Name: _____ TA:_____

7.013 Problem Set 3 FRIDAY October 8th, 2004

Problem set answers must be inserted into the box outside

Problem sets will NOT be accepted late. Solutions will be posted on the web.

Question 1

a) Below is a schematic of an inset electron micrograph showing simultaneous transcription and translation of a gene in *E. coli.* Select the names from the following list of structures depicted in the schematic and write them on the lines adjacent to their symbol. (Use each term only once; you do not need to use all of them.) Some of the terms are explained in the book, i.e. page 265.

1.	N-terminal of polypeptide	5.	5′ end of mRNA	9.	messenger RNA	13.	DNA
2.	C-terminal of polypeptide	6.	ribosome	10.	GTP cap	14.	5′ end of DNA
3.	3′ end of mRNA	7.	polypeptide	11.	3′ end of gene	15.	DNA polymerase
4.	Shine-Delgarno sequence	8.	RNA polymerase	12.	promoter	16.	stop codon

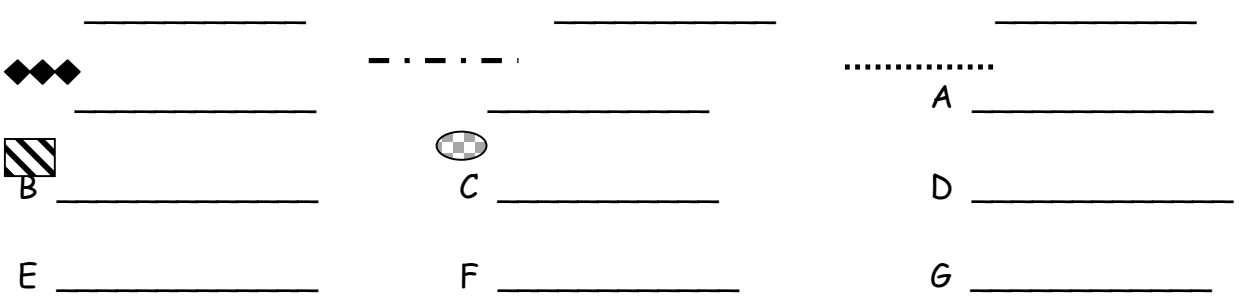


Image removed due to copyright reasons.

b) According to the schematic, in which direction are the structures symbolized by the checkered ovals moving?

up down not moving left right

c) In which direction are the structures symbolized by the striped boxes moving?

up down not moving left right

Question 2

Assume mRNA is being transcribed starting from the far left side of the following double stranded DNA template.

5'GTGCTAGCGGGAATGAGCTGGGATACTAGTAGGGCT3'
3'CACGATCGCCCTTACTCGACCCTATGATCATCCCGA5'

a) What are the first five nucleotides of the mRNA sequence? _____

b) What are the first 5 amino acids encoded? _____

c) The following sequences show (in bold) different mutations affecting the above DNA sequence. Assume none affect the expression of the mRNA synthesis.

A. 5'GTGCT**G**AGCGGGAATGAGCTGGGATACTAGTAGGGCT3'
3'CACGA**C**TCGCCCTTACTCGACCCTATGATCATCCCGA5'

B. 5'GTGCTAGCGGGAATGAGCTG**C**GGATACTAGTAGGGCT3'
3'CACGATCGCCCTTACTCGAC**G**CCTATGATCATCCCGA5'

C. 5'GTGCTAGCGGGAATGAGCTG**A**GATACTAGTAGGGCT3'
3'CACGATCGCCCTTACTCGAC**T**CTATGATCATCCCGA5'

D. 5'GTGCTAGCGGGAATGAGCTGGGA**A**ACTAGTAGGGCT3'
3'CACGATCGCCCTTACTCGACCCT**T**TGATCATCCCGA5'

E. 5'GTGCTAGCGGGAATGAGCTGGGA**C**ACTAGTAGGGCT3'
3'CACGATCGCCCTTACTCGACCCT**G**TGATCATCCCGA5'

F. 5'GTGCTAGCGGGAATGAGCTGG**C**ATACTAGTAGGGCT3'
3'CACGATCGCCCTTACTCGACC**G**TATGATCATCCCGA5'

WT 5'GTGCTAGCGGGAATGAGCTGGGATACTAGTAGGGCT3'
3'CACGATCGCCCTTACTCGACCCTATGATCATCCCGA5'

i) For the above mutations, fill in the following box.

Sequence	Type of mutation Choose from insertion, deletion, substitution.	Effect on protein Choose from missense, nonsense, frameshift, silent.
A		
B		
C		
D		
E		
F		

ii) Order the mutations according to the likelihood that they will result in an inactive protein, from most likely to least likely. If you think two mutations have an equal likelihood of resulting in an inactive protein, write an equal sign between them. Your answer should be in the format X>Y>Z=V=WT(wild type).

The Genetic Code

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

Question 3

Part I

In *E. coli*, when glucose is present, galactose is used to make a component of the bacterial cell wall. When glucose is absent, galactose is MOSTLY used to make energy, but a little is used to make the cell wall. The enzymes for the conversion of galactose into a cell wall component are transcribed as an operon. The operon is transcribed at a high level when glucose is present and at a low level when glucose is absent. In many organisms, this type of regulation is accomplished by having two promoters in front of the operon, one that is "strong", meaning it causes a high level of transcription, and one that is "weak", meaning it causes a low level of transcription. In this problem, we will be learning how the transcriptional regulation of this type of operon occurs.

You isolate mutants that do not respond appropriately to the presence or absence of glucose, as summarized below. **Assume each mutant has a single point mutation. Assume all mutations are disabling. The repressor is under its own promoter, and this promoter is not affected by any of the mutations.**

	Galactose incorporation into cell wall	
strain	+ Glucose	– Glucose
Wild type	High	Low
Mutant 1	High	High
Mutant 2	High	High
Mutant 3	Low	Low
Mutant 4	None	None
Mutant 5	High	None
Mutant 6	None	None

a) Which mutant(s) could have a mutation in either promoter? *Keep in mind what the phenotypes of the mutants would be if there were both a strong promoter and a weak* promoter.

_____

b) Which mutant(s) could have a mutation in an operator? _____

c) In a repressor? _____

d) In a protein coding region? _____

Part II

To determine which elements are cis-acting and which are trans-acting, you consruct *E. coli* that is diploid for the DNA encoding this operon (merodiploid), and observe how the mutations behave.

In this problem, the mutant number will correspond to the element that contains the mutation. An element in this problem can be a gene, promoter, operator, or a repressor. For example, mutant 6 would have an mutation in element 6. If the data indicated that mutant 6 has a mutation in an operator, then for part i) below, that operator would be numbered 6.

Measuring the activity of galactose incorporation into the cell wall, you observe the following results. A "+" means that the element corresponding to a particular mutant is wild type, whereas a "–" means that the element is mutated. For example, 3- 2+ means the DNA sequence has a mutation in element 3, is wild type in element 2, and all other parts of the operon and its regulatory elements are present and wild type on both DNA copies.

	Incorporation of Galactose into cell wall	
Strain	+ Glucose	- Glucose
$\frac{4^-\ 6^+}{4^+\ 6^-}$	High	Low
$\frac{3^-\ 4^+\ 6^+}{3^+\ 4^-\ 6^-}$	Low	Low
$\frac{1^-\ 4^+\ 6^+}{1^+\ 4^-\ 6^-}$	High	High
$\frac{5^-\ 4^+\ 6^+}{5^+\ 4^-\ 6^-}$	High	None
$\frac{2^-\ 4^+\ 6^+}{2^+\ 4^-\ 6^-}$	High	Low

_____ a) Which elements are *cis*-acting?

_____ b) Which elements are *trans*-acting?

_____ c) Which element is the strong promoter?

_____ d) Which element is the weak promoter?

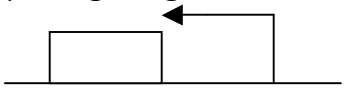
_____ e) Which element is the operator?

_____ f) Which promoter does it regulate?

_____ g) Which element is the repressor?

_____ h) How many proteins are encoded by this operon?

i) Draw a picture of this operon, with numbers corresponding to the elements identified by the mutant strains. Remember that the repressor has its own promoter and is in a separate operon. Also place the number corresponding to the repressor in the box depicting its gene drawn below.



Repressor

Question 4

You are studying several compounds (A, B, C, D, and E) that were shown to affect expression of secreted proteins in canine kidney cells. You've established an experimental system by fusing a signal sequence to green fluorescent protein (GFP) that you've introduced into canine cells. With this fusion in the strain, you'll be able to measure green fluorescence in the growth medium after removal of the cells. Your goal is to determine how each compound affects expression of the GFP gene fusion. To this end you assay individual cell cultures treated with each of your compounds in the following ways.

You measure fluorescence in the medium.

- All cultures treated individually with each compound fail to show fluorescence in the medium. Untreated cultures exhibit fluorescence in medium.

You measure intracellular GFP mRNA.

- The mRNA is detected in all of your cultures except the one treated with compound A.

You fractionate intracellular subcompartments and assay for the presence of GFP mRNA.

- After separating the cytoplasmic fractions from the nuclear fractions of the cells, you fail to detect GFP mRNA in the cytoplasm of cells treated with compound B.

You observe the cells under a fluorescent microscope.

- You fail to see fluorescence in cells treated with compound C.
- You see GFP fluorescence *within* the cytoplasm of cells treated with compound D.
- You see GFP fluorescence *within* the endoplasmic reticulum of cells treated with compound E.

Your advisor is able to continue this molecular analysis that will precisely determine the gene expression defect caused by each compound, but this research is expensive. You got a D in 7.012 last semester and the only way you can get into Harvard Medical School is to rise to the occasion and shine by accurately predicting how each compound affects gene expression. This will ensure your name on a publication of this research project. Your advisor awaits your hypotheses.

a) What is the most likely mechanism affected by compound A to halt GFP expression in your canine cells?

b) What process is affected by compound B? Why would this prevent secretion of GFP?

c) What process is affected by compound C?

d) Give an example of what might be a target for compound D.

e) What process is affected by compound E?