA, A always pairs only with T. In RNA, A pairs with U. G always pairs only with C.

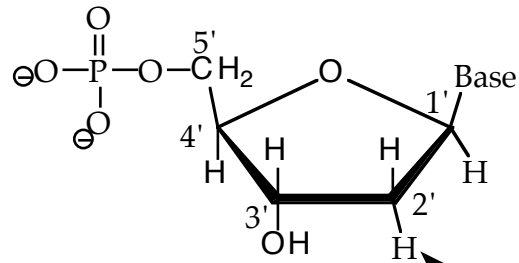
Polymerization – New nucleotides are always added only to the 3' end of a chain. Thus, polymerization always occurs in a 5' to 3' direction.

DNA and RNA Structure:

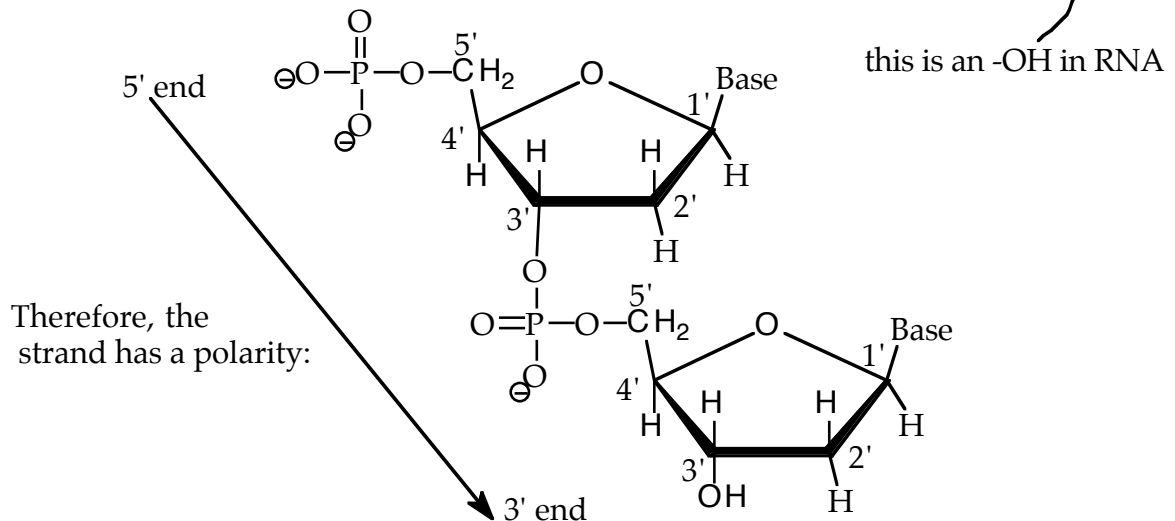
1) Building blocks of DNA are nucleotides:

Base =

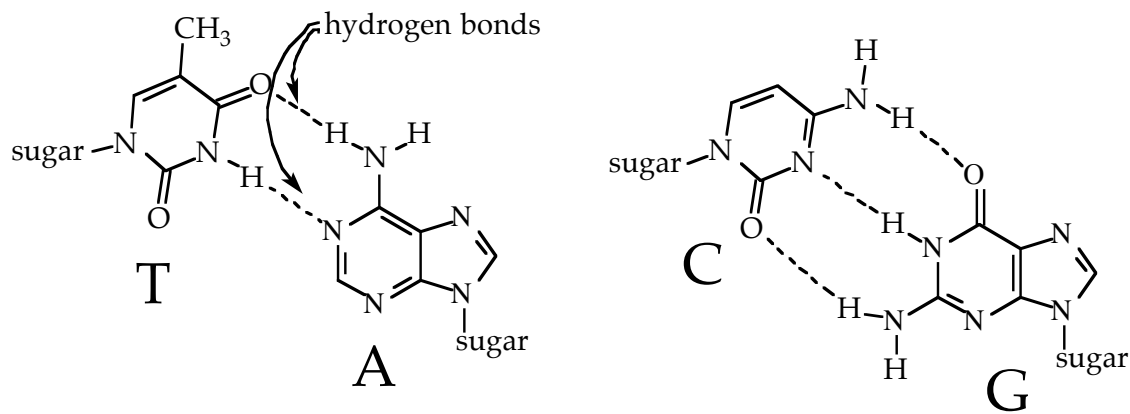
Adenine (A)	}	purines
Guanine (G)		
Cytosine (C)	}	pyrimidines
Thymine (T)		



2) The nucleotides in DNA are joined by 3' – 5' phosphodiester bonds.



- 3) The DNA has two strands held together by hydrogen bonds between the bases on opposite strands.



- 4) The two strands have opposite polarity.
 5) The two strands are twisted into a right-handed double-helix.
 6) RNA is single-stranded, but it can fold on itself to create regions where hydrogen bonds form between the bases.

Diagnostic Question:

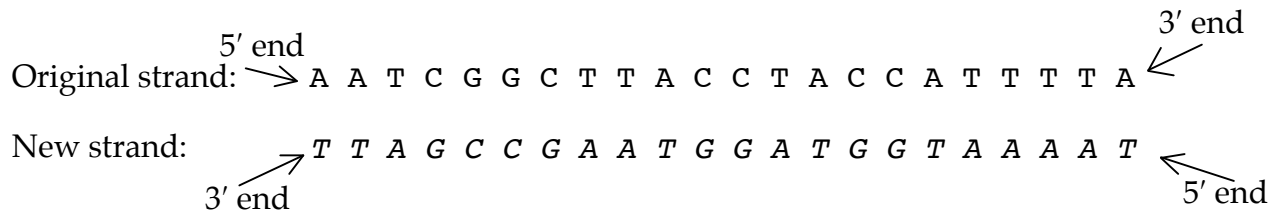
You are given the sequence of one strand of a DNA molecule:

5' end
 → A A T C G G C T T A C C T A C C A T T T T A ← 3' end

- a) Directly below the sequence above, give the sequence of the second strand of DNA. Label all 3' and 5' ends.
- b) What chemical group is usually found on the 3' end?
- c) What chemical group is usually found on the 5' end?
- d) What holds the two strands to each other?
- e) A new strand of DNA is made in what direction?

Answer to Diagnostic Question:

a)



b) What chemical group is usually found on the 3' end? *There is usually a hydroxyl at the 3' end.*

c) What chemical group is usually found on the 5' end? *There is usually a phosphate group at the 5' end.*

d) What holds the two strands to each other? *Hydrogen bonds hold the two strands together.*

e) A new strand of DNA is made in what direction? *A new strand of DNA is made in the 5' to 3' direction.*

(C1) Computer Activity 1. This activity will guide you through the three-dimensional structure of DNA.

Access “Molecules in 3-d” at this site <http://intro.bio.umb.edu/MOOC/jsMol/> and click on the link for this problem “Molecular Bio C1”.

This software has many useful features, so we will take some time now to describe its use in detail. This software allows you to get information from the image in several ways:

- Rotating the molecule: This is the best way to get an idea of the molecule’s three-dimensional structure. You can click and drag on any part of the molecule and it will rotate as though you had grabbed it.
- Zooming in or out: This helps to get close-up or “big-picture” views of the molecule. Hold the shift key down while dragging the cursor up (to zoom out) or down (to zoom in) the image.
- Identifying the atom you are looking at: You can find information on the atoms in the molecule in one of two ways:
 - By clicking on an atom and looking at the lower left of the “Molecules in 3-dimensions” window. A small line of text will appear there with information on the atom you just clicked.
 - By putting the cursor over the atom you are interested in and waiting a few seconds for the information to pop up. The program will then display information on the atom in a little pop-up window. The information in the pop-up window is more detailed than the above but rather cryptic. Try putting the cursor over some other atoms to see what you get. Note that this does not always work, especially on Macintosh computers.

In addition to the above, atoms are also identified by their color. The color scheme is shown to the right of the molecule images.

Here are a few important notes about the way these molecules are displayed in this problem:

- The atoms are colored according to the scheme at the right of the window.
- Only covalent bonds are shown.
- All covalent bonds are shown as single lines; single, double, and triple bonds are all shown identically. You have to figure out the bond type based on your knowledge of covalent bonding and DNA structure.
- Hydrogen atoms are not shown in these structures. This is because these are actual protein structures determined by X-ray crystallography. Hydrogen atoms are not visible in X-ray crystallograms and are therefore not shown in these structures. You have to infer the hydrogens yourself based on your knowledge of covalent bonds and DNA structures.

a) Click the button marked “Load First DNA Molecule.” You will see a black window with a DNA molecule shown in “spacefill” mode where atoms are shown as solid spheres at their actual sizes. You can click on the “Show atoms as ball and stick” button to change the representation to “ball and stick” where atoms are shown as balls and the covalent bonds connecting them are shown as rods. You will find it useful to switch back and forth between the two views. Note that you may sometimes need to click this button three times to get the view to change.

- The ball and stick view shows covalent bonds as rods and is most useful for determining which atoms are *covalently bonded* to each other.
- The spacefill view shows atoms as joined spheres of their approximate actual size in the molecule. It is most useful for determining which atoms are *closest together*. Because it does not show covalent bonds, it can sometimes be hard to figure out which atoms are covalently bonded.

i) Click on the “Show the two strands” button and select the “spacefill” view. In this view, the atoms are colored as follows:

- One sugar-phosphate **backbone** strand is **pink**.
- The other sugar-phosphate **backbone** strand is **light yellow**.
- The **5' carbons** at the ends of the two backbone strands are **purple**.
- The **3' carbons** at the ends of the two backbone strands are **white**.
- The **bases** are **green**.

The two sugar-phosphate backbone strands run next to each other. Based on the structure, are they parallel (both run 5' to 3' in the same direction) or antiparallel (run 5' to 3' in opposite directions)?

ii) Switch to a “ball and stick” view. Based on this view, are there any covalent bonds between the bases on the two different strands?

iii) Click on the “Color-code bases and the two strands” button. This is the same as (i) and (ii), except the bases are colored to identify which type they are.

- **Adenine** is **yellow**.
- **Thymine** is **red**.
- **Cytosine** is **blue**.
- **Guanine** is **green**.

Based on this view, which bases pair with which? By clicking on individual atoms, you can identify the different bases as you click along the chain. Using this technique, determine the sequence of as much of both DNA strands as you can.

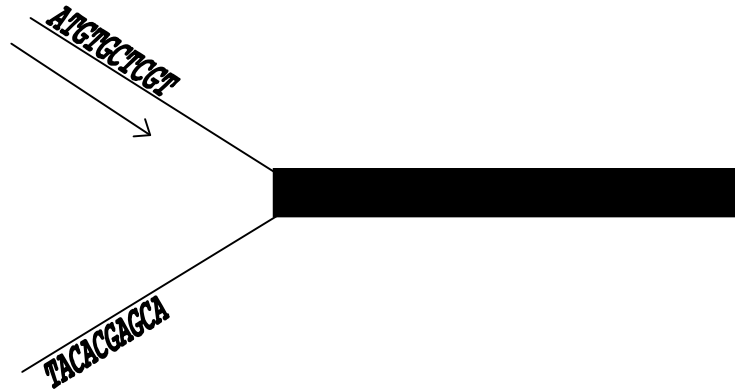
b) Click the “Load Second DNA Molecule” button. You will see a short (3-base-pair) double-stranded DNA. This illustrates the details of DNA structure. You should switch to the “ball and stick” view to help you see this more clearly. The atoms are colored using the color scheme on the right of the window **except**:

- **5' carbons** are **purple**.
- **3' carbons** are **white**.

i) Look at the 5' and 3' carbons and determine the 5' to 3' direction of each strand. Note that every nucleotide contains both a 5' and a 3' carbon. If you look at the strands, you will see that, in one strand, the 5' carbons in each sugar come first, while in the other it is reversed.

ii) Determine the sequence of both strands of the short DNA molecule shown. The “spacefill” view may be best for this. Hint: you don't have to look at every atom in the bases: the bases are colored as described. Also, you can click on atoms in each base and they will be identified in the information line at the bottom of the window.

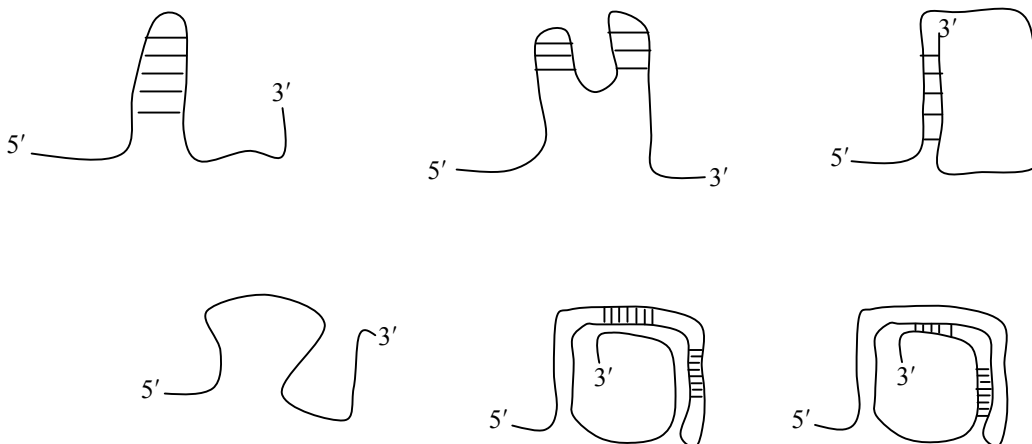
(2.2) Consider the following DNA segment that has begun to unwind for replication. The arrow represents the first nucleotides of the newly formed DNA. The arrow point marks the site where a new nucleotide will be added.



- What is the sequence of the newly formed DNA?
- Label all 5' and 3' ends.
- Draw an arrow on the bottom strand such that the arrow point marks the site where a new nucleotide will be added.

(2.3) Look at the following schematics of RNA molecules. The hydrogen bonds between the bases are indicated by the lines.

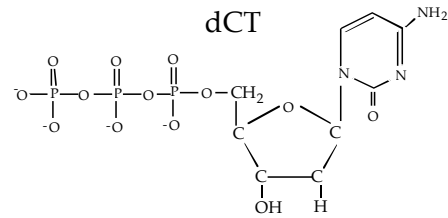
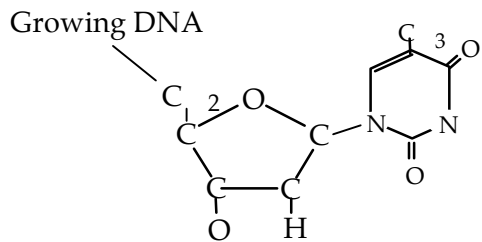
- Given that base pairing requires the strands to be antiparallel, circle the RNA molecules that could form.



- If the sequence in a region were 5'... ACGGACGC...3', what would be the sequence of the DNA that is hydrogen bonded to it?

(2.4) The next nucleotide to be added to a growing DNA strand is dCTP (shown).

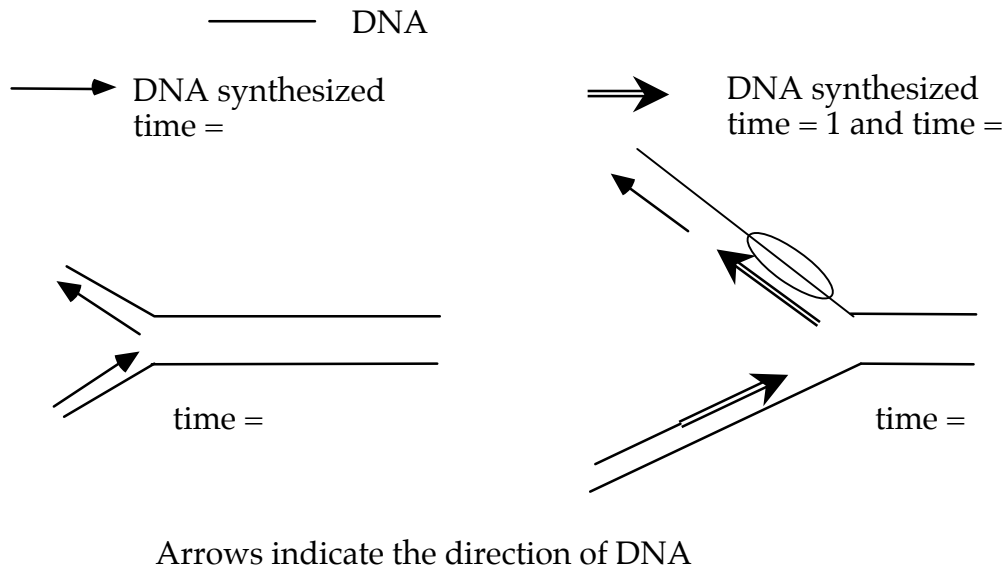
- Circle the part of the growing DNA chain to which the next base is attached.
- Circle the part of the dCTP that is incorporated into the growing DNA chain.



(3) DNA REPLICATION

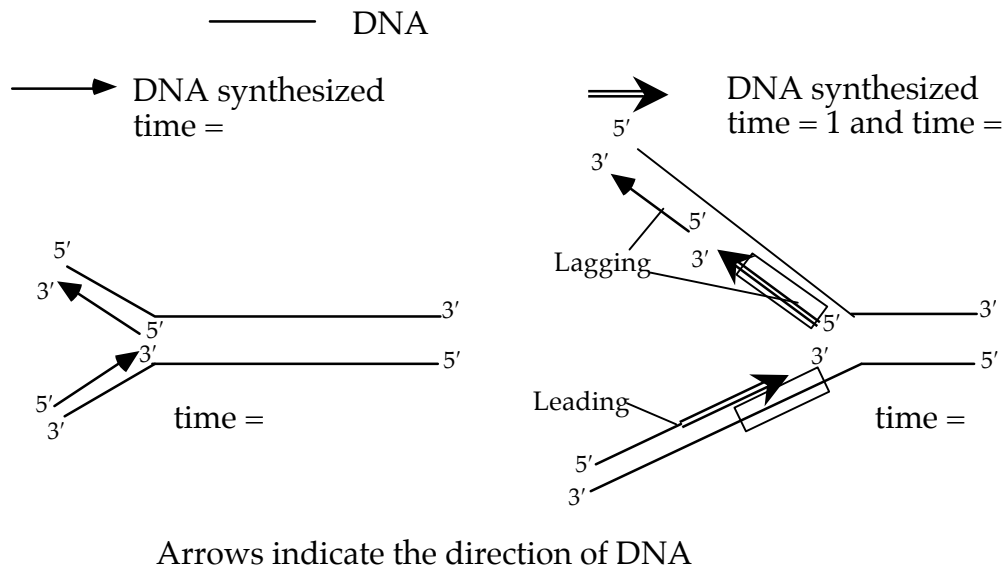
Diagnostic Question:

Shown below are two sequential close-up views of the same replication fork.



- Label the 5' and 3' ends of all the DNA molecules shown.
- The sequence within the circled region on the top template strand was as follows:
5' ... ATTCCG ...3'
 - Give the sequence of the complementary DNA.
 - Box the area where this complementary sequence is found.
- As the two template strands are pulled apart, synthesis of new DNA using the bottom template strand is continuous. Why is synthesis of new DNA using the top strand discontinuous?
- Label the leading strand and the lagging strand.

Answer to Diagnostic Question:



The answer to this question is really a restatement of some of the concepts given at the beginning of this section.

- **Directionality** – All DNA strands have a direction. This is specified in terms of the 5' and 3' carbons in the sugar backbone.
- **Antiparallel** – In order to form proper base pairs, the two strands must run 5' to 3' in opposite directions.
- **Base pairing** – A always pairs only with T. G always pairs only with C.
- **Polymerization** – New nucleotides are always added only to the 3' end of a chain. Thus, polymerization always occurs in a 5' to 3' direction.

a) See above.

b) The sequence of the complementary DNA would be: 3'...TAAGGC...5'. This sequence is boxed in two places on the diagram above.

c) DNA replication requires a template, and new nucleotides are added only to the 3' end of a chain. As the two original strands are pulled apart, a template becomes available and is copied such that the new DNA is made in the 5' to 3' direction. At time = 2, an additional template is available and is copied such that the new DNA is made in the 5' to 3' direction. The directional constraints of DNA replication cause the top template strand to be copied in a discontinuous fashion.

d) See above.

(3.1) DNA replication involves many different enzymatic activities. Match each enzyme activity listed below with the function(s) that it has in the replication process. The first one is done for you.

Enzyme Activity	Function
Topoisomerase	k
Primase (synthesizes primer)	
DNA polymerase to elongate new DNA strand	
Helicase to unwind DNA	
DNA polymerase to replace RNA with DNA	
Processivity factor	

Choose from:

- a) 3' \Rightarrow 5' growth of new DNA strand
- b) 5' \Rightarrow 3' growth of new DNA strand
- c) 3' \Rightarrow 5' exonuclease
- d) 5' \Rightarrow 3' exonuclease
- e) Makes RNA primer complementary to the lagging strand
- f) Makes RNA primer complementary to the leading strand
- g) Makes peptide bonds
- h) Separates the two DNA strands
- i) Maintains DNA polymerase on template
- j) Provides 3'-hydroxyl for initiation of DNA polymerization
- k) Untangles super-coiled DNA

(3.2) The following diagram shows a replication bubble within a DNA strand. A primer for DNA replication is 5'-GUACGUUG-3'.

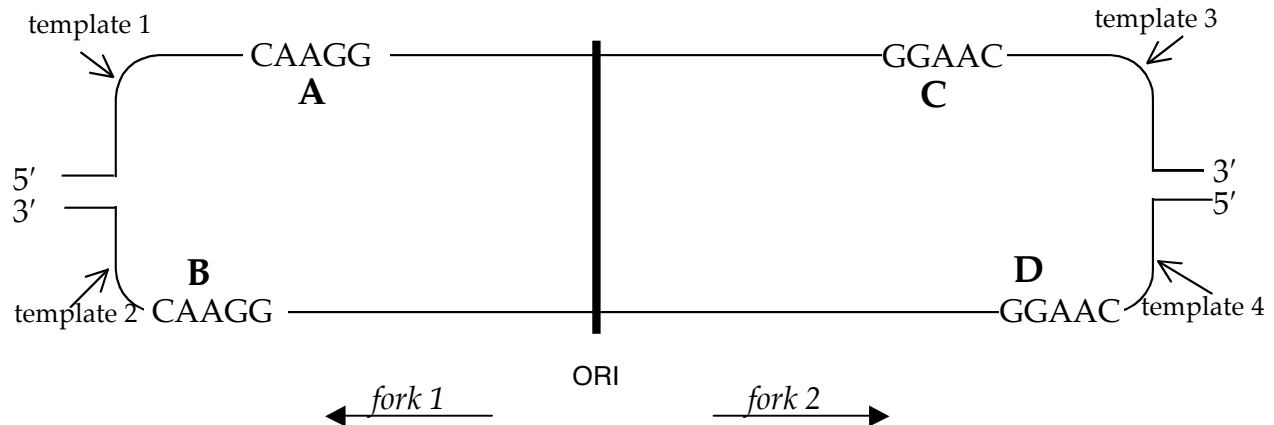
a) Draw the primer where it would anneal in the replication bubble and indicate the direction of replication by an arrow.



b) If the replication fork moves to the left, will this primer be used to create the leading strand or the lagging strand? Please explain your answer.

c) If you answered leading strand, explain why replication is continuous. However, if you answered lagging strand, explain why replication is discontinuous.

(3.3) Shown is a representation of an origin of replication. Synthesis of new DNA occurs on both strands and in both directions.



a) For the following, use sites A and B with respect to fork 1 and sites C and D with respect to fork 2.

i) On which strand(s) will replication be continuous?

template 1

template 2

template 3

template 4

ii) To which site or sites (A, B, C, or D) can the primer 5'-GUUCC-3' bind to initiate replication?

iii) When DNA ligase is inhibited, it differentially affects the synthesis from the leading and the lagging strands. Explain which strand (leading or lagging) is more affected by the lack of DNA ligase and why.

(3.4) From this origin, replication is occurring on both strands and the two forks are moving away from each other.

■
origin of

a) Label the 3' and 5' ends of the five DNA strands shown. Indicate which strands are Okazaki fragments.

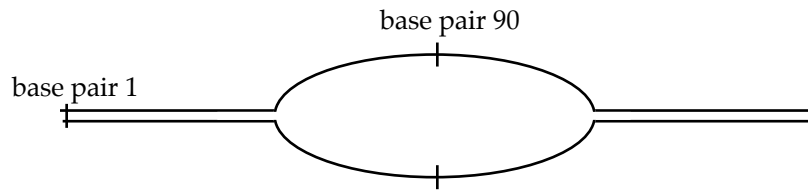
b) What enzyme is required at C in the diagram above?

c) To which site (**A** or **B** or **both**) can the primer, 5'-CAAGG-3' bind to initiate replication?

d) For each site chosen (in iii), what is the direction of elongation (**left** or **right**) of the daughter DNA strand?

e) For each site chosen (in iii), is DNA synthesis performed in a **continuous** or a **discontinuous** fashion relative to the nearest replication fork?

(3.5) Imagine that the DNA sequence adjacent to position 90 functions as an origin of replication.



The sequence in these regions is:

	80		100
5' ...	G C A T G C G T A C A A T A G T T C G A C		...3'
3' ...	C G T A C G C A T G T T A T C A A G C T G		...5'

- What is the sequence of the RNA primer that binds to the top strand at base pair positions 80-90? Indicate the 5' and 3' ends of your primer.
- Would DNA synthesis from the primer in (a) above be continuous or discontinuous?
- What is the sequence of the RNA primer that binds to the bottom strand at base pair positions 90-100? Indicate the 5' and 3' ends of your primer.
- Would DNA synthesis from the primer in (c) above be continuous or discontinuous?

(4) TRANSCRIPTION AND TRANSLATION

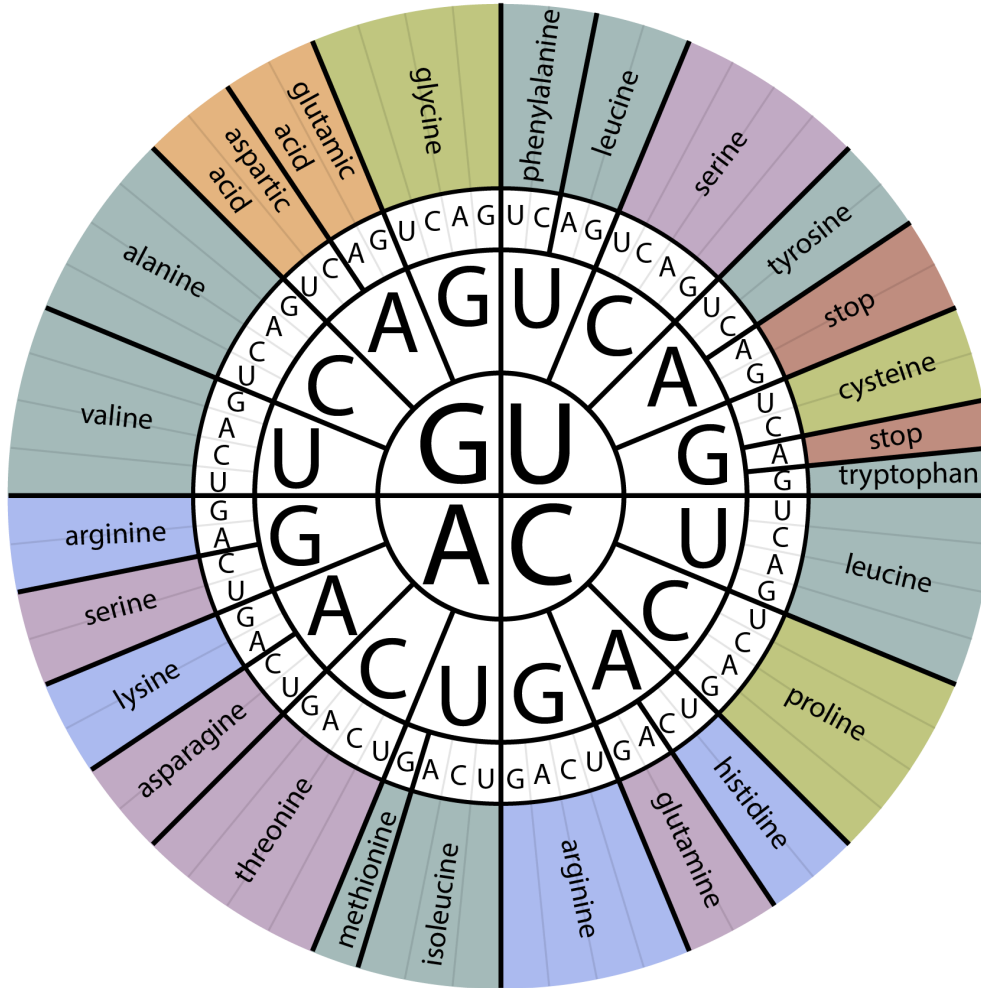
(4.1) Transcription and translation in prokaryotes

Some important notes:

- The promoter dictates where transcription starts. The promoter specifies the beginning point and the direction of transcription.
- The promoter is a site on the double-stranded DNA molecule where RNA polymerase binds.
- The concepts governing DNA synthesis also apply to RNA synthesis. Review directionality, base pairing, and polymerization.
- Translation starts with the AUG closest to the 5' end of the mRNA.* This AUG need not be at the 5' end of the mRNA, nor does it have to be a multiple of three nucleotides from the 5' end of the mRNA.
- Translation ends with the first in-frame stop codon, even if more nucleotides remain in the mRNA.
- Translation can restart at the next start codon following a stop codon.

* Please note that this is an oversimplification used in the context of these problems. In real genes, the start codon must be preceded by a particular sequence in order to be recognized as a start codon.

Here are two different ways to show the genetic code:



	U	C	A	G	
U	UUU phe (F) UUC phe (F) UUA leu (L) UUG leu (L)	UCU ser (S) UCC ser (S) UCA ser (S) UCG ser (S)	UAU tyr (Y) UAC tyr (Y) UAA STOP UAG STOP	UGU cys (C) UGC cys (C) UGA STOP UGG trp (W)	U C A G
C	CUU leu (L) CUC leu (L) CUA leu (L) CUG leu (L)	CCU pro (P) CCC pro (P) CCA pro (P) CCG pro (P)	CAU his (H) CAC his (H) CAA gln (Q) CAG gln (Q)	CGU arg (R) CGC arg (R) CGA arg (R) CGG arg (R)	U C A G
A	AUU ile (I) AUC ile (I) AUA ile (I) AUG met (M)	ACU thr (T) ACC thr (T) ACA thr (T) ACG thr (T)	AAU asn (N) AAC asn (N) AAA lys (K) AAG lys (K)	AGU ser (S) AGC ser (S) AGA arg (R) AGG arg (R)	U C A G
G	GUU val (V) GUC val (V) GUA val (V) GUG val (V)	GCU ala (A) GCC ala (A) GCA ala (A) GCG ala (A)	GAU asp (D) GAC asp (D) GAA glu (E) GAG glu (E)	GGU gly (G) GGC gly (G) GGA gly (G) GGG gly (G)	U C A G

Diagnostic Question:

The statement “DNA goes to RNA goes to protein” is used as a shorthand description of the flow of information in the cell.


a) “DNA goes to RNA” is meant to describe the process of _____.

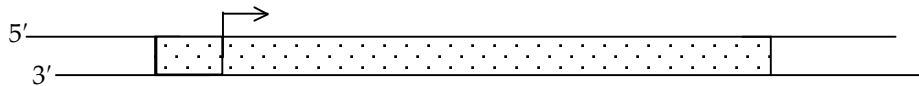
For this process to occur, we need an enzyme called _____ that uses DNA as a template to make _____.

b) “RNA goes to protein” is meant to describe the process of _____.

Three different types of RNA are used in this process. _____ RNA is used as the template to make a _____. _____ RNA is part of the complex called a _____ required for protein synthesis. _____ RNA is a small RNA molecule that acts as an adapter between RNA and protein.

c) Shown below is a schematic of a prokaryotic gene. The stippled area represents the region that is transcribed.

The  represents the promoter, and the direction of transcription is shown by the arrow.



a) Which strand of the DNA is being used as a template for transcription, the top or the bottom? Why?

b) On the diagram, indicate where the start codon would be. What sequence would it have?

c) On the diagram, indicate where the stop codon would be.

d) On the diagram, indicate where the transcription stop would be found.


Answer to Diagnostic Question:

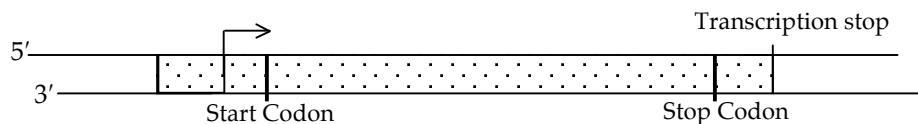
The statement “DNA goes to RNA goes to protein” is used as a shorthand description of the flow of information in the cell.

a) “DNA goes to RNA” is meant to describe the process of Transcription. For this process to occur, we need an enzyme called RNA polymerase that uses DNA as a template to make messenger RNA.

b) “RNA goes to protein” is meant to describe the process of Translation. Three different types of RNA are used in this process. Messenger RNA is used as the template to make a polypeptide (protein). Ribosomal RNA is part of the complex called a ribosome required for protein synthesis. Transfer RNA is a small RNA molecule that acts as an adapter between RNA and protein.

c) Shown below is a schematic of a prokaryotic gene. The stippled area represents the region that is transcribed.

The  represents the promoter, and the direction of transcription is shown by the arrow.



a) Which strand of the DNA is being used as a template for transcription, the top or the bottom?

The bottom strand is used as a template for transcription. RNA polymerase binds at the promoter and moves in the direction of the arrow. RNA can be made only in the 5' to 3' direction, antiparallel and complementary to the template. Thus, the template must be the bottom strand.

b) On the diagram, indicate where the start codon would be. What sequence would it have?

The exact position cannot be determined. In general, the start codon will be near the promoter and the sequence will be AUG.

c) On the diagram, indicate where the stop codon would be.

The exact position cannot be determined. In general, the stop codon will be near the transcription stop and could have the sequence UAA, UAG, UGA.

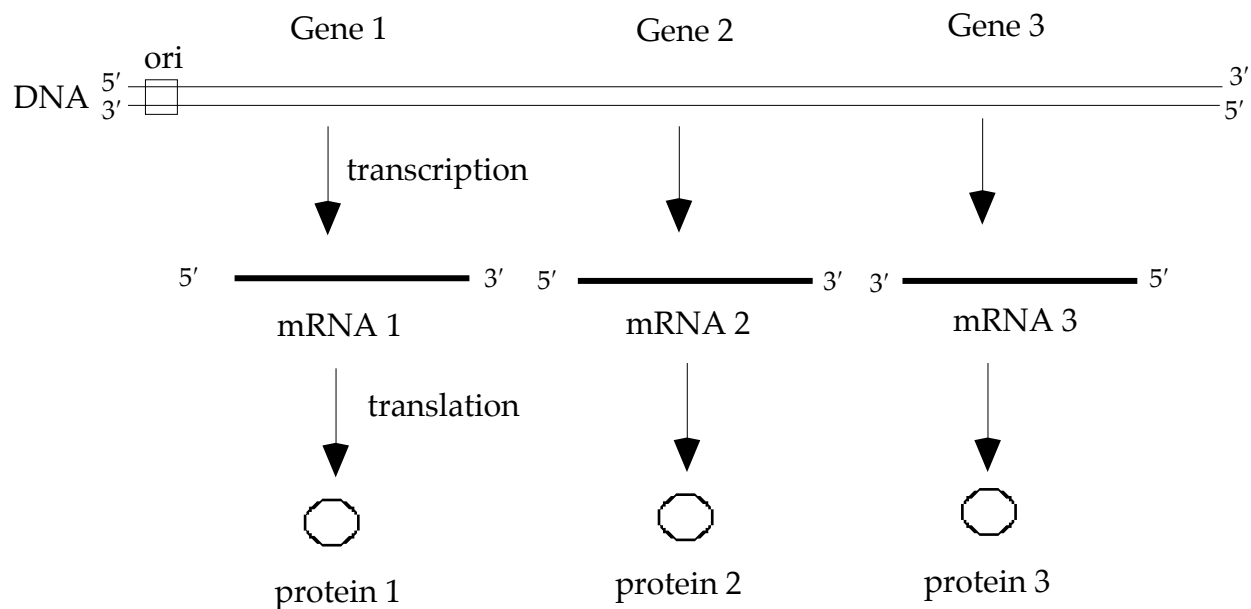
d) On the diagram, indicate where the transcription stop would be found.

Because you were told that the stippled area is the region transcribed, the exact position can be determined.

(4.1.1) Shown below are three genes (gene 1, gene 2, and gene 3) located on the same bacterial chromosome.

- a) Indicate where on the diagram you would find the following for each gene:
- Promoter (p1 for gene 1, p2 for gene 2, and p3 for gene 3)
 - Transcription termination site (tts1, tts2, and tts3)
 - Start codon (start1, start2, and start3)
 - Stop codon (stop1, stop2, and stop3)
 - Template strand (ts1, ts2, and ts3), the DNA strand that directs RNA synthesis

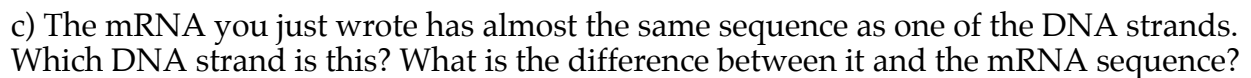
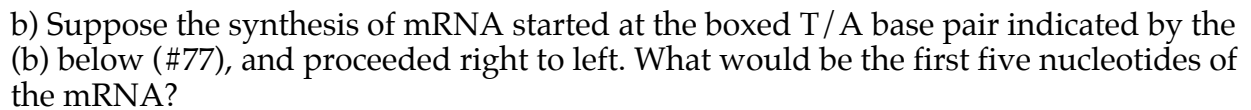
Be sure to indicate the component on the appropriate molecule (DNA or RNA). The origin of replication is shown as a sample: the replication complex recognizes the origin of replication as a double-stranded DNA sequence (not an RNA sequence), so the origin (ori) is shown as a small region of the DNA.



b) If a mutation inactivates the promoter for gene 2, do you expect protein 2 to be produced? Briefly explain.

c) After mutating the promoter for gene 2, you discover that protein 2 is still being made. Further analysis reveals that a spontaneous mutation has occurred in the transcription termination site in gene 1. How can you explain the presence of protein 2?

a) Synthesis of the mRNA starts at the boxed A/T base pair indicated by the (a) below (#11) and proceeds left to right on the sequence below. Write the sequence of the first 10 nucleotides of the resulting mRNA.



e) Does translation terminate at the UAA in the mRNA corresponding to the boxed bases at positions 48-50? Why or why not?

g) Mutations can add base pairs, delete base pairs, or change base pairs. The sequence shown below is the same as in part (a), except a G/C base pair has been added between 30 and 31. Note that the overall length of the DNA is now 81 base pairs due to the addition. This is a frame-shift mutation.



mutation.

5'-TGTGTGTGGAATTGTGAATGGATAAAATGTGACACAGGAAACAGCTAAGACCATGTTTTGACCAAGCTCGGAATTAAC-3'

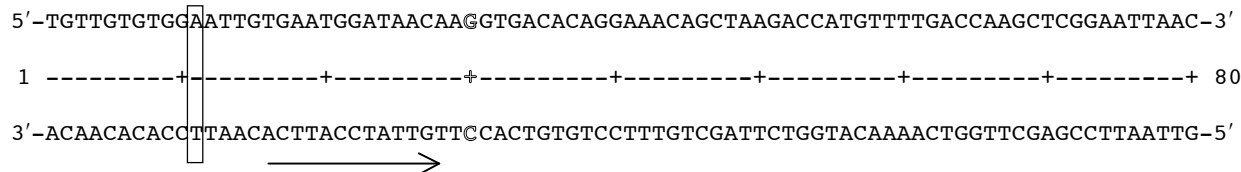
1 -----+-----+-----+-----+-----+-----+-----+----- 79

3'-ACAACACACCTTAACACTTACCTATTTTACACTGTGTCCTTTGTGCGATTCTGGTACAAAACCTGGTTCGAGCCTTAATTG-5'

C/G deleted from here

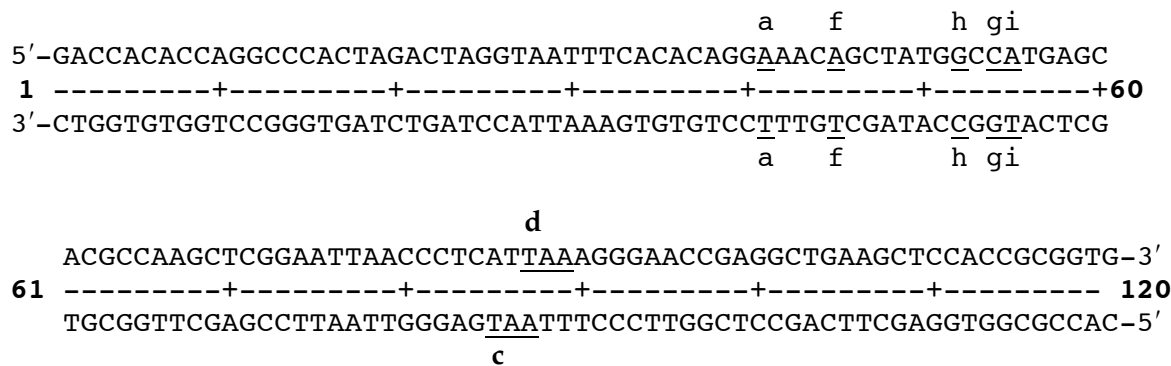
What will the sequence of the resulting protein be?

i) The sequence shown below is the same as in part (a), except the T/A base pair at 30 has been changed to a G/C base pair.



What will the sequence of the resulting protein be? Is this a silent mutation (no change in amino acid sequence), missense mutation (one amino acid changed to a different amino acid), or nonsense mutation (one amino acid changed to a stop codon)?

(4.1.3) The sequences of both strands of a DNA duplex are shown below. The top strand reads 5' to 3' left to right (1 to 120) and the bottom, complementary, strand reads 5' to 3' right to left (120 to 1). The letters above or below the underlined nucleotides are keyed to the particular questions that follow.



a) Assume that transcription (synthesis of mRNA) begins at the underlined A/T (written as top strand/bottom strand) at base pair 41 and proceeds to the right as shown on the diagram. What are the first 15 nucleotides of the resulting mRNA synthesized?

b) Given the mRNA synthesized in part (a), what are the first six amino acids of the resulting protein?

c) Given the mRNA synthesized in part (a), does protein translation end with the underlined TAA sequence on the bottom strand at nucleotides 85, 86, and 87? Why or why not?

d) Given the mRNA synthesized in part (a), does protein translation end with the underlined TAA sequence on the top strand at nucleotides 87, 88, and 89? Why or why not?

e) Given the mRNA synthesized in part (a), what would be the last three amino acids of the resulting protein?

f) Given the mRNA synthesized in part (a), what would the changes to the sequence of the resulting protein be if the underlined A/T base pair (written as top strand/bottom strand) at position 45 was deleted from the DNA sequence by mutation? Explain briefly.

g) Given the mRNA in part (a), what would the changes to the sequence of the resulting protein be if the underlined C/G base pair (written as top strand/bottom strand) at position 54 was changed to an A/T base pair? Explain briefly.

h) Given the mRNA synthesized in part (a), what would the changes to the sequence of the resulting protein be if the underlined G/C base pair (written as top strand/bottom strand) at position 52 was changed to a C/G base pair? Explain briefly.

i) Given the mRNA in part (a), what would the changes to the sequence of the resulting protein be if the underlined A/T base pair (written as top strand /bottom strand) at position 55 was deleted by mutation? Explain briefly.

(4.1.4) You are studying a protein and wish to determine the nucleotide sequence that encodes its first four amino acids. The sequence of the first four amino acids of the **normal** protein is as follows:

N – Met – Ser – Cys – Trp – – C

a) What possible mRNA sequences could encode the first four amino acids shown above? You need not write out all combinations; just write out all the codons that could encode each amino acid in their proper order.

HINT: there are six codons for serine (Ser). Be sure to indicate the 5' and 3' ends as appropriate.

b) You generate a mutant that produces an altered version of this protein due to a single base substitution in codon 2. The resulting protein sequence (differences from normal are **bold underlined**) is:

N – Met – **Gly** – Cys – Trp – – C

What does this tell you about the nucleotide sequence of the mRNA that encodes the serine (Ser) in the normal protein shown at the top of this page?

c) You generate a different mutant that produces an altered version of the protein shown at the top of this page. This mutation is due to a single base deletion of the first nucleotide of codon 2. The resulting protein sequence is shown below (differences from normal are **bold underlined**). Note that the mutation from part (b) is **not** present in this mutant.

N – Met – **Val** – **Ala** – **Gly** – – C

Based on this information, what is the nucleotide sequence of the mRNA that encodes the first four amino acids of the **normal** protein?

(4.1.5) You are studying a bacterium-like organism discovered on the wreck of an alien spacecraft. You discover that the creature has DNA and RNA like Earth organisms, except they have four different bases, which you name W, X, Y, and Z to differentiate them from the familiar A, T, G, and C. You also discover that the creature has transcription and translation processes almost identical to Earth cells – genes are double-stranded DNA, transcribed to mRNA, which is translated using tRNA and ribosomes.

You do notice one unusual feature: this creature's proteins are composed of only 14 amino acids, instead of the usual 20. You know that the organism's codons contain no more than three nucleotides. The amino acids used are as follows:

arginine	aspartic acid	glutamic acid	glutamine
isoleucine	leucine	lysine	methionine
phenylalanine	proline	serine	threonine
tryptophan	valine		

You decide to decipher the genetic code on this organism. First, you find conditions where you can react synthetic mRNAs with alien ribosomes, tRNA, and all the components needed to promote translation. You find that, under these unusual in vitro conditions, translation initiation can occur without a start codon and translation termination can occur without a stop codon. This means that initiation and termination can randomly occur on the mRNAs, leading to short oligopeptides.

First, you must find out how many nucleotides per codon there are.

For your first round of experiments, you add synthetic RNA to an alien translation reaction mix and sequence the protein that is synthesized. The results are as follows:

<u>Synthetic RNA added</u>	<u>Protein produced</u>
W-W-W...(W) _n	Met-Met-Met...(Met) _m
(X) _n	(Val) _m
(Y) _n	(Thr) _m
(Z) _n	(Leu) _m

a) Can you conclude anything about the number of nucleotides per codon from this information?

You perform a second round of experiments using the same procedure:

<u>Synthetic RNA added</u>	<u>Protein produced</u>
(WX) _n	a mixture of (Ile) _m and (Glu) _m
(WY) _n	(Lys) _m
(WZ) _n	a mixture of (Arg) _m and (Phe) _m
(XY) _n	a mixture of (Asp) _m and (Gln) _m
(XZ) _n	a mixture of (Pro) _m and (Ser) _m
(YZ) _n	(Trp) _m

b) From this information you are able to deduce the number of nucleotides per codon. What is this number and how did you deduce it?

You perform a final round of similar experiments:

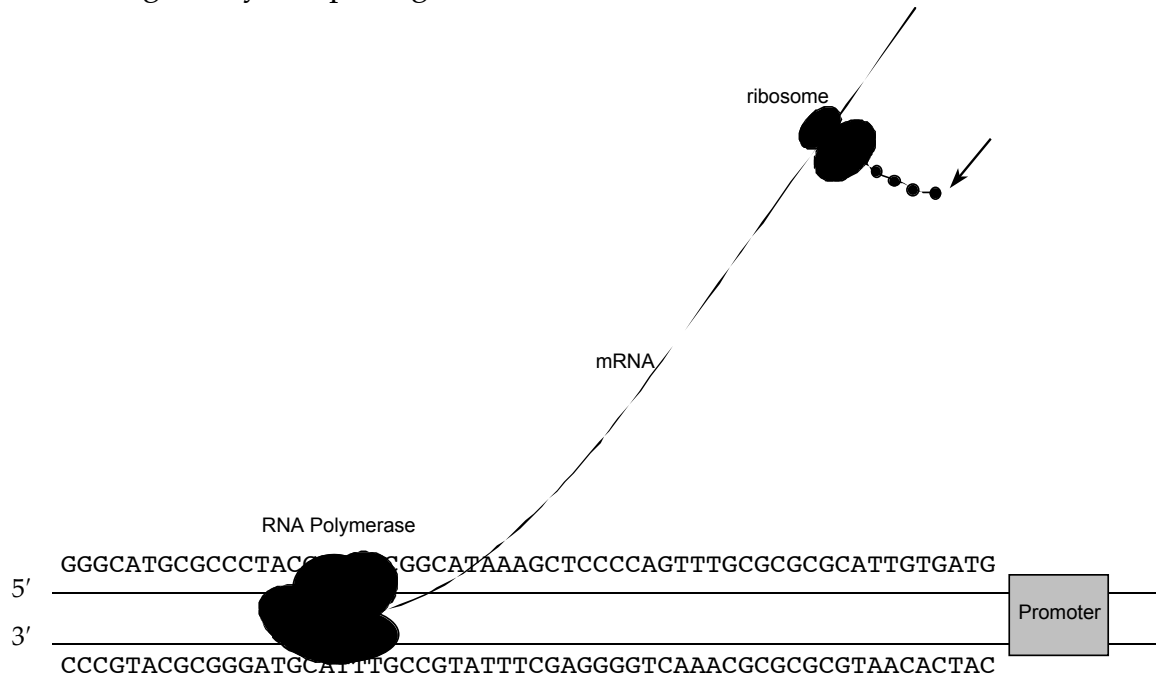
<u>Synthetic RNA added</u>	<u>Protein produced</u>
(WXY) _n	(Glu-Lys-Gln) _m [note: this is equivalent to (Gln-Glu-Lys) _m and (Lys-Gln-Glu) _m]
(WWX) _n	(Ile-Glu-Met) _m [same note]
(XYZ) _n	(Ser-Trp-Gln) _m [same note]
(WXZ) _n	(Pro-Glu-Arg) _m [same note]
(WZY) _n	Phe-Lys
(XWY) _n	Ile-Asp

Note that the last two are short peptides of only two amino acids.

c) From all the experiments above, you are able to deduce the genetic code in this organism. What is it? List all the codons and the amino acids they encode. Include stop codons.

(4.1.6) The following transcription and translation graphic is incomplete.

a) Finish the diagram by completing i to v.



- i) Label 5' and 3' on the mRNA.
 - ii) Label the arrow with either the N (amino) termini of the protein being made or the C (carboxyl) termini of the protein being made.
 - iii) As drawn, label the template strand for transcription.
 - iv) Box the three bases encoding the first amino acid of the protein being made.
 - v) Circle the part of the schematic where tRNAs would bind.
- b) Give the anticodon used in the tRNA encoding Trp. Be sure to label the 5' and 3' ends.
- c) Would a substitution within a codon for Trp always change the resulting protein sequence? Explain your answer.
- d) Would a substitution within a codon for Thr always change the resulting protein sequence? Explain your answer.

(4.2) Transcription, RNA processing, and translation in eukaryotes

Some important notes:

The processes of DNA replication, transcription, and translation are highly conserved in prokaryotes and eukaryotes. An important difference, which is reflected in the following problems, is the structure of eukaryotic genes.

Eukaryotic genes often have DNA that is transcribed but not translated. This DNA is termed intervening sequences or introns. The DNA sequences that are represented by the mRNA that is translated into protein are considered exons.

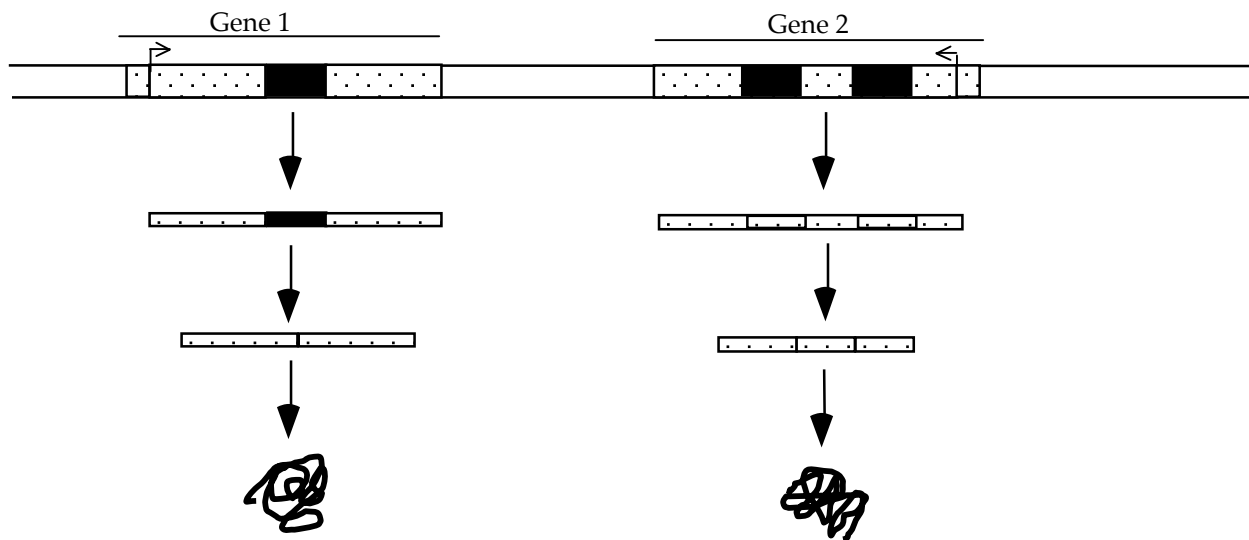
Diagnostic Question:

a) Shown below is a schematic of the production of a polypeptide. At the top is the chromosomal arrangement found in the cell; a schematic of the polypeptide is shown below it.

i) Label the process indicated by each arrow.

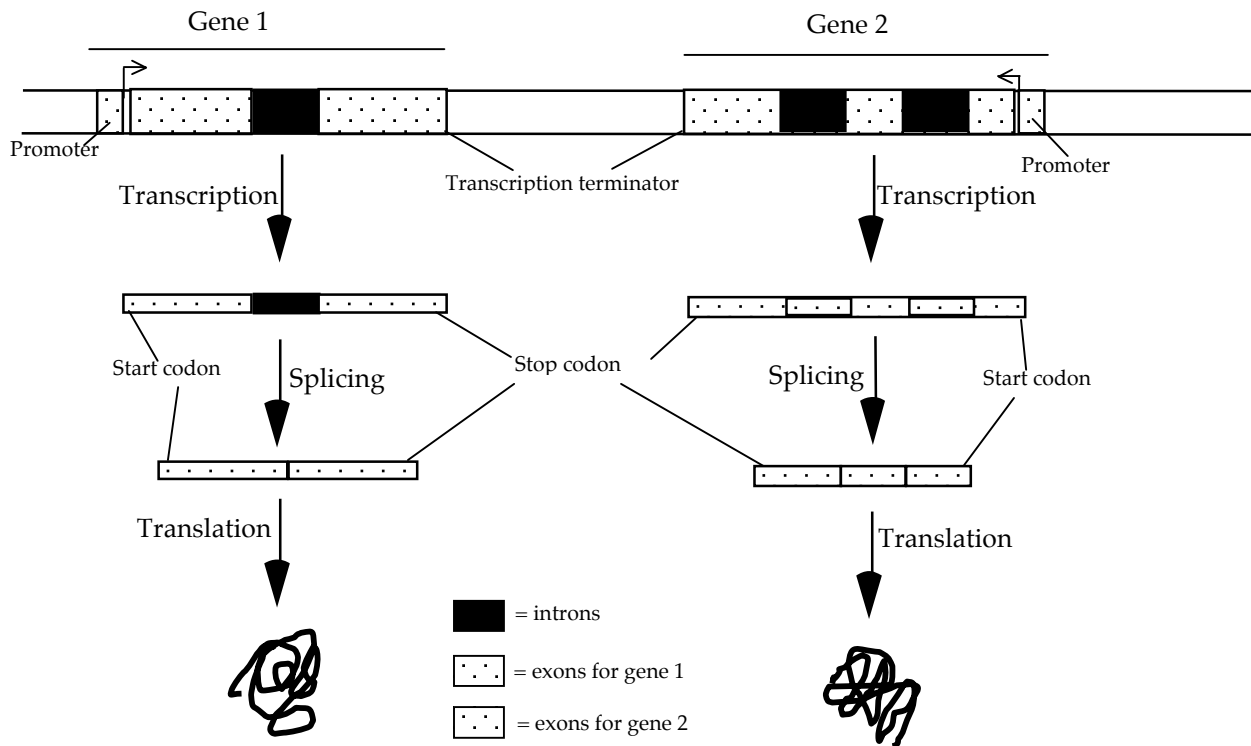
ii) Indicate on the diagram below where you would expect to find each of the following components:

- Promoter
- Transcription terminator
- Start codon
- Stop codon
- Introns
- Exons

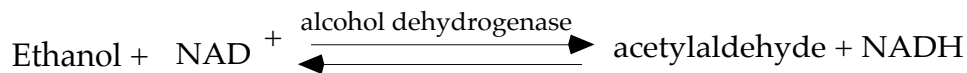


Answer to Diagnostic Question:

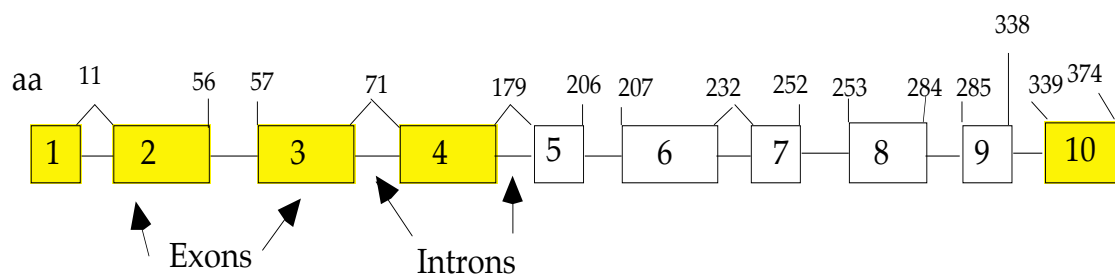
a) Shown below is a schematic of the production of a polypeptide. At the top is the chromosomal arrangement found in the cell; a schematic of the polypeptide is shown below it.



(4.2.1) The alcohol dehydrogenase enzyme functions in the breakdown of alcohol in the liver by converting it to acetaldehyde. The reaction is outlined below:



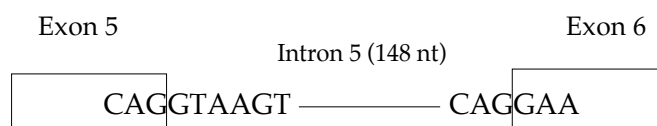
The enzyme binds the cofactor NAD^+ in order to carry out the oxidation of ethanol. The utilization of NAD^+ is crucial to the activity of the enzyme. The gene that encodes the 374-amino-acid enzyme is made up of 10 exons and 9 introns. Amino acid (aa) residues encoded by exons 5 to 9 are involved in binding the NAD^+ cofactor, while amino acid residues encoded by exons 1, 2, 3, 4, and 10 are involved in catalysis. A schematic diagram of the alcohol dehydrogenase gene is shown below.



Note that codons can be interrupted by introns and later formed when the intron is excised and two exons are joined together. As an example, see amino acid 11.

a) Using a format similar to the one given in the diagram above, draw the fully processed mRNA of the alcohol dehydrogenase gene. Indicate the 5' and 3' ends of the mRNA.

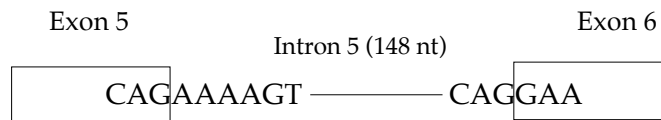
b) The DNA structure and partial codong-strand sequence of an exon/intron boundary from a region in the gene are shown below. (Note: not all 148 nucleotides in the intron are shown.)



i) Using the format given above, indicate the structure and partial sequence of the mRNA at this exon/intron boundary from this region of the gene before and after splicing of the mRNA.

ii) What are amino acids 206 and 207 in the alcohol dehydrogenase enzyme?

c) Certain individuals possess a defective alcohol dehydrogenase gene. This defective gene produces a mutant enzyme. The defect is due to a mutation at the 5' splice site in intron 5. The structure and partial sequence of this region in the DNA are shown below.



i) Using a format similar to the one used in part (b), indicate the structure of the entire spliced mRNA that encodes this mutant enzyme.

ii) What two possible effects can this mutation have on the structure of the protein?

(4.2.2) The figures on the next two pages present a small gene that was found in the eukaryotic fungus *Neurospora crassa*. This is intended as a real-world example of a eukaryotic gene. Note that most genes in eukaryotes are substantially larger.

```

5'-CGGTGAATAAATACGTCATGACGGTGCTGTCAGCATCATCGATAGGTAGGAGCGAACAACAACCTAACATCGGATTGCA
1  +-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
3'-GCCACTTATTTATGCAGTACTGCCACGACAGTCGTAGTAGCTATCCATCCTCGCTTGTGTTGTTGGATTGTAGCCTAACGT

      GGACCGCGGGGCGAGGATTGCTCCGGGCTGTTTCATGACTTGTTCAGGTGGGATGACTTGGATGGAAAAGTAGAAGGTCATG
81  +-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
      CCTGGCGCCCCGTCCTAACGAGGCCCGACAAAGTACTGAACAGTCCACCCTACTGAACCTACCTTTTCATCTTCCAGTAC

      GGGTGGCCAACTTGGGCGAGAAAAGGTATATAAAGGTCTCTTGCTCCCATCAACTGCCTCAAAAGTAGGTATTCCAGCAG
161 +-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
      CCCACCGGTTGAACCCGCTCTTTTCCATATATTTCCAGAGAACGAGGGTAGTTGACGGAGTTTTTCATCCATAAGGTCGT

      ATCAGACAACCAACAACACACTTCATTTCCCAAGACATCACTCACAACAACCAACCTCTTCCAATCCAACCACAAACA
241 +-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
      TAGTCTGTTGGTTTGTGTTGTGAAGTAAGGGTCTGTAGTGAGTGTTGTTGGTTGGAGAAGGTTAGGTTGGTGTGTTGT

      AAAATCAGCCAATATGTCCGACTTCGAGAACAAAGACCCCAACAACGTCCTTGGCGGACACAAGGCCACCCTTCACAACC
321 +-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
      TTTTAGTCGGTTATACAGGCTGAAGCTCTTGTTCTTGGGGTTGTTGCAGGAACCGCTGTGTTCCGGTGGGAAGTGTTGG

      CTAGTATGTATCCTCCTCAGAGCCTCCAGCTTCCGTCCTCGTCGACATTTCTTTTTTTTCATATTACATCCATCCAAG
401 +-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
      GATCATACATAGGAGGAGTCTCGGAGGTCGAAGGCAGGGAGCAGCTGTAAAGGAAAAAAAAAGTATAATGTAGGTAGGTTC

      TCCCAACAATCCATGACTAACCAGAAATATCACAGATGTTTCCGAGGAAGCCAAGGAGCACTCCAAGAAGGTGCTTGAAAA
481 +-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
      AGGGTGTTAGGTACTGATTGGTCTTTATAGTGTCTACAAAGGCTCCTTCGGTTCCTCGTGAGGTTCTTCCACGAACTTTT

      CGCCGGCGAGGCCTACGATGAGTCTTCTTCGGGCAAGACCACCACCGACGACGGCGACAAGAACCCCGGAAACGTTGCGG
561 +-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
      GCGGCCGCTCCGGATGCTACTCAGAAGAAGCCCGTTCTGGTGGTGGCTGCTGCCGCTGTTCTTGGGGCCTTTGCAACGCC

      GAGGATACAAGGCCACCCTCAACAACCCCAAAGTGTCGACGAGGCCAAGGAGCACGCCAAGAAGAAGCTTGACGGCCTC
641 +-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
      CTCCTATGTTCCGGTGGGAGTTGTTGGGGTTTCACAGGCTGCTCCGGTTCCTCGTGCGGTTCTTCTTCGAACTGCCGGAG

      GAGTAAGCTCAGAGTTCACGAAAGAACCATTTCGACGAGGGGAAGCACGGGGTTATCTCGTTTCGAAACATGGGCCTGGTTA
721 +-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
      CTCATTGAGTCTCAAGTGCTTCTTGGTAAGCTGCTCCCTTCGTGCCCAATAGAGCAAGCTTTGTACCCGGACCAAT

      ATGCAAATGCATAATGGGGAGGATAATGAATCATGAGGTGTACGATATGGACGATATTGACGGATCTTAATTTGATGACA
801 +-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
      TACGTTTACGTATTACCCCTCCTATTACTTAGTACTCCACATGCTATACCTGCTATAACTGCCTAGAATTAAACTACTGT

      GTAATGAAATCACACCATAGT-3'
881 +-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
      CATTACTTTAGTGTGGTATCA-5'

```

Figure 1: Genomic DNA sequence of con-6 gene from *Neurospora crassa*. The sequence of both strands (5' to 3' on top, 3' to 5' on bottom) is shown above with nucleotides numbered 1 to 901. The dashed lines are interrupted every tenth nucleotide with a "+."

```

GENOMIC DNA: 251 CAAACAAACACACTTCATTCCCAAGACATCACTCACAACCAACCAACCTC 300
                |||
mRNA: 1 @AAACAAACACACUUCAUCCCAAGACAUCACUCACAACCAACCAACCUC 49

GENOMIC DNA: 301 TTCCAATCCAACCACAAACAAAAATCAGCCAATATGTCCGACTTCGAGAA 350
                |||
mRNA: 50 UUCCAAUCCAACCACAAACAAAAAUCAGCCAAUAUGUCCGACUUCGAGAA 99

GENOMIC DNA: 351 CAAGAACCCCAACAACGTCCTTGGCGGACACAAGGCCACCTTCACAACC 400
                |||
mRNA: 100 CAAGAACCCCAACAACGUCCUUGGCGGACACAAGGCCACCUUCACAACC 149

GENOMIC DNA: 401 CTAGTATGTATCCTCCTCAGAGCCTCCAGCTTCCGTCCCTCGTCGACATT 450
                |||
mRNA: 150 CUA..... 152

GENOMIC DNA: 451 TCCTTTTTTTTCATATTACATCCATCCAAGTCCCACAATCCATGACTAAC 500
mRNA: .....

GENOMIC DNA: 501 CAGAAATATCACAGATGTTTCCGAGGAAGCCAAGGAGCACTCCAAGAAGG 550
                |||
mRNA: 153 .....AUGUUUCCGAGGAAGCCAAGGAGCACUCCAAGAAGG 188

GENOMIC DNA: 551 TGCTTGAAAACGCCGCGAGGCCCTACGATGAGTCTTCTTCGGGCAAGACC 600
                |||
mRNA: 189 UGCUUGAAAACGCCGCGAGGCCUACGAUGAGUCUUCUUCGGGCAAGACC 238

GENOMIC DNA: 601 ACCACCGACGACGGCGACAAGAACCCCGGAAACGTTGCGGGAGGATACAA 650
                |||
mRNA: 240 ACCACCGACGACGGCGACAAGAACCCCGGAAACGUUGCGGGAGGAUACAA 288

GENOMIC DNA: 651 GGCCACCCCTCAACAACCCCAAAGTGTCCGACGAGGCCAAGGAGCACGCCA 700
                |||
mRNA: 289 GGCCACCCUCAACAACCCCAAAGUGUCCGACGAGGCCAAGGAGCACGCCA 338

GENOMIC DNA: 701 AGAAGAAGCTTGACGGCCTCGAGTAAGCTCAGAGTTCACGAAAGAACCAT 750
                |||
mRNA: 339 AGAAGAAGCUUGACGGCCUCGAGUAAGCUCAGAGUUCACGAAAGAACCAU 388

GENOMIC DNA: 751 TCGACGAGGGGAAGCACGGGGTTATCTCGTTCGAAACATGGGCCTGGTTA 800
                |||
mRNA: 389 UCGACGAGGGGAAGCACGGGGUUAUCUCGUUUGAAACAUGGGCCUGGUUA 438

GENOMIC DNA: 801 ATGCAAATGCATAATGGGGAGGATAATGAATCATGAGGTGTACGATATGG 850
                |||
mRNA: 439 AUGCAAUGCAUAAUGGGGAGGAUAAUGAAUCAUGAGGUGUACGAUAUGG 488

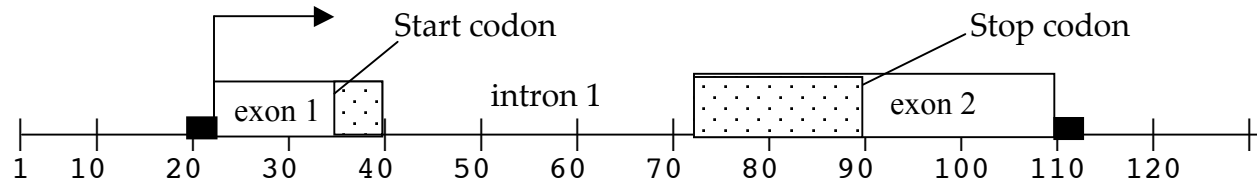
GENOMIC DNA: 851 ACGATATTGACGGATCTTAATTTGATGACAGTAATGAAATCACACCATAG 900
                |||
mRNA: 489 ACGAUUUGACGGAUCUUAUUUGAAAAAAAAAAAAAAAAAAAAAAAAAAAA 538

```

Figure 2: Sequence alignment of *con-6* genomic DNA and mRNA sequences. The top line of each pair of sequences is the sequence of *con-6* genomic DNA. The genomic DNA nucleotides are numbered as in Figure 1. The bottom line is the sequence of a *con-6* mRNA isolated from *Neurospora crassa*. The nucleotide numbers of the mRNA begin at the 5' end with #1, and end with #539 at the 3' end. Vertical dashes indicate nucleotides identical in both sequences (not base pairs!). Dots indicate nucleotides in the genomic sequence that are not found in the mRNA sequence (@ represents 5' G-cap).

You should use the information to make a map of the con-6 gene that follows the format shown below.

Maps have the following format; the numbers correspond to the numbers on the DNA strands. The map shown below is **for illustration purposes only** and does not correspond to the gene we are studying.



Here is what this map shows (and a list of all the features that a map must contain); note that all the map coordinates are approximate.

- The promoter is indicated by the small black rectangle at position 20.
- The start of transcription is indicated by a bent arrow; in this gene it is at position 22 (roughly).
- Exon 1 is indicated by a labeled box; it starts at 22 and ends at 40.
- Intron 1 is indicated by a labeled blank space; it starts at 41 and ends at 72.
- Exon 2 is indicated by a labeled box; it starts at 73 and ends at 110.
- The end of transcription is indicated by the end of the last exon; here it is at 110.
- The terminator is indicated by a small black rectangle around position 110.
- The start codon is indicated by the start of the hatched region in exon 1; it is at position 35.
- The stop codon is indicated by the end of the hatched region in exon 2; it is at position 90.
- The coding region, the region that encodes the protein, is indicated by the hatched parts of the exons; it extends from 35 to 90. Note that it **does not** include the intron.

a) Make your map of the con-6 gene based on Figures 1 and 2 on the preceding pages; be sure to include all the features shown on the example. If there are more than two exons and one intron, be sure to include them as well.

b) What are the first five and last five amino acids of the con-6 protein?

(4.2.3) Shown below is the sequence of a short fictitious eukaryotic gene. Both strands of DNA are shown.

After exhaustive studies, you have determined the following:

- Transcription in this organism always starts at the sixth nucleotide after the TATAA. That is, given the following sequence, the first nucleotide of the mRNA would be X: TATAAnnnnnX (where n can be any nucleotide).
- The intron splice sites have been well defined. Introns always begin with GUAUGU and end with CAG or UAG. That is, given the following sequences, the bases in **bold** will be in the mature mRNA while the underlined nucleotides will be spliced out as an intron:

XXXXGUAUGU(X)_nCAGXXXXXX
XXXXGUAUGU(X)_nUAGXXXXXX

- The intron-splicing machinery processes the RNA from 5' to 3'. It finds the first GUAUGU sequence and the first subsequent (moving 5' to 3') CAG or UAG and removes the intron between them. (Note that these rules and sequences are similar to those of real eukaryotic organisms.)
- This organism adds poly(A) tails immediately after the sequence GAAUAAAU. Poly(A) tails are usually about seven nucleotides long.
- RNA is processed in this sequence: transcription, then splicing, then polyadenylation and capping.

```

5'-GAAGCTAGAGGTCAATACCTGTATAAATGAAAAGGCGCTGGTATGTCCGAATAGCATGCA
1:  -----+-----+-----+-----+-----+-----+-----+ 60
3'-CTTCGATCTCCAGTTATGGACATATTTACTTTTCCGCGACCATACAGGCTTATCGTACGT

      GAACATGCCTCTGTATGTATTACTGTAGCTTTAAGGTACTACGTATGTCCGTATGTAATA
61:  -----+-----+-----+-----+-----+-----+ 120
      CTTGTACGGAGACATACATAATGACATCGAAATTCATGATGCATACAGGCATACATTAT

      AATAACTGTACAGTAACTAATGATGGTTGACGATACCCTCGGAATAAATGCGCATACGTA-3'
121: -----+-----+-----+-----+-----+-----+ 180
      TTATTGACATGTCATTGATTACTACCAACTGCTATGGGAGCCTTATTTACGCGTATGCAT-5'

```

a) The promoter sequence is TATAA. Why wouldn't the sequence TATA (or even TATATA, for that matter) work as a promoter sequence? (Hint: remember that a promoter is not just a place for RNA polymerase to bind; a promoter must indicate which direction RNA polymerase must read the DNA.)

b) What is the sequence of the mature mRNA from this gene?

c) What is the sequence of the protein produced from this gene?

For the following mutations, describe the changes to the mRNA sequence (either list the new mRNA sequence or just list that of the altered areas) and give the sequence of the protein produced by the mutant gene. Base pairs are numbered and “C/G base pair” means: C on the top strand, G on the bottom. (The base pairs in question are highlighted in bold.) Consider each mutation separately.

d) T/A base pair 52 is changed to A/T.

e) G/C base pair 41 is changed to T/A.

f) T/A base pair 86 is changed to G/C.

g) T/A base pair 127 is changed to A/T.

h) A/T base pair 24 is deleted.

(4.2.4) This question is based on some experiments that were described by Chan, A.C. et al. in the journal *Science* (volume 264, page 1599; 1994). Shown below is an internal portion of the genomic DNA sequence of a gene which produces a protein kinase (we will talk about these later in the course) essential for proper immune system function, shown as double-stranded base-paired DNA:

```

      1824          1840          1860          1880
      | . * . * . * . * . * . *
5' . . CCCTACAAGGCAGGCGCGGGCAGAGGCAGGTGGGCGGTGTGGTGGGGAGGGGGATGA
      |||||
3' . . GGGATGTTCCGTCCGCGCCCGTCTCCGTCCACCCGCCACACCACCCCTCCCCCTACT

      1881          1900          1920          1940
      | . * . * . * . * . *
GGAGGAGGACACTGGTCACTCACAGGTGTCTCTGCCCCGGCTTGAGCAGAAGATGAAAGGG
      |||||
CCTCCTCCTGTGACCAGTGAGTGTCCACAGAGACGGGCCGAACTCGTCTTCTACTTTCCC

      1941
      | . *
CCGGAGGTCATG . . . 3'
      |||||
GGCCTCCAGTAC . . . 5'

```

Starting with nucleotide 1824 and proceeding to the right, this region encodes the following amino acid sequence:

N-....Pro-Tyr-Lys-Lys-Met-Lys-Gly-Pro-Glu-Val-Met...C

a) The region contains a single intron. Using the sequence data above, locate the region encoding the intron within the above genomic DNA sequence. Using the numbering scheme used above, what are the first and last nucleotides of the DNA region that encodes the intron?

b) Does the intron you have identified in part (a) follow the “GT.....AG rule” described in most textbooks (the first two nucleotides of an intron are usually GU and the last two are usually AG)? Which is more compelling evidence for the intron’s boundaries, the “rule” or the protein sequence data?

c) A mutant form of this gene is known; it carries the recessive phenotype of complete immune deficiency. The mutation is a single-base-pair substitution in the region encoding the intron. The protein sequence of the corresponding region of the mutant protein is shown below with differences from wild-type underlined; the mutation results in the insertion of three amino acids into the middle of the protein.

N-...Pro-Tyr-Lys-Leu-Glu-Gln-Lys-Met-Lys-Gly-Pro-Glu-Val-Met-...C

i) Using the numbering scheme used above, what are the first and last nucleotides of the DNA region that encodes the intron in the mutant?

ii) Assuming that introns always end with AG and that the splicing machinery proceeds from 5' to 3', what is the mutation in these individuals (use the numbering scheme used above)?

(C2) Computer Activity 2: Gene Explorer (GeneX). This computer simulation of eukaryotic gene expression will help you to understand:

- More about transcription and translation
- Eukaryotic gene structure
- Exons and introns
- The effects of mutations on gene expression

Introduction:

Up until now, you have worked with genes of increasing complexity on paper. This can be time-consuming and error-prone. We will now move on to examine the structure and function of the genes using a computer simulation that takes care of the tedious details.

This simulation allows you to explore gene expression in a hypothetical simplified eukaryote. Eukaryotic genes have promoters and terminators for controlling transcription as well as start and stop codons for controlling translation. Although promoter and terminator sequences are different in different organisms, the genetic code, including the start and stop codons, is **identical in all organisms**. The promoter and terminator sequences used in the hypothetical organism simulated by the gene explorer are:

a promoter:

5'-TATAA-3'

|||||

3'-ATATT-5'

Transcription begins with the first base pair to the *right* of this sequence and continues to the right.

a terminator:

5'-GGGGG-3'

|||||

3'-CCCCC-5'

Transcription ends with the first base pair to the *left* of this sequence.

Therefore, a gene would look like this:

```
5'-TATAAXXXXXXXXXXXXXXXXXXXXXXXXXXXXXGGGGG-3'
      |||
3'-ATATTXXXXXXXXXXXXXXXXXXXXXXXXXXXXCCCCC-5'
```

Where the region shown as X's would be transcribed into pre-mRNA.

In addition, eukaryotic genes have a few features that prokaryotic genes do not have. These are:

- Transcription produces an mRNA called a pre-mRNA that is not yet ready for translation.
- This pre-mRNA is then processed in several steps to produce the mature mRNA ready for translation:
 - The introns are removed and the exons are joined; this is called mRNA splicing. This is controlled by splice signal sequences. In real organisms, these sequences are not well known. In general, introns start with 5'-GU-3' and end with 5'-AG-3'. In the hypothetical organism simulated by the Gene Explorer, introns start with 5'-GUGCG-3' and end with 5'-CAAAG-3'.
 - A modified G nucleotide is added to the 5' end of the mRNA; this is called the "cap." In the Gene Explorer, this is not shown.
 - Many A's are added to the 3' end of the mRNA; this is called the poly(A) tail. In real organisms, as many as 400 A's can be added at a specific signal sequence; the Gene Explorer adds 13 A's as a tail to the 3' end of any mRNA. Note that these A's do not correspond to T's in the DNA.

In previous problems, you did the work of expressing a gene by hand. Now that you are familiar with how these processes work, the Gene Explorer will do all the tedious work of:

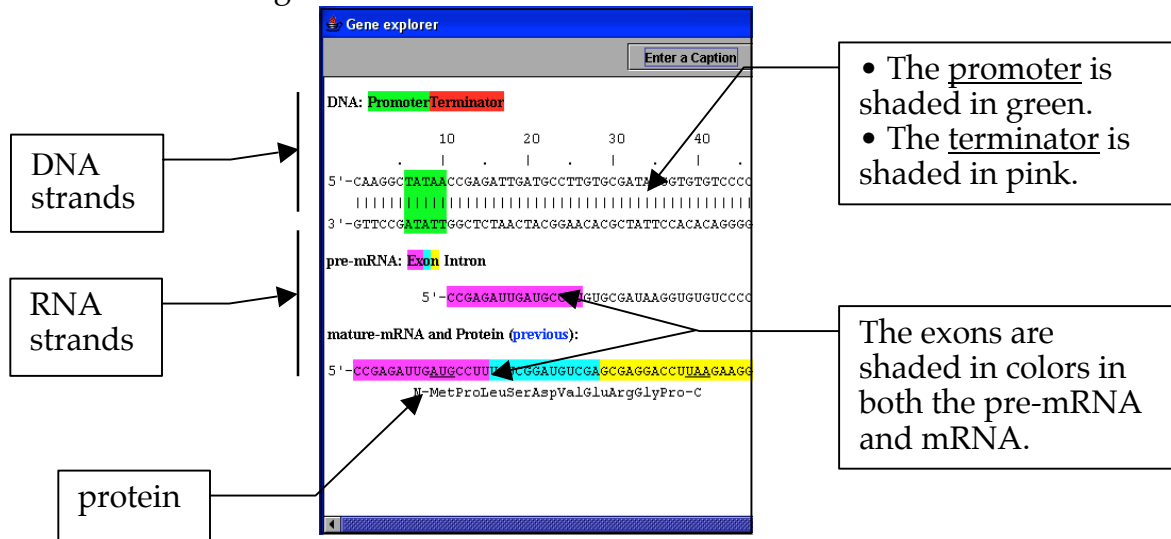
- Finding the promoter and terminator
- Reading the DNA sequence to produce the pre-mRNA
- Finding the splice sites
- Splicing and tailing the mRNA
- Finding the start codon
- Translating the mRNA

The Gene Explorer will then allow you to make specific mutations in a gene sequence, and it will then calculate and display their effects on the mRNA and protein. You do not have to deal with all the details listed above; the Gene Explorer will take care of it all. Researchers use tools like this to analyze the genes they are studying.

Procedure

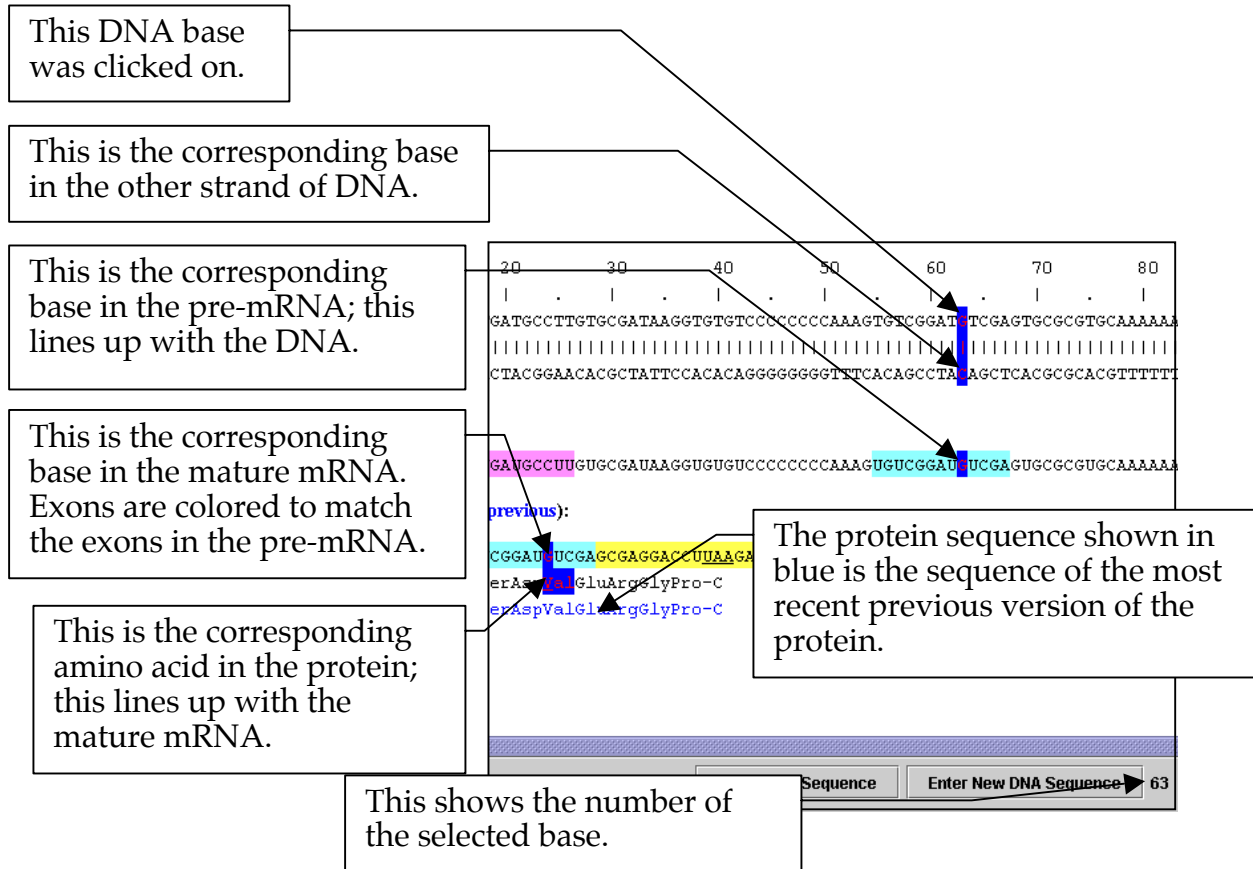
Part I: The Gene Explorer

- 1) Access the “**Gene Explorer**” from this site
http://intro.bio.umb.edu/MOOC/jsGX/JsGenex_C2.html.
- 2) You can also download a stand-alone version of Genex from
<http://intro.bio.umb.edu/GX/>
- 3) You will see something like this:



- 4) You can select a base in the DNA strand and the Gene Explorer will show the corresponding bases in the mRNA and the corresponding amino acid in the protein, if applicable. To select a base, here are some notes:
 - You can select bases only in the top strand of DNA. If you click anywhere else on the gene panel, nothing will happen.
 - It can be tricky to select a particular base on the first click; here are some tips:
 - You can use the right-arrow key to move one base to the right or the left-arrow key to move one base to the left.
 - To hit a particular base, click on the space between it and the next base. For example, to select 25, click on the space between 25 and 26.
 - The number of the selected base is shown at the bottom right of the Gene Explorer.

An example is shown below. Here, base 63 has been selected.



5) You can edit the DNA in several ways:

- To delete the selected base, use the "delete" or the "backspace" key.
- To replace the selected base with another base, type a *lowercase* letter (a, g, c, or t).
- To insert bases to the left of the selected base, type an *uppercase* letter (A, G, C, or T).

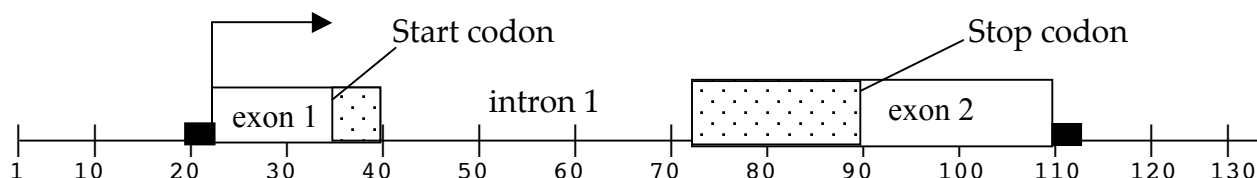
6) When you change the DNA sequence, the pre-mRNA, mature mRNA, and protein sequences are automatically updated. The *previous protein sequence*, the sequence of the protein before the latest change, is shown in blue for comparison purposes.

7) There are several useful buttons on the Gene Explorer:

- (Button that looks like a keyboard) – Click this to bring up some buttons for entering and editing DNA – this works on tablets including the iPad.
- Reset DNA Sequence – Click this to reset the DNA to the starting gene sequence.
- Enter New DNA Sequence – Click this to enter a new DNA sequence.
- Evaluate Answer – this is not used

Legend

Gene maps like the one on the preceding page have the following format: the numbers correspond to the numbers on the DNA strands. The map shown below is **for illustration purposes only** and does not correspond to the gene we are studying.



Here is what this map shows (and a list of all the features that a map must contain); note that all the map coordinates are approximate.

- The promoter is indicated by the small solid black rectangle at position 20.
- The start of transcription is indicated by a bent arrow; in this gene it is at position 22 (roughly).
- Exon 1 is indicated by a labeled box; it starts at 22 and ends at 40.
- The start codon is indicated by the start of the hatched region in exon 1; it is at position 35.
- The coding region, the region that encodes the protein, is indicated by the hatched parts of the exons; it extends from 35 to 40 and 73 to 90. Note that it **does not** include the intron.
- Intron 1 is indicated by a labeled blank space; it starts at 41 and ends at 72.
- Exon 2 is indicated by a labeled box; it starts at 73 and ends at 110.
- The stop codon is indicated by the end of the hatched region in exon 2; it is at position 90.
- The end of transcription is indicated by the end of the last exon; here it is at 110.
- The terminator is indicated by a small solid black rectangle around position 110.

Part II: A eukaryotic gene

1) Look at the Gene Explorer Display; it shows the unmodified gene.

2) Click on a base in the DNA and look at the parts of the other strands that are highlighted; these correspond to the base you clicked on. You can click on any base you like or use the arrow keys to move one base to the left or right. The number of the currently selected base is shown at the bottom of the Gene Explorer window.

Use these features to explore the normal gene. For each of the following, give a DNA nucleotide number that fits the description and give the name of that part of the gene (intron, exon, untranslated region, coding region, etc.). If that type of DNA base is impossible, put impossible. There may be more than one right answer for each.

a) A DNA base that corresponds to bases in the pre-mRNA, mature mRNA, and protein.

DNA base number_____

Part of gene_____

b) A DNA base that corresponds to bases in the mature mRNA, and protein but **not** in the pre-mRNA.

DNA base number_____

Part of gene_____

c) A DNA base that corresponds to a base in the pre-mRNA, but not in the mature mRNA.

DNA base number_____

Part of gene_____

d) A DNA base that corresponds to bases in the pre-mRNA, and the mature mRNA but **not** in the protein.

DNA base number_____

Part of gene_____

3) There are several important things to notice about this gene.

a) Note that the poly(A) tail, the string of A's at the 3' end of the mature mRNA, does not have any corresponding bases in the DNA or the pre-mRNA. How did they get there?

b) A common misconception is that introns do not split inside codons. This is not true, as you can see if you look at introns 1 and 2. Why doesn't it matter if an intron occurs in the middle of a codon?

c) Another common misconception is that an intron has to be a multiple of three nucleotides long. This is also not true, as you can see if you measure the length of introns 1 and 2 or if you try inserting some bases in an intron and see that it has no effect. Why is it that introns can be any number of nucleotides long?

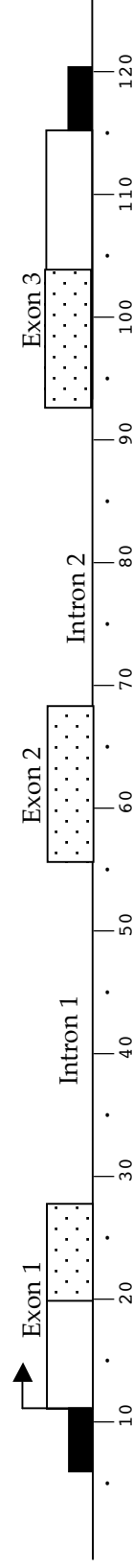
Part III: Mutations

You will now make several mutations in the gene and explore their effects.

- 1) Click "Reset DNA Sequence."
- 2) Try making mutations in different parts of the normal gene; these can be insertions, deletions, or substitutions of **one DNA base**. Be sure to click "Reset DNA Sequence" before making a new mutation; that way, you can see the effects of one mutation at a time. Remember that the protein sequence corresponding to the current gene is shown in black; the previous protein sequence is shown in blue for comparison purposes.
- 3) Make a map of the parts of the gene that can be changed **without changing the protein sequence** (other features of the gene can change). Choose one type of mutation (insertion, deletion, or substitution) to use in your studies. Using the map below, mark off the parts of the gene that can suffer a mutation without affecting the protein sequence.

Type of mutation: _____

- 4) Draw your map of regions that are insensitive to this type of mutation on the figure below:



- 5) Why are some regions insensitive to mutation while others are sensitive to mutation?

6) Click "Reset DNA Sequence."

7) Now do mutation 1: deleting the T/A base pair at position 26.

- Select base number 26; be sure that you have selected the right base by looking at the number in the lower right of the screen.
- Type the delete or backspace key once.
- Look at the sequence of the DNA strands; it should look like this: GCCTGTGC
|||||||
- If it does not look like this, go back to step 6 and try again. Note that base pair 26 is now missing.

8) What is the amino acid sequence of the mutant protein? How does it differ from the original sequence? What kind of mutation is this? Remember that the previous protein sequence is shown in blue; if you follow the directions above exactly, it will show the original protein sequence.

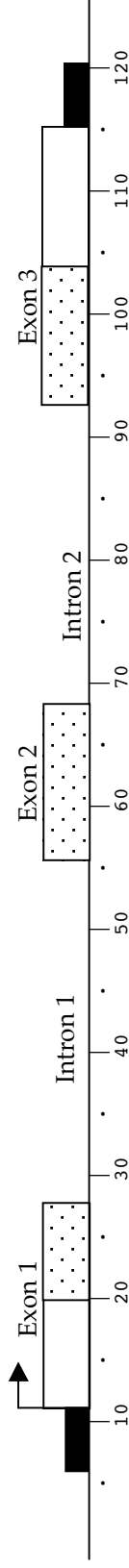
Original Protein Sequence: N-Met-Pro-Leu-Ser-Asp-Val-Glu-Arg-Gly-Pro-C

Mutant Protein Sequence:

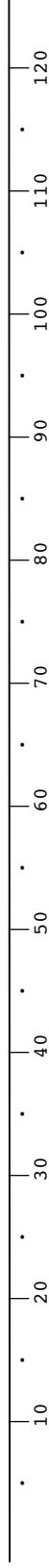
229

9) Using the line below, draw a map of the mutant gene. You can click on various bases in the DNA to help locate important parts of the gene. Indicate the location of the mutation in the DNA sequence with an asterisk (*). How does the structure of the mutant gene differ from the original gene?

Original Gene Map:



Mutant Gene Map:



10) Describe the differences between the original and mutant gene maps and explain how the mutant protein is longer even though the mutant gene is shorter.

11) Click "Reset DNA Sequence."

12) Now do mutation 2: changing the A/T base pair at position 51 to a T/A base pair.

- Select base number 51; be sure that you have selected the right base by looking at the number in the lower right of the screen.
- Type "t" (be sure it is lowercase).
- Look at the sequence of the DNA strands; it should look like this: GGC~~ATT~~TCA
- Note that base pair 51 is now T in the top strand and A in the bottom strand.
- If it does not look like this, go back to step 11 and try again.

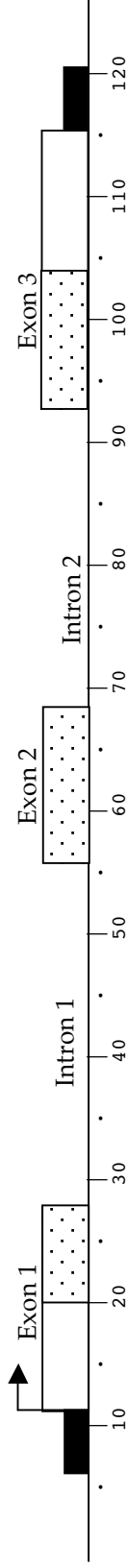
13) What is the amino acid sequence of the mutant protein? How does it differ from the original sequence?

Original Protein Sequence: **N-Met-Pro-Leu-Ser-Asp-Val-Glu-Arg-Gly-Pro-C**

Mutant Protein Sequence:

14) Using the line below, draw a map of the mutant gene. Indicate the location of the mutation in the DNA sequence with an asterisk (*). How does the structure of the mutant gene differ from the original gene?

Original Gene Map:



Mutant Gene Map:



15) Explain why the pattern of introns and exons is different in the mutant gene.

16) Click "Reset DNA Sequence."

17) Now do mutation 3: changing the T/A base pair at position 21 to a G/C base pair.

- Select base number 21; be sure that you have selected the right base by looking at the number in the lower right of the screen.
- Type "g" (be sure it is lowercase).
- Look at the sequence of the DNA strands; it should look like this: `ACTCCGGA`
- Note that base pair 21 is now G in the top strand and C in the bottom strand.
- If it does not look like this, go back to step 16 and try again.

18) What is the amino acid sequence of the mutant protein? How does it differ from the original sequence?

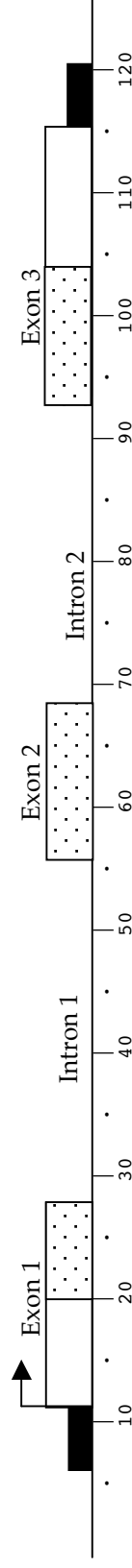
Original Protein Sequence: **N-Met-Pro-Leu-Ser-Asp-Val-Glu-Arg-Gly-Pro-C**

Mutant Protein Sequence:

231

19) Using the line below, draw a map of the mutant gene. Indicate the location of the mutation in the DNA sequence with an asterisk (*). How does the structure of the mutant gene differ from the original gene?

Original Gene Map:



Mutant Gene Map:



20) Explain why both the start and stop codons have now moved.

21) Click "Reset DNA Sequence."

22) Invent a nonsense mutation of your own.

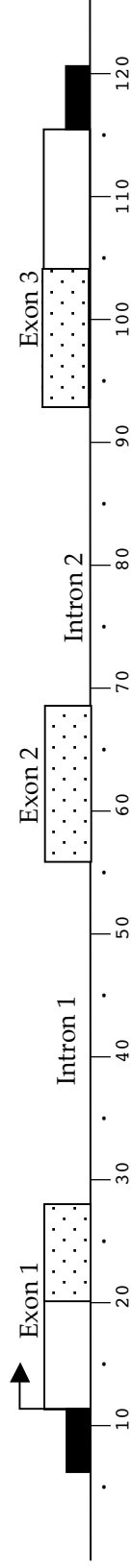
- Which base pair did you change? _____
- What did you change it to? _____

Original Protein Sequence: N-Met-Pro-Leu-Ser-Asp-Val-Glu-Arg-Gly-Pro-C

Mutant Protein Sequence:

- Draw a map of the resulting gene:

Original Gene Map:



Mutant Gene Map:



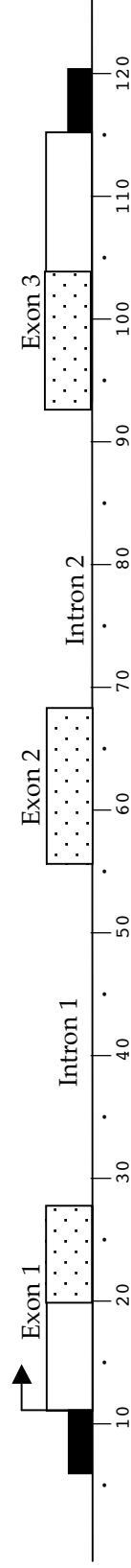
23) Explain how the mutation caused the change you diagrammed in part (22).

24) Click "Reset DNA Sequence."

25) Invent a mutation where a **deletion of one base** in the DNA causes the mature mRNA to be **longer**. Note that the normal mRNA [including the poly(A) tail] is about 65 nucleotides long; you can use the tick marks on the DNA strand to estimate the length of the mature mRNA.

- Which base pair did you delete? _____
- How long is the mature mRNA in the mutant? _____
- Is the mutant protein sequence the same as the original? _____
- Draw a map of the resulting gene:

Original Gene Map:



Mutant Gene Map:



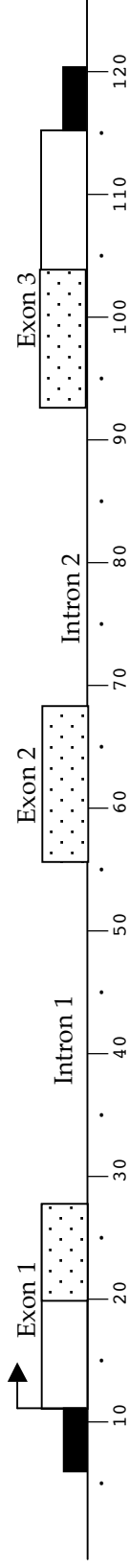
26) Explain how the mutation caused the change you diagrammed in part (25).

27) Click "Reset DNA Sequence."

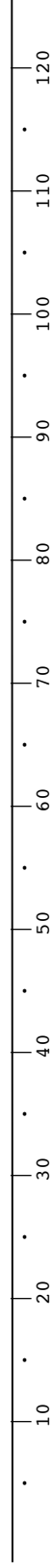
28) Invent a mutation that alters only one DNA base (insertion, deletion, or substitution) that results in no mRNA or protein being made at all.

- Which base pair did you change? _____
- What did you change it to? _____
- Draw a map of the resulting gene:

Original Gene Map:



Mutant Gene Map:



29) Explain how the mutation caused no mRNA to be made and how it caused no protein to be made.

This is a very challenging problem.

30) Click "Reset DNA Sequence."

31) A single base mutation (one base inserted, deleted, or changed) in the starting gene results in the following protein sequence:

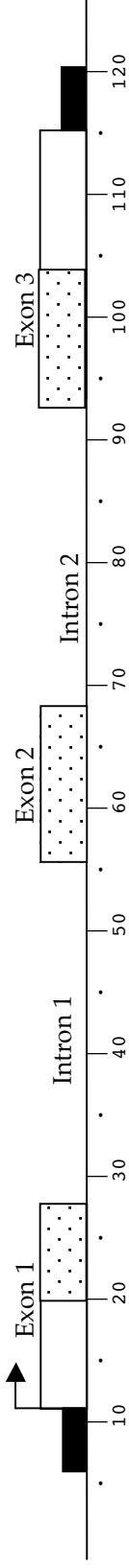
N-MetProLeuSerAspValAspAlaArgAlaLysLysAsnLysGlyGluAspLeuLysLysVal-C

Note that the first six amino acids are the same as the normal protein (underlined).

What mutation led to this? (There may be more than one right answer; give only one.)

- Which base pair was changed? _____
- What was the change? _____
- Draw a map of the resulting gene:

Original Gene Map:



Mutant Gene Map:



32) Explain how the mutation caused the change you diagrammed in part (31).

33) Here are some other mutations to try. For each one, explain how the mutation has the effect described.

- a) Make a mutation where one base is deleted that causes the protein to be longer.
- b) Make a mutation where one base is inserted that causes the protein to be shorter.
- c) Make a mutation that causes Exon 2 to be absent from the mature mRNA.

34) Design an entirely new gene that you have invented. Using the new Gene Explorer, this gene should (you can make it more challenging if you like):

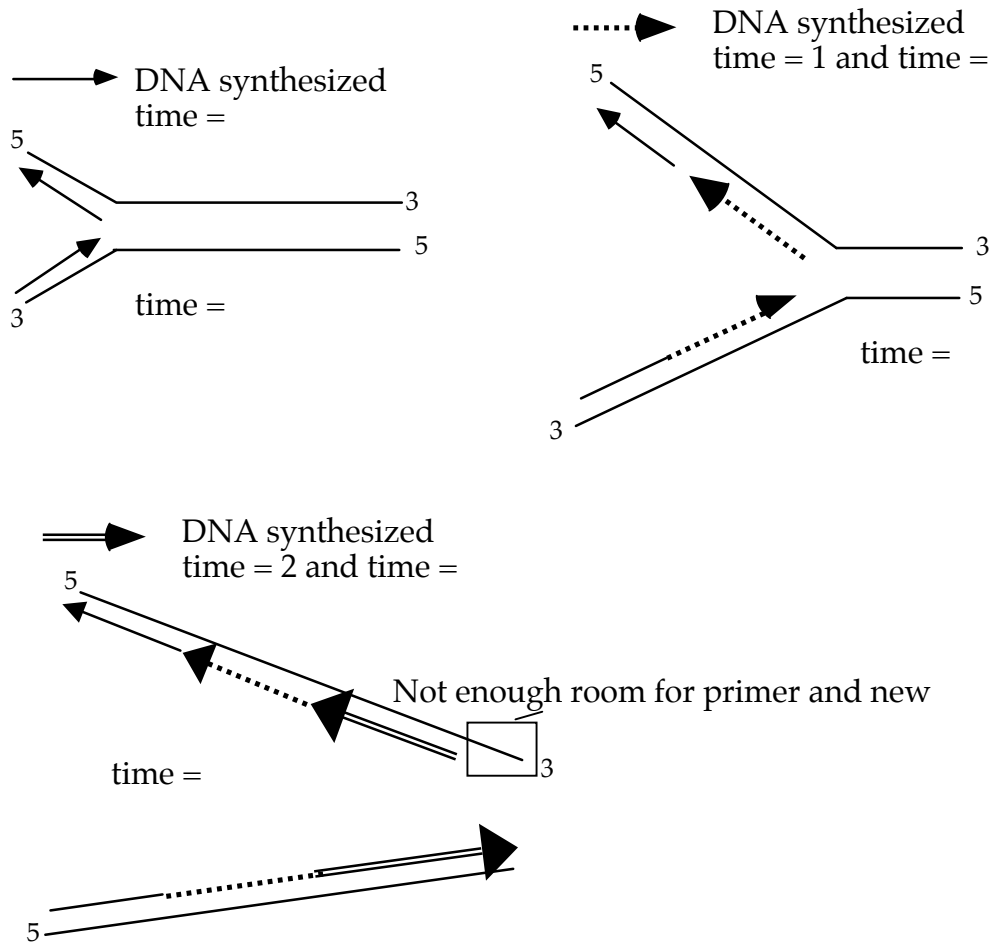
- Produce a protein of at least five amino acids (including the N-terminal Met).
- Contain at least one intron.

Tips

- Use the “Enter New DNA Sequence” button and delete the starting sequence from the entry blank.
- Type in a promoter, a little DNA, and a terminator; be sure your RNA is made.
- Click on your gene and add the start codon, coding region, and stop codon; be sure your protein is made. Type slowly so that the program can keep up.
- Similarly, add an intron in the coding region and be sure your gene works.

(5) CHALLENGE PROBLEMS

(5.1) DNA synthesis cannot begin *de novo*. It requires a free 3'-OH group. This free OH is provided by an RNA primer. The RNA polymerase that makes this primer does not require an end on which to build. DNA polymerase's requirement for a primer has an interesting effect on DNA replication. The final 3' end of the lagging strand cannot be replicated, because there is no DNA left from which to make the RNA primer.



If this problem were left uncorrected, linear chromosomes (which are long stretches of DNA) would shrink with each successive replication. You might guess correctly that this would have a detrimental effect on an organism. Therefore, organisms have evolved ways to combat this problem.

a) Some organisms, such as bacteria and viruses, have circular, not linear, chromosomes. Explain how having a circular chromosome could solve the shrinkage problem explained above.

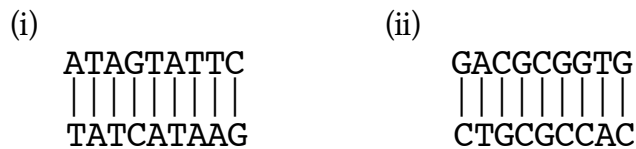
b) Another strategy that organisms take to combat shrinkage is to have linear chromosomes with telomeres at either end. A telomere is simply a long stretch of repeated nucleotides. For example, in yeast (*S. cerevisiae*), there is a telomere composed of many (TGTGTGTG)_n repeats present at the end of each chromosome, where *n* can equal several hundred. A special enzyme called telomerase periodically extends the length of this repeat sequence without requiring a template. How could having a telomere solve the problem of shrinking chromosomes?

(5.2) Compare DNA and RNA.

a) List three key differences between DNA and RNA structures.

b) What molecular interaction allows base pairing to occur between two of the strands of DNA in a double-stranded DNA helix?

c) DNA denaturation is the separation of the two strands of a DNA molecule. Consider the two DNA sequences shown below. The symbol “|” indicates the molecular interactions between the base pairs. Which sequence (i) or (ii) would you expect to denature at a higher temperature? Briefly explain your reasoning.

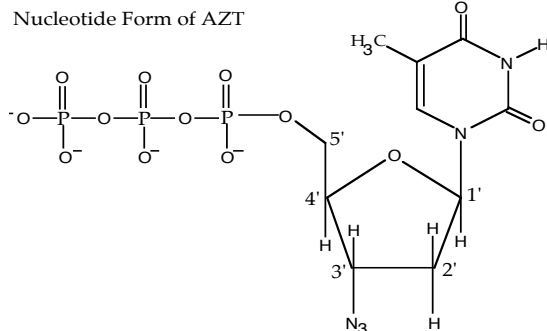


d) A DNA helix is formed from two DNA molecules with properly aligned base pairs. Aside from base pairing, what other forces might contribute to the stability of the DNA helix?

e) The usual base pair relationships are A-T and C-G. Which bases are purines and which are pyrimidines? How might the DNA double helix be affected by the base pair mismatch of A-G or C-T.

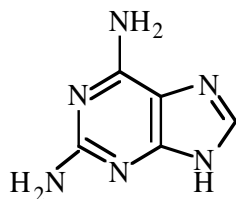
f) In a cell that is at physiological pH, what would be the overall charge (positive or negative) of a double-stranded DNA molecule? Briefly justify your answer.

g) You carry out a DNA replication reaction using a single-stranded DNA template, DNA polymerase, a primer, and the four deoxyribonucleoside triphosphates. You then add the nucleotide form of AZT (azidothymidine) to the reaction mixture. The structure of AZT is very similar to deoxythymidine except that in AZT, the 3'-hydroxyl (OH) group on the deoxyribose ring has been replaced by an azido (N_3) group. The nucleotide form of AZT is shown on the next page.



- What would you expect to happen to DNA replication when you add the AZT nucleotide to the reaction mixture? Briefly explain your reasoning.
- Why might AZT help individuals who have cancer or who are infected with HIV (human immunodeficiency virus)?

h) Consider the structure of the base, 2,6-diaminopurine, shown below.



2,6-diaminopurine

With which of the normal pyrimidines (C or T) would 2,6-diaminopurine be able to base pair? Draw the base pair and indicate the hydrogen bonds that would be formed. (You do not need to draw the structure of the sugar or the phosphate groups.)

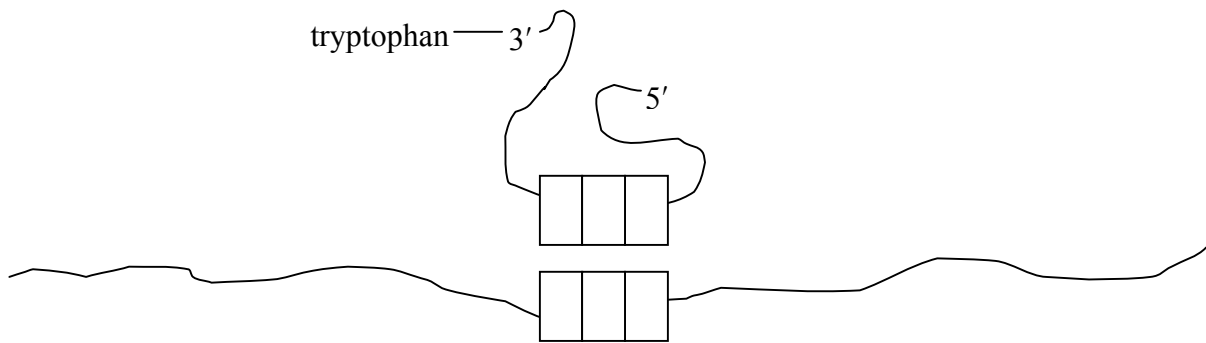
(5.3) You are studying translation in a bacterial species that has a single gene encoding the tryptophan tRNA (the *trnA-trp* gene). The wild-type sequence of this gene is shown below. The portion of the gene that encodes the anticodon of the tRNA is boxed.

```

5' GTACCTGCACTGCATGCCTAGCTAGCCCTAGCCAAGCCTAGCTAGCTAGCACCAA3'
3' CATGGACGTGACGTACGGATCGATCGGGATCGGTTCGGATCGATCGATCGTGTT5'

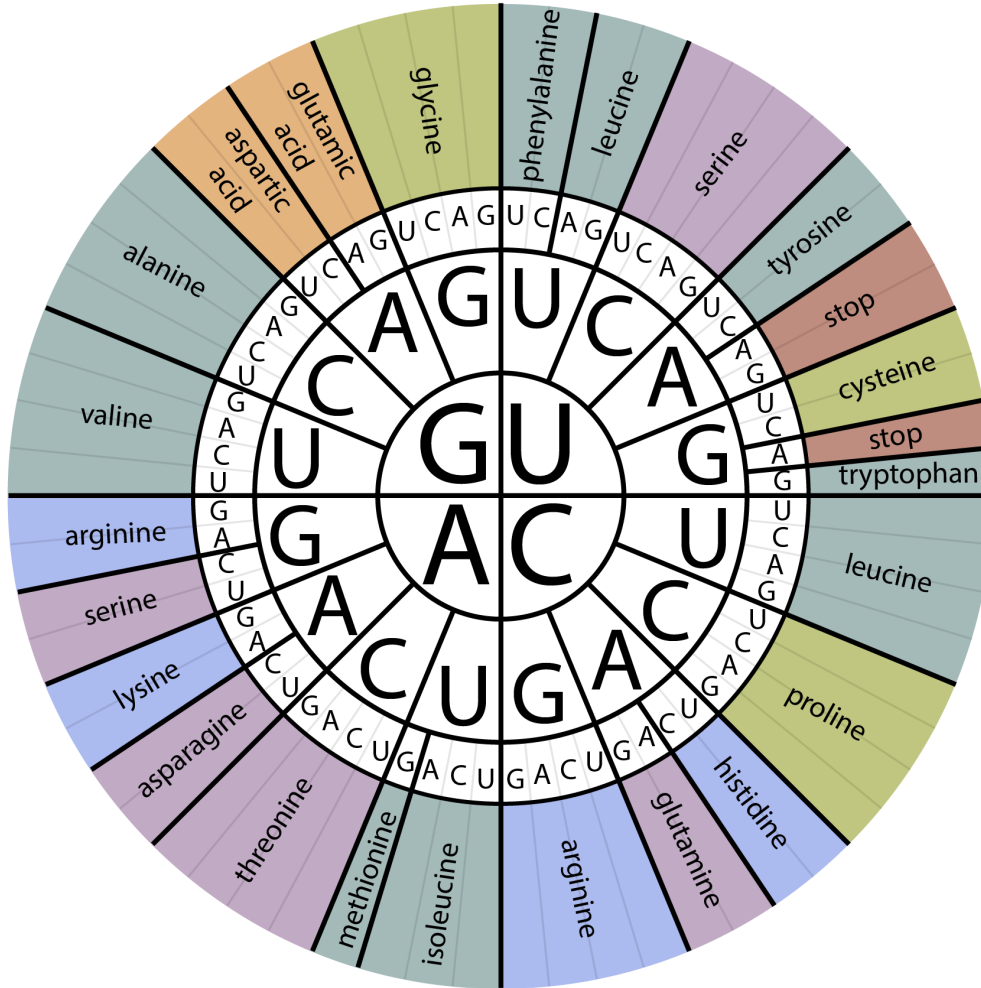
```

a) Below is drawn the folded tryptophan-tRNA (produced from this *trnA-trp* gene). It is shown base pairing with an mRNA containing the codon that this tryptophan-tRNA recognizes. Fill in the three boxes on the tRNA with the correct nucleotide sequence of its anticodon. Then fill in the three boxes on the mRNA with the correct nucleotide sequence of the codon currently being read. Be sure to label the ends of the mRNA to show directionality.



b) Which strand of the double-stranded *trnA-trp* gene is used as a template when the tryptophan tRNA is transcribed, the upper strand or the lower strand? Remember that tRNAs are transcribed directly from genes; there is no mRNA intermediate made during the production of a tRNA from its DNA sequence.

Here are two different ways to show the genetic code:



	U	C	A	G	
U	UUU phe (F) UUC phe (F) UUA leu (L) UUG leu (L)	UCU ser (S) UCC ser (S) UCA ser (S) UCG ser (S)	UAU tyr (Y) UAC tyr (Y) UAA STOP UAG STOP	UGU cys (C) UGC cys (C) UGA STOP UGG trp (W)	U C A G
C	CUU leu (L) CUC leu (L) CUA leu (L) CUG leu (L)	CCU pro (P) CCC pro (P) CCA pro (P) CCG pro (P)	CAU his (H) CAC his (H) CAA gln (Q) CAG gln (Q)	CGU arg (R) CGC arg (R) CGA arg (R) CGG arg (R)	U C A G
A	AUU ile (I) AUC ile (I) AUA ile (I) AUG met (M)	ACU thr (T) ACC thr (T) ACA thr (T) ACG thr (T)	AAU asn (N) AAC asn (N) AAA lys (K) AAG lys (K)	AGU ser (S) AGC ser (S) AGA arg (R) AGG arg (R)	U C A G
G	GUU val (V) GUC val (V) GUA val (V) GUG val (V)	GCU ala (A) GCC ala (A) GCA ala (A) GCG ala (A)	GAU asp (D) GAC asp (D) GAA glu (E) GAG glu (E)	GGU gly (G) GGC gly (G) GGA gly (G) GGG gly (G)	U C A G

Challenge Problems

