

# Molecular Biology Solutions

## (1) PROBLEMS EXPLORING CLASSIC EXPERIMENTS

### (1.1)

a) The negative control shows that either  $R_{(II)}$  or heat-killed  $S_{III}$  alone is nonvirulent. Without this result, you could argue that one or the other killed the mice, and it would not be necessary to invoke transformation to explain the results of the central experiment.

If the negative control experiment resulted in dead mice, the result of the central experiment would have been meaningless.

The positive control shows that the bacteria do not change type once inside the mouse. Without this result, you could argue that the bacteria randomly change type. It also shows that the bacteria are still lethal to the mouse. This control shows more directly that  $S_{III}$  is the cause of death. Without this result, you could argue that the mouse died of some other cause and just happened to harbor  $S_{III}$  bacteria.

If the positive control experiment resulted in mice that lived, you could not have done the central experiment because there would be no difference between injecting the mixture or injecting either  $R_{(II)}$  or heat-killed  $S_{III}$  alone.

b) It was essential to use a mixture of bacteria that are derived from different strains, i.e.,  $R_{(II)}$  and heat-killed  $S_{III}$  instead of a mixture of  $R_{(II)}$  and heat-killed  $S_{II}$ . This excludes the possibility that a few  $R_{(II)}$  bacteria had reverted back to  $S_{II}$  and these  $S_{II}$  were responsible for the death of the mice. In this case, it would not be necessary to invoke transformation to explain the results of the central experiment. However, because  $R_{(II)}$  and  $S_{III}$  are derived from different strains, there can be no reversion. So when  $R_{(II)}$  and heat-killed  $S_{III}$  were injected, and virulent  $S_{III}$  are found, we are forced to consider transformation to explain the results of the central experiment.

### (1.2)

a) They presumed that the transforming substance was genetic material, because the transformation from R to S was heritable. Once an R cell was transformed into an S cell, all the progeny from that cell and successive generations were S cells.

b) Even the smallest trace of protein in their preparations of the transforming substance could support model (1). It could be that the DNA is simply a carrier for the proteins that are the true genetic material. In that case, you might need only a very little protein to make a gene.

c) i) In the four base pairs shown, there are 30 N's and 8 P's. To get the mass ratio, you must multiply by the atomic weights of each atom ( $N = 14$ ,  $P = 31$ ):

$$\frac{30 \times 14}{8 \times 31} = 1.69$$

ii) Protein contamination should increase the N/P ratio. Protein contains N, but no P.

iii) These data (with the exception of #38B) show a ratio of 1.69 or below. So it appears that the preparations were not contaminated with protein. However, it is not clear why two of the N/P ratios are  $\leq 1.69$ .

d) i) These data support the model that genes are made of DNA and not protein. If the genetic material were protein, then treatment with these enzymes should destroy its activity. If it were DNA, they should have no effect.

ii) This is a negative result; there was no effect. Because of this, you can never be sure that the enzymes were working properly or if somehow the proteins that make up genes are resistant to these enzymes.

While this work was compelling, it was not completely conclusive. Avery, McCarthy, and MacLeod did several additional experiments. The accumulation of data, all consistent with model (2), gradually convinced the scientific community. This is typical of the process of science; there is almost never one crucial experiment that answers a question once and for all.

### (1.3)

a) i) If the phage injected protein, not DNA, then in experiment 1 you would expect to find most of the  $^{32}\text{P}$  in the supernatant because the  $^{32}\text{P}$ -labeled DNA was never injected into the bacterium.

If the phage injected protein, not DNA, then in experiment 2 you would expect some  $^{35}\text{S}$  in both supernatant (sheared-off phage heads) and the pellet (transferred genetic material). The progeny phage would have some  $^{35}\text{S}$  from the reused genetic material.

ii) If the phage injected genetic material that was mostly DNA and a little protein, then in experiment 1 you would find little  $^{32}\text{P}$  in the supernatant and a lot in the pellet since it has been injected into the bacteria.

If the phage injected genetic material that was mostly DNA and a little protein, then in experiment 2 you would expect  $^{35}\text{S}$  in both supernatant (sheared-off phage heads) and a little in the pellet (transferred genetic material). The progeny phage would have some  $^{35}\text{S}$  from the reused genetic material.

iii) If the phage injected DNA and protein, but protein is the genetic material, and DNA is only a scaffold, then in experiment 1 you would find little  $^{32}\text{P}$  in the supernatant and a lot in the pellet since it has been injected into the bacteria. If protein is the genetic material, and DNA is only a scaffold, then in experiment 2 you would expect some  $^{35}\text{S}$  in both supernatant (sheared-off phage heads) and in the pellet (transferred genetic material). The progeny phage would have some  $^{35}\text{S}$  from the reused genetic material.

b) i)  $^{32}\text{P}$  in the supernatant represents phage DNA that was not injected into the bacteria (perhaps the phages were sheared off before they injected their DNA) or DNA present in phage that did not attach to the cells in the first place. It is not crucial that

this number be zero because their model did not require that all the DNA of all the phages had to be injected.

ii)  $^{35}\text{S}$  in the pellet represents protein that remained associated with the bacteria. Perhaps not all phages are sheared off the blender, or perhaps a small portion of the phage is left attached to the bacteria.  $^{35}\text{S}$  in the pellet makes the experiments less conclusive, since it could also represent protein that was injected along with the DNA, which could be genetic material or other proteins. This undermines the model that only DNA injection is required for infection and that therefore DNA is the genetic material.

c) You can rule out only the possibility that the genetic material injected was protein and the DNA remained in the phage head (model 1). This model predicts that very little  $^{32}\text{P}$  would go into the bacteria, but that is not what the data show. Both other models are consistent with the data.

**(1.4)**

a)

- Nucleic acid will be labeled with  $^{32}\text{P}$ ?
- All macromolecules will be labeled with  $^3\text{H}$ ?
- Proteins will be labeled with  $^{35}\text{S}$ ?

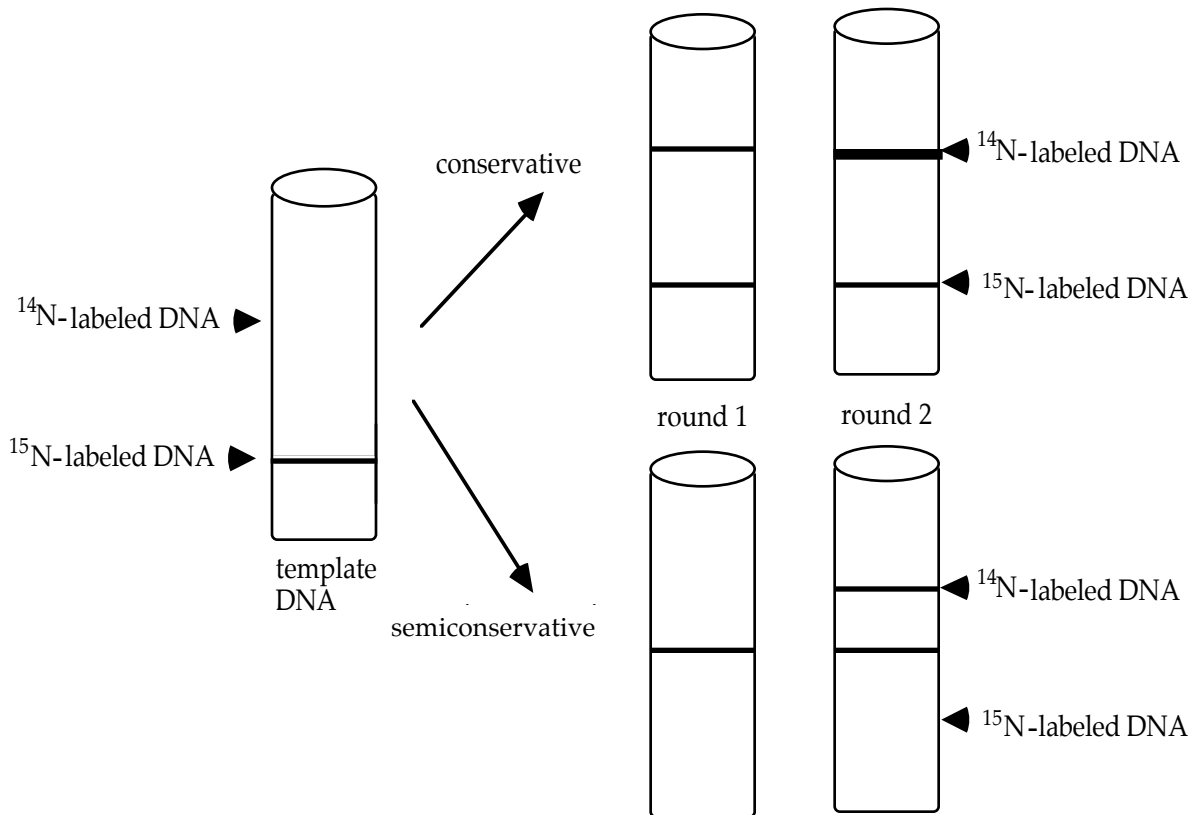
b) This virus is carrying double-stranded DNA as shown by the  $\%A = \%T$ ,  $\%C = \%G$ . The trace of uracil is contaminating RNA.

c) In a manner analogous to the Hershey-Chase experiment, you would infect cells with  $^{32}\text{P}$  labeled virus. After a short time, you would separate the virus particles from the cells and examine whether the  $^{32}\text{P}$ -labeled DNA is found in the virus particles (in the supernatant) or in the bacteria (in the pellet).

d) i) An in vitro system for DNA replication would include template DNA, primer, dNTPs, DNA polymerase, ligase, helicase, topoisomerase, single-stranded binding protein.

ii) Nitrogen is found in the bases of 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

### (Computer Activity 1)

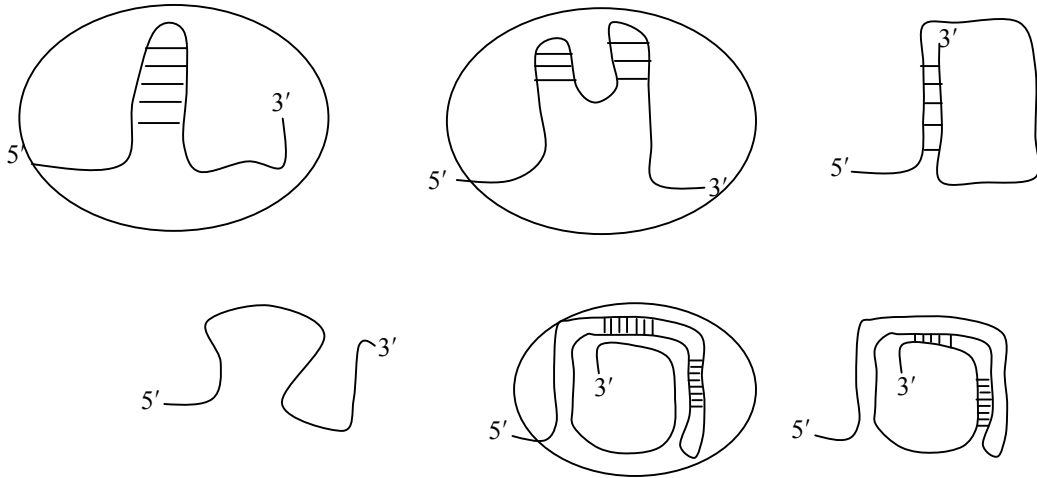
- a) i) The strands are antiparallel. They run 5' to 3' in opposite directions. If you reset the view and click the (i) button:
- The strand that starts at the top left (light yellow) runs 5' to 3' top to bottom.
  - The strand that starts at the top right (light purple) runs 3' to 5' top to bottom.

ii) There are no covalent bonds between the bases, only hydrogen bonds. Therefore, this view shows two molecules of DNA and the entire object should be called "a double-stranded DNA complex" ("double-stranded DNA molecule" is, strictly speaking, incorrect). These are two DNA strands.

In the DNA double helix, the bases are on the inside of the helix; in a protein alpha-helix, the side chains are on the outside.

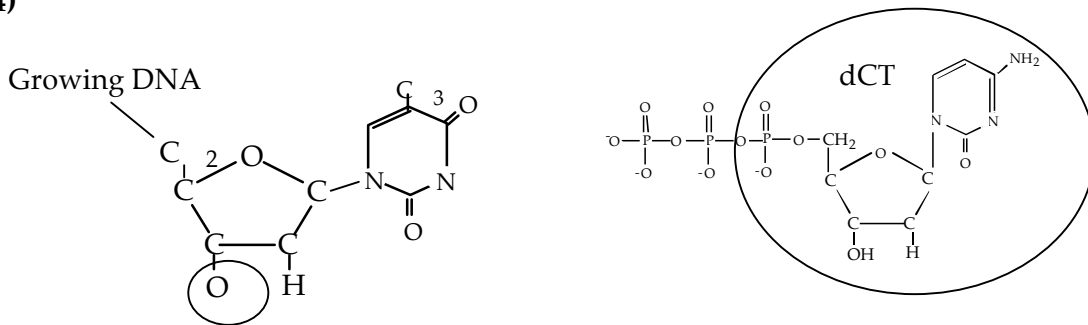


(2.3)  
a)



b) 3' ...TGCCTGCG...5'

(2.4)



### (3) DNA REPLICATION

#### (3.1)

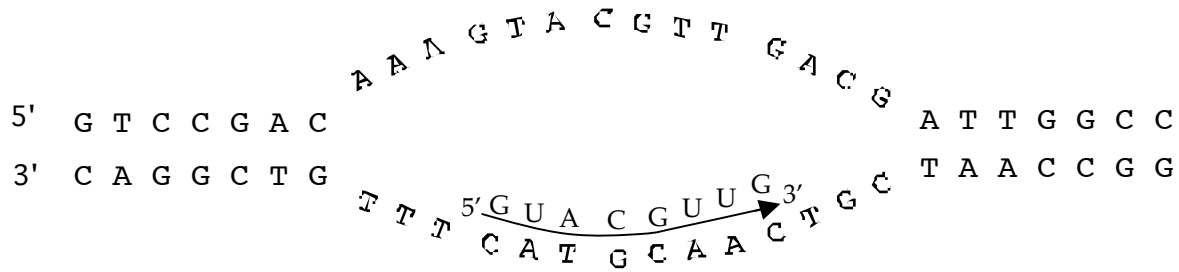
Enzyme activity	Function(s)
Topoisomerase	k
Primase (synthesizes primer)	e, f, j
DNA polymerase to elongate new DNA strand	b, c
Helicase to unwind DNA	h
DNA polymerase to replace RNA with DNA	d
Processivity factor	i

Choose from:

- a) 3' ⇒ 5' growth of new DNA strand
- b) 5' ⇒ 3' growth of new DNA strand
- c) 3' ⇒ 5' exonuclease
- d) 5' ⇒ 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)

a)



b) If the replication fork moves to the left, this primer will be used to create the lagging strand because the direction of replication will be moving in the opposite direction of the movement of the origin of replication.

c) Replication is discontinuous because the origin is moving 3' to 5' relative to the lagging strand, but nucleotides can be added only in the 5' to 3' direction. On the lagging strand, each time a primer is created the replication moves in the opposite direction of the origin. Thus, as the replication fork moves and the DNA is unwound, more unreplicated DNA at the 5' end of the primer is revealed. Since DNA can only be replicated 5' to 3', new primers must constantly be added, making a discontinuous strand of DNA.

(3.3)

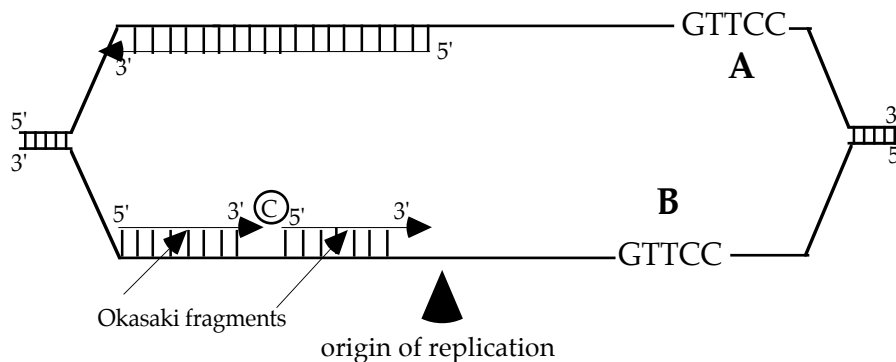
a) i) Replication will be continuous on templates 1 and 4

ii) The primer 5'-GUUCC-3' binds to and initiates replication at sites B and C.

iii) The lagging strand is more affected by the lack of DNA ligase. DNA replication on the lagging strand occurs in small stretches called Okazaki fragments. For replication of the lagging strand to be complete, a phosphodiester bond must be formed between the 3'-OH on one Okazaki fragment and the 5'-phosphate on the other. DNA ligase makes this bond.

(3.4)

a)





- b) DNA ligase is required at C in the above diagram.
- c) The primer 5'-CAAGG-3' binds at site B to initiate replication.
- d) From site B, the direction of elongation of the daughter DNA strand is to the right.
- e) From site B, DNA synthesis is performed in a continuous fashion relative to the nearest replication fork.

**(3.5)**

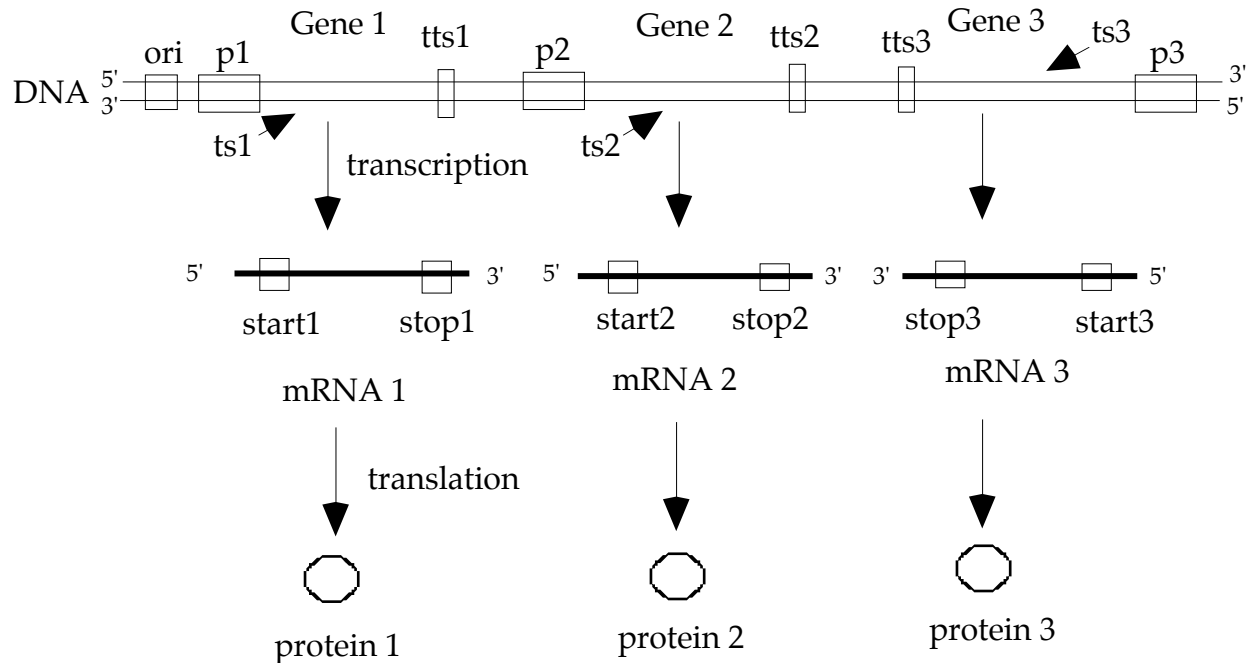
- a) The sequence of the RNA primer that binds to the top strand at base-pair positions 80–90 is 5'-UGUACGCAUGC-3'.
- b) DNA synthesis from the primer in (a) would be continuous to the left as the diagram is displayed.
- c) The sequence of the RNA primer that binds to the bottom strand at base-pair positions 90–100 is 5'-AUAGUUCGACG-3'.
- d) DNA synthesis from the primer in (c) above would be continuous to the right as the diagram is displayed.

## (4) TRANSCRIPTION AND TRANSLATION

### (4.1) Transcription and translation in prokaryotes

#### (4.1.1)

a)



**promoter** - DNA sequence recognized by RNA polymerase. Signals RNA polymerase to start transcription; always found at the 5' end of the transcribed region.

**transcription termination site** - DNA sequence recognized by RNA polymerase. Signals RNA polymerase to stop transcribing. This determines the 3' end of the message.

**start codon** - RNA sequence recognized by ribosome. Signals ribosome to begin translation; often found near the 5' end of mRNA.

**stop codon** - RNA sequence recognized by ribosome. Signals ribosome to end translation; often found near the 3' end of mRNA.

**transcribed strand** - DNA strand that RNA pairs with during transcription; has orientation opposite to mRNA produced.

b) No. If RNA polymerase II can't be recruited to the promoter, the mRNA for gene 2 cannot be transcribed, so no protein will be produced.

c) Because the transcription termination site in gene 1 was mutated, RNA polymerase II can continue making an mRNA molecule until it encounters the next available transcription termination site. Apparently, this was the one located in gene 2. Thus, one long mRNA contains the protein-coding regions for both gene 1 and gene 2. Each of these protein-coding regions is preceded by a start codon, so each will be translated by ribosomes as they scan the mRNA.

**(4.1.2)**

a) 5' AAUUGUGAAU....3'

b) 5' AAUCCGAGC...3'

c) The mRNA has the same sequence as the DNA strand that oriented 5' to 3', because all nucleic acids are made in the 5' to 3' direction. Synthesis is directed by a template that runs antiparallel to the newly synthesized molecule. The mRNA is the same except that wherever there is a T in the DNA, there is a U in the mRNA.

d)  $H_3N^+$ -Met-Asp-Asn-Asn-Val-Thr-Gln-Glu-Thr-Ala-Lys-Thr-Met-Phe- $COO^-$   
Protein synthesis begins at the start codon, usually the first AUG of the mRNA. Note that there can be a Met within the protein sequence; translation does not restart there. Also note that the stop codon does not specify an amino acid; the chain ends with the last amino acid before the stop codon.

e) No. Translation does not terminate there because the UAA is in a different reading frame; it is read as: GCU AAG A.

f)  $H_3N^+$ -Met-Asp-Asn-Asn-Val-Thr-Gln-Glu-Thr-Ala-Lys-Thr-Met-Phe- $COO^-$

For the next parts, the altered amino acids are shown in **bold and underlined** type:

g)  $H_3N^+$ -Met-Asp-Asn-Asn-**Gly-Asp-Thr-Gly-Asn-Ser**- $COO^-$ . Note that all the amino acids after the mutation are altered.

h)  $H_3N^+$ -Met-Asp-**Lys-Met**- $COO^-$ .

i)  $H_3N^+$ -Met-Asp-Asn-**Lys**-Val-Thr-Gln-Glu-Thr-Ala-Lys-Thr-Met-Phe- $COO^-$ .  
Note that only one amino acid is changed. This is a missense mutation.

**(4.1.3)**

a) 5'-AAACAGCUAUGGCCA .....-3'

b)  $H_3N^+$ -Met-Ala-Met-Ser-Thr-Pro-.....- $COO^-$

c) No. Translation of the mRNA does not terminate at this TAA, because nucleotides appear as AUU in the mRNA transcript. The TAA is on the noncoding strand.

d) No. Translation does not terminate there because the TAA chain-terminating codon would not be in the correct reading frame in the mRNA transcript; in the correct reading frame (set by the first AUG codon), they are read as AUU AAA.

e)  $H_3N^+$ -.....Asn-Arg-Gly- $COO^-$

f) A frameshift mutation as a result of the deletion would alter the sequence of the mRNA; however, the protein sequence would not be altered because the deletion is prior to the start codon.

g) There would be no change in the amino acid sequence of the protein because the codons GGC and GCA both code for the amino acid alanine.

h) The GCC in the sequence that codes for alanine would be changed to CCC, which would now code for proline.

i) The protein would be terminated prematurely because a new stop codon was created at the new nucleotide positions 55-57. As a result of this deletion, the new sequence of the protein would be:  $\text{H}_3\text{N}^+\text{-Met-Gly-COO}^-$ .

**(4.1.4)**

a)  $\text{NH}_3^+\text{-Met-Ser-Cys-Trp-}$   
 5' - AUG UCU UGU UGG ... 3'  
           UCA UGC  
               UCG  
           UCC  
           AGU  
           AGC

b) Gly is GGA, GGC, CCU, or GGG. The only way to get there via one base change is from AGU --> GGU or AGC --> GGC. So the original sequence for the Ser was AGU or AGC.

c) From (b), there are four possibilities for the normal sequence. Here they are as normal and with the first nucleotide of codon 2 deleted:

(1)	5' - AUG AGU UGU UGG...3'
with one base deleted:	5' - AUG GUU GUU GG.... Met Val Val Gly NOT RIGHT
(2)	5' - AUG AGC UGU UGG...3'
with one base deleted:	5' - AUG GCU GUU GG.... Met Ala Val Gly NOT RIGHT
(3)	5' - AUG AGC UGC UGG...3'
with one base deleted:	5' - AUG GCU GCU GG.... Met Ala Ala Gly NOT RIGHT
(4)	5' - AUG AGU UGC UGG...3'
with one base deleted:	5' - AUG GUU GCU GG.... Met Val Ala Gly RIGHT

Therefore, the original sequence was 5'-AUG AGU UGC UGG...3'.

**(4.1.5)**

a) No. You can model any length codon and get the same results with a homopolymer template.

b) One nucleotide per codon would only give four amino acids, so it can be ruled out. Two would work – 16 total codons could give 14 amino acids and two stops. Three is possible, but they would be degenerate or have many stop codons. To eliminate one of these two options, you can predict the results of translating  $(WX)_n$  in a two or three nucleotides per codon scheme:

two nucleotides per codon could be read as WX WX WX WX or W XW XW XW X, giving a mixture of  $(aa1)_m$  and  $(aa2)_m$ .

three nucleotides per codon could be read as WXW XWX WXW or XWX WXW XWX, giving  $(aa1-aa2)_m$  only.

Since we see a mixture of two homopolymers, there must be two nucleotides per codon.

c) You can work out the code as follows:

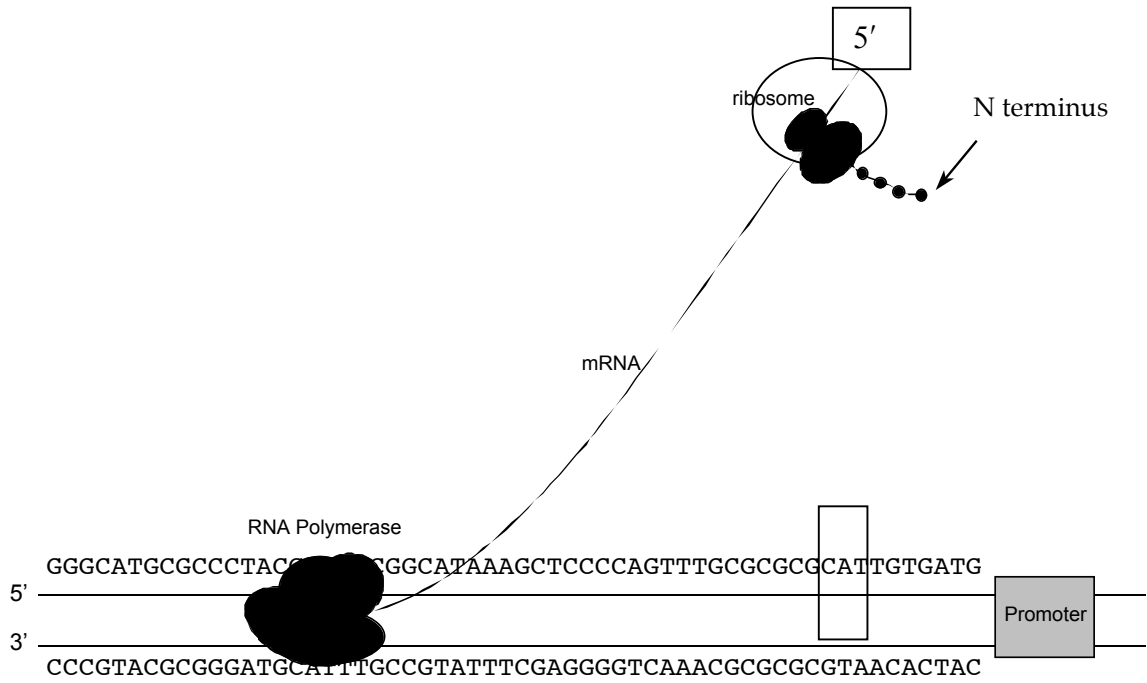
- From the first round of experiments we know that  $WW = \text{Met}$ ,  $XX = \text{Val}$ ,  $YY = \text{Thr}$ , and  $ZZ = \text{Leu}$ .
- From the next round you know that either  $WX$  or  $XW = \text{Ile}$  and the other is  $\text{Glu}$ . Also,  $(WY)_n$  encodes only one amino acid, so it must also encode a stop codon.
- Since from the final experiment you know that  $(WXY)_n$  gives  $\text{Glu}$  but not  $\text{Ile}$ , then  $WX = \text{Glu}$ . Since  $(WXY)_n$  encodes  $\text{Lys}$ , then  $YW = \text{Lys}$ , and  $WY = \text{stop}$ . By similar reasoning, the other codons can be determined:

Codon	Amino Acid
WW	Met
WX	Glu
WY	stop
WZ	Phe
XW	Ile
XX	Val
XY	Gln
XZ	Pro

Codon	Amino Acid
YW	Lys
YX	Asp
YY	Thr
YZ	Trp
ZW	Arg
ZX	Ser
ZY	stop
ZZ	Leu

(4.1.6)

a)



b) 3' ACC 5'

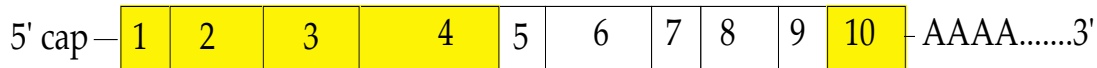
c) Yes. Any substitution will produce a different codon, and the only codon for Trp is 5' UGG 3'.

d) No. There are four codons for Thr. For example, 5' ACA 3' could be changed to 5' ACG 3' and still encode Thr.

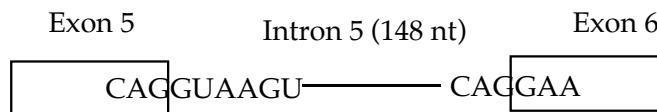
## (4.2) Transcription, RNA processing, and translation in eukaryotes

### (4.2.1)

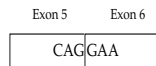
a)



b) i) before splicing:



after splicing:



ii) CAG = glutamine 206, GAA = glutamic acid 207

c) i)



ii) The mutation eliminates the 5' splice site in intron 5. As a result, this intron will no longer be spliced out of the processed message. Three possible outcomes (you only had to give two):

- 1) The unexcised intron will form a larger protein that will disrupt the protein structure.
- 2) The unexcised intron will introduce a stop codon that is in frame in intron 5 results in a truncated protein that is inactive since it will lack part of the NAD<sup>+</sup>-binding domain, as well as part of the catalytic subunit.
- 3) The unexcised intron sequence is translated, causing a frameshift mutation (148 nucleotides = 49 codons + 1 base). The larger protein will also be inactive since it will no longer be able to properly bind NAD<sup>+</sup>.







d) The 3' splice junction of intron 1 is lost. The next potential splice site is the CAG (nt 59-61). This changes the 5' end of the mRNA. Seven nucleotides have been lost between underlined bases; the rest of the mRNA remains the same.

5' : G<sub>cap</sub>-AAGGCGCUGAAC**AUG** CCU CUC UUU AAG GUA CUA CUA ACU AAU GAU  
GGU **UGA** CGAUACCCUCGGAAUAAAUA~~AAAAAA~~ :3'

This gives a protein sequence (the same sequence, just started at the second AUG):

H<sub>3</sub>N<sup>+</sup>-Met-Pro-Leu-Phe-Lys-Val-Leu-Leu-Thr-Asn-Asp-Gly-COO<sup>-</sup>

e) The 5' splice junction of intron 1 is lost. Therefore, intron 1 is not spliced out at all. The rest of the mRNA remains the same. This adds a new AUG codon in a different frame, which changes the protein sequence. Added nucleotides are underlined.

5' : G<sub>cap</sub>-AAGGC GCUGG UAUG,U CC,GAA UAGCA UGCAG AACAU GCCUC etc.

The resulting protein sequence: H<sub>3</sub>N<sup>+</sup>-Met-Ser-Glu-COO<sup>-</sup>

f) The 3' splice site in intron 2 is lost. Therefore, the splicing machinery will search for the next CAG or UAG it can find. The next CAG is at the 3' end of intron 3. This results in a deletion of all the sequences of exon 2. The deletion occurs between the underlined nucleotides:

5' : G<sub>cap</sub>-AAGGC GCUGC **AUG**,CA G,AAC,A UG,CCU CUU,AA C,**UAAU** GAUGG  
UUGAC GAUAC CCUCG GAAUA AAUAA AAAAA :3'

The resulting protein sequence (a truncated protein):

H<sub>3</sub>N<sup>+</sup>-Met-Gln-Asn-Met-Pro-Leu-Asn-COO<sup>-</sup>

g) This creates a new 3' splice site within the intron. This adds five nucleotides to the mRNA (underlined):

5' : G<sub>cap</sub>-AAGGC GCUGC **AUG**,CA G,AAC,A UG,CCU CUC,UU U,AAG,G UA,CUA  
CUA,CA G,**UAA**C UAAUG AUGGU UGACG AUACC CUCGG AAUAA AUAAA  
AAAA :3'

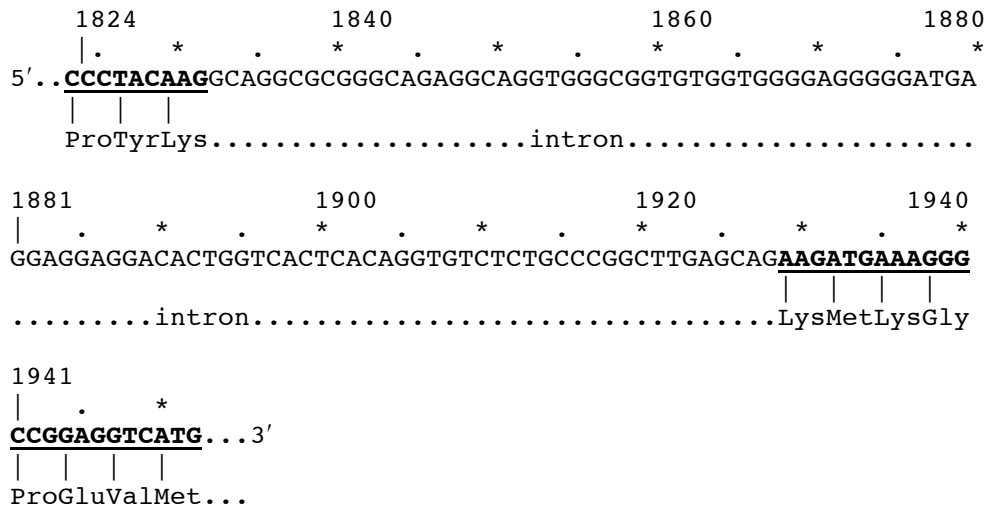
This causes premature termination of the protein chain:

H<sub>3</sub>N<sup>+</sup>-Met-Gln-Asn-Met-Pro-Leu-Phe-Lys-Val-Leu-Leu-Gln-COO<sup>-</sup>

h) This mutation destroys the promoter TATAA sequence. Therefore, there will be no transcription of the gene. No mRNA will be made, and no protein will be made.

(4.2.4)

a) If you start with the first three nucleotides, CCC, they encode proline, then tyrosine (TAC), and lysine (AAG). However, the next three nucleotides, GCA, do not encode lysine. The next lysine codon (AAA or AAG) starts at position 1929. After that point, the coding sequence is uninterrupted for the region of the protein shown. This is shown below (coding sequences shown **bold and underlined**). The noncoding sequence must be the intron, which begins at position 1833 and ends with position 1928.



b) The intron starts and ends with (5') GC.....intron.....AG (3'). Even though this does not follow the "rule," the protein sequence data are more compelling, since it bears directly on where the splice sites are located. The "rule" is only a consensus based on many sequences; there are always variations.

c) i) Since a single base-pair substitution caused three amino acids to be inserted in the region from which the intron was removed, it is logical to suppose that some alteration of the splicing signals has caused less of the intron to be removed than would be removed in wild type. Therefore, the start of the intron (5' splice site) must have moved to the right or the end of the intron (3' splice site) has moved to the left. Looking at the sequence, the intron 3' splice site must have moved nine nucleotides to the left to encode the amino acids given in the problem. This is shown below (the new coding region is shown **bold and underlined**). The first and last nucleotides of DNA that encode the mutant intron are 1833 and 1919, respectively.

ii) To make the 3' splice site be at position 1919, nucleotide 1918 must have changed from a G to an A, making a new 3' splice site. This base is shown in *outline type* on the next page.

```

      1824          1840          1860          1880
      | . * . * . * . * . * . *
5' ..CCCTACAAGGCAGGCGCGGGCAGAGGCAGGTGGGCGGTGTGGTGGGGAGGGGATGA
      | | |
      ProTyrLys.....intron.....

1881          1900          1920          1940
| . * . * . * . * . *
GGAGGAGGACACTGGTCACTCACAGGTGTCTCTGCCCAGCTTGAGCAGAAAGATGAAAGGG
      | | | | | | | |
.....intron.....LeuGluGlnLysMetLysGly

1941
| . *
CCGGAGGTCATG...3'
| | | |
ProGluValMet...

```

### Computer Activity 2: Gene Explorer (GeneX)

1) There are no solutions for this part.

2) a) Any of the following bases: 20-26, 55-67, or 92-101. These are in the “coding region”; you could also have called this part the “exon,” but that term is less precise.

b) This is not possible. Except for the poly(A) tail, all the bases in the mature mRNA must have counterparts in the pre-mRNA. In the normal gene, the poly(A) tail is not translated.

c) Any of the following bases: 27-54 or 68-91. These are in the introns.

d) Any of the following bases: 11-19 or 102-114. These are the untranslated regions of the exon.

3) a) They are not encoded in the DNA; they are added by an enzyme called poly(A) polymerase without using base-pairing. This is an interesting exception to the DNA/RNA rules: these A’s are added to the 5’ end of the mRNA in a 5’ to 3’ direction; they are also added without base-pairing—there are no T’s in the DNA that they pair with.

b) Because the splice junctions are seamless—there is no sign that the intron was there—so the ribosomes read right over where the splice junction was.

c) Intron 1 is 28 nucleotides long (not a multiple of 3); Intron 2 is 24 (a multiple of 3). Since introns are spliced out before being read by the ribosome, the length of the intron does not need to be a multiple of three nucleotides.

### Part III

4) and 5) You can make a single-base insertion, deletion, or substitution without changing the protein sequence in the following locations:

- Before the promoter: bases 0-5. This is not part of the gene.
- After the terminator: bases 120-127. This is also not part of the gene.
- In the introns, as long as your mutation doesn't alter the splice sites at the start and end of the intron or create a new splice site: bases 32-49 and 73-86. These do not code for the protein since they are not part of the mature mRNA.
- In the untranslated parts of the exons, before the start codon or after the stop codon, as long as they don't make a new start codon: bases 11-19 and 105-114. These do not code for protein.

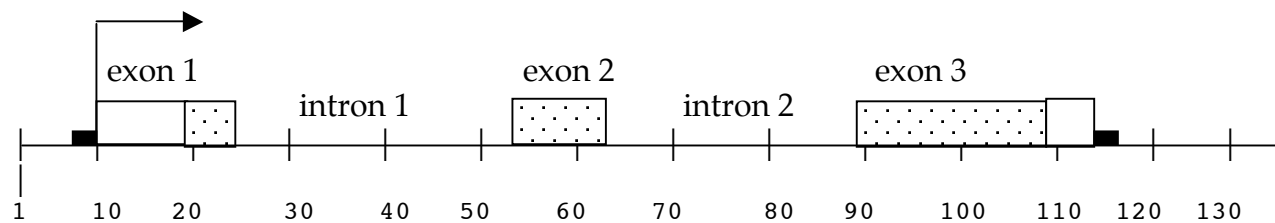
You cannot make insertion or deletion mutations in the coding region without changing the protein sequence. Substitution mutations are allowed at certain sites in the coding region if they don't change the amino acid (silent mutations): for example, changing base 101 from T to A.

8) N-Met-Pro-Cys-Arg-Met-Ser-Ser-Glu-Asp-Leu-Lys-Lys-Val-C

Everything after the Pro at position 2 is altered.

This is a FRAMESHIFT mutation.

9) This is the same map as the normal gene with only the position of the stop codon changed.



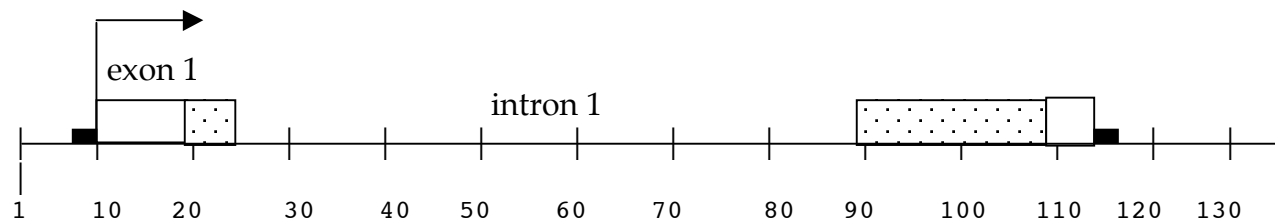
10) The deleted base changed the reading frame.

13) N-Met-Pro-Cys-Glu-Asp-Leu-Lys-Lys-Val-C

Everything after the Pro at position 2 is altered.

This type of mutation does not have a name.

14) Note that this map is very different from the normal gene.



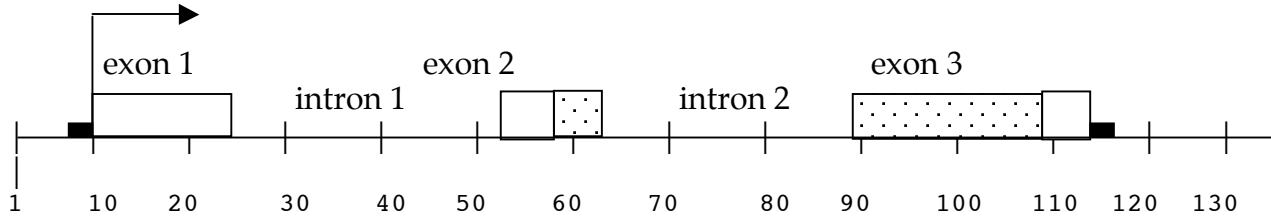
15) The mutation destroyed the "end of intron 1" (a.k.a. "3' splice site") sequence so the splicing machinery went looking for the next downstream "end of intron" sequence. It therefore skipped exon 2 entirely!

18) N-Met-Ser-Ser-Glu-Asp-Leu-Lys-Lys-Val-C

Everything is altered.

This type of mutation does not have a name.

19) (same map as before with just the position of start and stop changed)



20) This mutation destroyed the start codon so the ribosome went looking for the next AUG. This makes a totally new protein with a different reading frame and a different stop codon. Note that exon 1 is completely nontranslated.

(22) and (23) To have this effect, the mutation must change a regular amino acid codon to a stop codon. There are several such mutations, for example, changing 55 from T to A; changing 58 from C to A; and changing 96 from G to T.

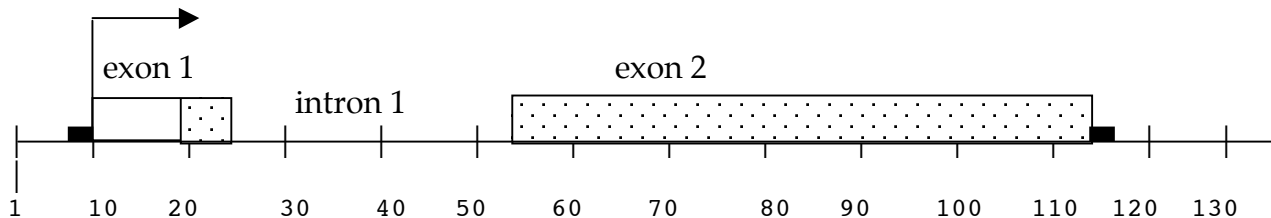
(25) and (26) To have this effect, the mutation must be in the “start intron” signal at the 5’ end of an intron. That is:

- Changing bases 27-31. These eliminate the “start intron” signal for intron 1; this gives a mature mRNA about 92 nt long because intron 1 is not spliced out.
- Changing bases 68-72. These eliminate the “start intron” signal for intron 2; this gives a mature mRNA about 87 nt long because intron 2 is not spliced out.

Note that altering the “stop intron” signal at the 3’ end of an intron (50-54 and 87-91) causes the following exon to be skipped so the mutant mRNA is shorter than normal.

(28) and (29) Only a mutation in the promoter or terminator will completely abolish mRNA and protein synthesis from the gene.

31) Delete base 68. This knocks out the start of intron 2 so it is not spliced out. This adds many more amino acids to the protein. The resulting gene is:



33) For each of these, there are several possibilities; we will describe only one.

a) Delete 26. This causes a frameshift. The resulting reading frame does not stop until the last codon in exon 3 and produces a protein that is three amino acids longer than normal.

b) Click on 60 and insert a T. This produces a stop codon that results in a shorter protein.

c) Delete 52. This inactivates the “end intron 1” sequence. This causes the splicing machinery to skip over exon 2 to the next “end intron” sequence.

## (5) CHALLENGE PROBLEMS

### (5.1)

a) With a circular chromosome, the DNA is continuous—it has no end. This means that there will always be DNA from which to make the RNA primer for the lagging strand.

b) A small piece of the telomeric DNA is lost during replication, so telomeres do not solve the problem of shrinking chromosomes but lessen the impact. The loss of telomeric DNA does not cause harm to the organism, because the telomeric DNA does not encode any genes. Its function is to protect the rest of the chromosome from being slowly lost (from the ends inward) during successive rounds of replication. The shrinkage of the telomere is later compensated by the action of telomerase.

### (5.2)

a) Three key differences between DNA and RNA are:

1) DNA is usually double stranded whereas RNA is mainly single stranded.

2) DNA contains thymine and RNA contains uracil.

3) The ribose component in DNA lacks a 2'-OH group; however, this hydroxyl group is present in the ribose moiety found in RNA.

b) Hydrogen bonds allow base pairing to occur between the two strands of DNA in a double-stranded DNA helix.

c) You would expect sequence (ii) to denature at a higher temperature. It contains a higher number of G/C base pairs. Each G/C base pair is held together by three hydrogen bonds, as opposed to two hydrogen bonds for A/T base pairs.

(i)

```
ATAGTATTC
|||||
TATCATAAG
```

(ii)

```
GACGCGGTG
|||||
CTGCGCCAC
```

d) The hydrophobic portions of each base interact with the hydrophobic regions of bases stacking above and below in the helix. These van der Waals and hydrophobic interactions contribute significantly to the structural stability of a double-stranded DNA molecule.

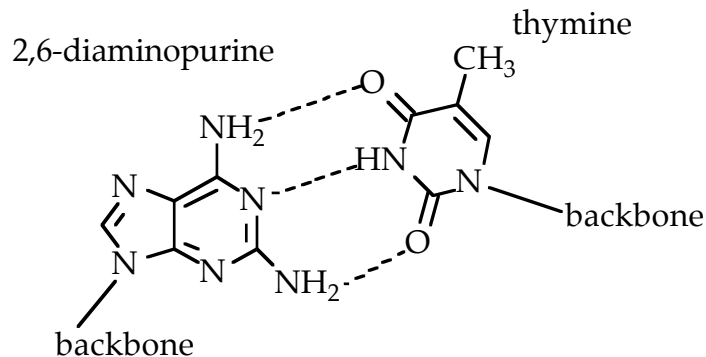
e) The A and G bases are purines that have two aromatic rings and are larger than the pyrimidines. The T and C bases are pyrimidines that have one aromatic ring. The distance between the backbones of each strand of the double helix accommodates a purine hydrogen bonded to a pyrimidine. A base-pair mismatch of A-G would increase the distance between the backbones, and a base-pair mismatch of C-T would decrease the distance between the backbones.

f) In a cell that is at physiological pH, the overall charge of a double-stranded DNA molecule is highly negative because of the charged phosphate groups on the DNA backbone. (The pKa of the phosphate group is  $\sim 1.0$ ; hence, the reason why the molecule is called deoxyribonucleic acid. In a cell the negative charges are neutralized by NaCl.)

g) DNA replication should halt or be slowed down by the addition of the AZT nucleotide. Since the AZT nucleotide lacks a 3'-OH group, every time it is incorporated into an elongating DNA strand it stops DNA replication after that point.

If AZT is incorporated into the DNA of cells that are undergoing cell division, it would inhibit DNA replication, thus leading to cell death. Cancer cells divide rapidly and thus would be most highly affected by AZT. The HIV nucleic acid polymerase preferentially incorporates AZT into its genome, thus inhibiting the replication of the viral DNA.

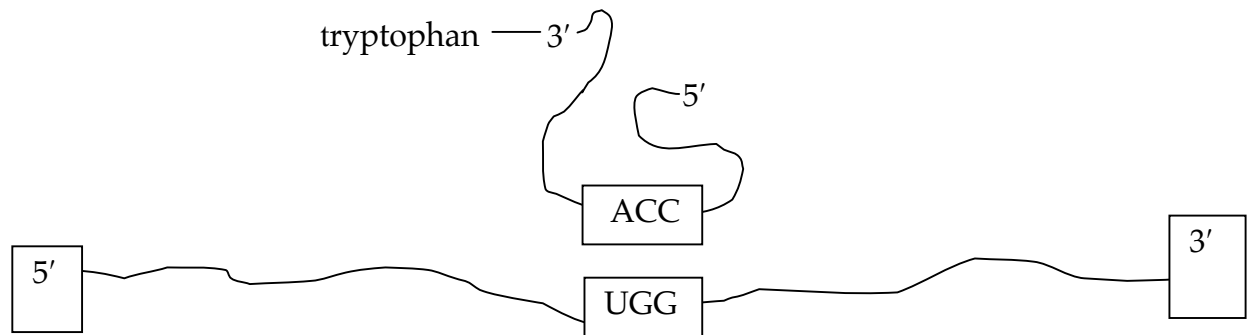
h) The base 2,6-diaminopurine would most likely form a base pair with thymine.





(5.3)

a)



b) The lower strand is used as the template strand. When transcribed, the sequence for the anticodon in the tRNA should be 5' CCA 3'. Therefore, the lower strand must be used as the template:

3' ...GGT... 5' used as template

5' ...CCA... 3' tRNA