# Solution key- 7.013 Problem Set 4- 2018

Question 1 (3pts)
Cardiovascular disorders such, as heart attack and stroke are among the top ten causes of death in the western hemisphere. Stroke patients, under certain circumstances, have better outcomes when treated with tissue plasminogen activator (TPA).

a) TPA is **a serine protease** that catalyzes the **hydrolysis** of plasminogen to plasmin. Plasmin is an essential enzyme for clot breakdown (hydrolysis of fibrin protein to Fibrinogen degradation product (FDP)).
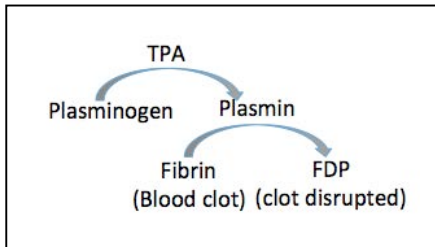
i. How would you categorize the TPA catalyzed reaction: **exergonic or endergonic**? *Exergonic, since peptide bonds between the amino acids are being hydrolyzed (0.25pts)*

ii. Which covalent bond in plasminogen does TPA hydrolyze: **ester bond/ peptide bond/ glycosidic bond/ phosphodiester bond**? *(0.25pts)*

b) In humans, the *PLAT* gene encodes the TPA protein.

i. Would you expect all the cell types in the body to have the *PLAT* gene? **Why or why not?**
Yes, ALL cell types in the body originate through somatic cell divisions starting with a zygote. Therefore almost all cell types (debarring the spontaneous mutations) have the same genome. But each cell type expresses its own unique set of genes. (0.5pts with 0.25 for explanation)

ii. The *PLAT* gene is expressed only in endothelial cells that line the blood vessels. How might *PLAT* gene expression be regulated in the endothelial cells? Propose **one** explanation.
-Endothelial cells have the PLAT gene specific transcription factors unlike other cell types.
-Part of the chromatin structure that includes PLAT gene lacks the histones and hence is open state
- The bases in the transcribed or regulatory region of PLAT gene are not methylated, which promotes histone removal and modification. (0.5pts with 0.25 for explanation)

c) Considering the therapeutic potential of TPA, you want to produce a large amount of recombinant human TPA (rTPA) in bacterial cells using the recombinant technologies that you have just learned about in 7.013.

i. You suggest cloning the *PLAT* gene in bacteria to express it and isolate PLAT protein. Your UROP supervisor suggests that you instead clone the *PLAT* cDNA. Which strategy is better? **Explain** why. *(0.5pts with 0.25 for explanation)*
The PLAT gene would have untranscribed regulatory sequence (such as promoter) and transcribed region that includes both introns and exons. In bacteria, the introns will not be spliced out, so the resulting protein will likely be nonfunctional. The mature mRNA will only have the exons and no introns.

ii. Which human cell types would you use for preparation of human PLAT cDNA and **why**?
The endothelial cells, since they express PLAT gene and should therefore have mature PLAT mRNA. (0.5pts with 0.25 for explanation)

iii. Give the sequence of the 5 bases long primer that you would use to make single-stranded *PLAT* cDNA and label its 5' and 3' ends. How can you be sure that this primer will base pair with the mRNA?
*All mRNAs (pre-spliced and spliced), in a cell will have 3'Poly A tail with which the **oligo-T primer** can undergo complementary and anti-parallel base pairing. The 3'OH end of the last T in the oligo-dT primer can then be extended to make the first DNA strand of cDNA that is complementary to the mature mRNA. (0.5pts with 0.25 for explanation)*

Question 2 (3pts)

You generate *PLAT* cDNA and clone it into the plasmid diagrammed below.
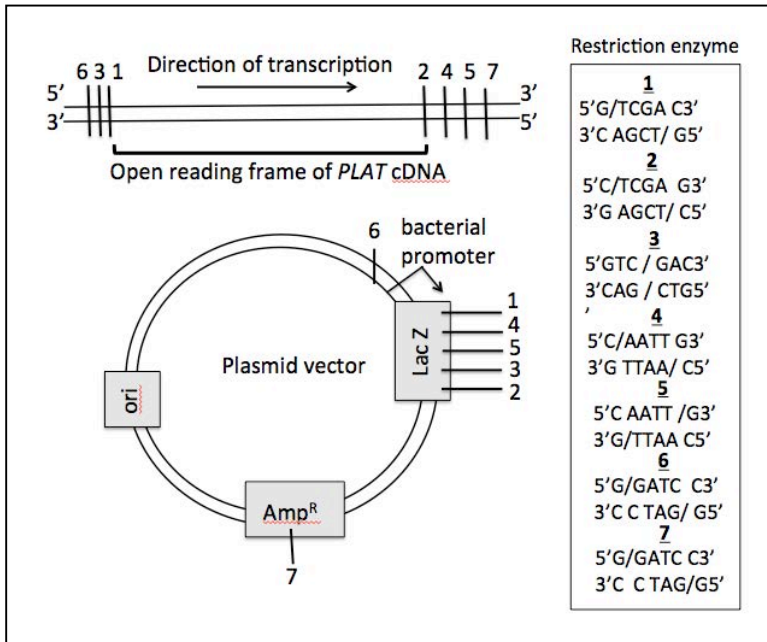
Note: *Both the PLAT cDNA and plasmid vector have recognition sites for restriction enzymes 1-7. A slash (/) represents the cut site for each restriction enzyme. The plasmid has a bacterial origin of replication (ori), an ampicillin resistance gene (Amp^R) that degrades ampicillin antibiotic. It also has LacZ gene that encodes the β-gal enzyme, which converts white X-gal substrate into a blue product.*

a) The table below shows different restriction enzyme pairs that you can use to digest the *PLAT* cDNA and clone it in the bacterial plasmid. Complete the table below for each restriction enzyme pair.

Restriction enzyme pair used to cut the *PLAT* cDNA	Will you be able to….		
	Column 1: Clone the digested *PLAT* cDNA into the plasmid: **Yes/ No**?	**Column 2:** Amplify the recombinant plasmid in the ampicillin sensitive bacterial cells: **Yes/ No**?	**Column 3:** Express the *PLAT* cDNA in the bacterial cells: **Yes/ No/ maybe**?
1 & 5	*Yes*	*Yes*	*Yes*
1 & 2	*Yes*	*Yes*	*Maybe*
2 & 6	*Yes*	*Yes*	*No*
3 & 7	*Yes*	*No*	*No*

(0.5pts for each row, 2pts total)

b) If you selected "maybe" for any enzyme pair in Column 3 in the table above, please provide an explanation for your selection. *(0.5pts for explanation)*

Restriction enzymes 1 and 2 generate complementary overhangs. So the PLAT cDNA using this restriction enzyme pair can be cloned in the plasmid vector in two ways: one that is oriented correctly with respect to the bacterial promoter and will express PLAT cDNA and the other which is oriented in a direction that is opposite to the promoter and will therefore not express PLAT cDNA.

c) After transformation, you grow the bacterial cells. Briefly explain how would you find the bacterial cells that are transformed with plasmid containing the *PLAT* cDNA insert.

You would plate the bacterial cells transformed with the ligation mix on a growth medium that has X-gal and ampicillin. The bacterial colonies that will grow and will be white are the ampR and LacZ- and therefore will have the cloned PLAT cDNA insert. (0.5pts)

Question 3 (3pts)

Your fellow classmate plans to construct a GFP-PLAT fusion protein. ***Note:*** *GFP stands for the Green fluorescent protein.*
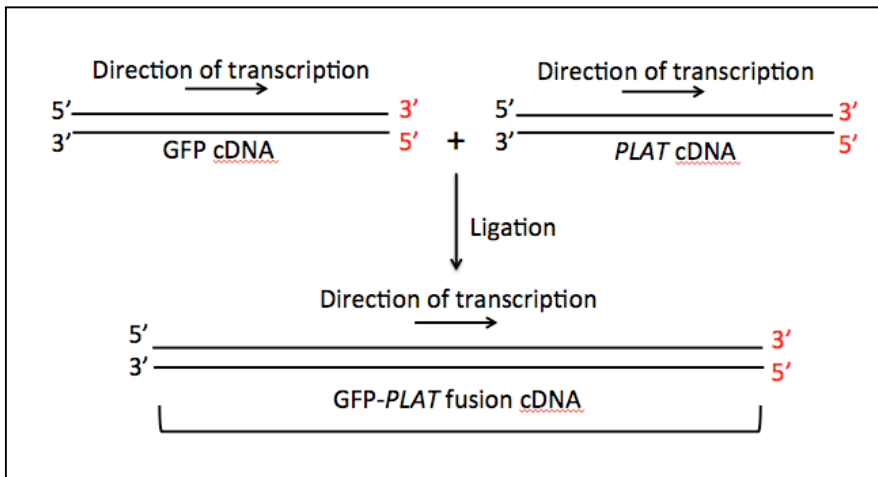
The following is the partial cDNA sequence encoding the C- terminus of the GFP gene. ***Note:*** *The DNA corresponding to the stop codon is bold and underlined. The sequence specifically recognized by each restriction enzyme is shown in gray. Each codon is separated from the next by a space.*

```
                                     1              2
                                 ─────────      ─────────
GFP:    5'AAG AAA ATT GCC GCG AAT TCC CCC TCG AGC ATG TAG3'
        3'TTC TTT TAA CGG CGC TTA AGG GGG AGC TCG TAC ATC5'
```

The following is the partial cDNA sequence encoding the N- terminus of the *PLAT* gene. ***Note:*** *The DNA corresponding to the start codon is bold and underlined. The recognition sequence for each restriction enzyme is shown in gray. Each codon is separated from the next by a space.*

```
             1           3             4
        ─────────    ─────────    ─────────
PLAT:   5'G GAA TTC GCA ATT GGA GGG TCG ACC ATG CCA....... 3'
        3'C CTT AAG CGT TAA CCT CCC AGC TGG TAC GGT........5'
```

The recognition sequences and the cleavage sites (*indicated by /*) for each enzyme are given below.

```
        3                  1                  2                  4
5'C/AATT  G3'      5'G/AATT C3'       5'C/TCGA G3'       5'G/TCGA C3'
3'G TTAA/ C5'      3'C TTAA/G5'       3'G AGCT/C5'       3'C AGCT/G5'
```

a) To make a *GFP-PLAT* fusion gene that expresses the GFP-TPA fusion protein, which restriction enzyme would you use to cut the…

i.   C terminus of the GFP gene? ***Enzyme 2*** *(1pt)*

ii.   N- terminus of the *PLAT* gene? ***Enzyme 4*** *(1pt)*

3

Question 3 continued

This GFP-*PLAT* fusion cDNA has restriction enzyme Y sites at each end. Your friend digests both the *GFP-PLAT* fusion gene and the plasmid vector diagrammed below with the restriction enzyme Y. She incubates the restriction enzyme Y digested DNA fragments with DNA ligase, transforms the Amp^S bacterial cells with the ligation mix and grows them on ampicillin containing plates. She decides to further analyze two bacterial colonies (Colony A and Colony B) both of which have the recombinant plasmid with the *GFP-PLAT* cDNA insert. **Note:** *The plasmid has recognition sites for restriction enzymes Y and X.*

To further examine the structure of the recombinant plasmid, your friend isolates the recombinant plasmid from Colony A and Colony B and digests it with restriction enzyme "X". She then resolves the sizes of the digested DNA fragments on a DNA gel and gets the profile shown below.
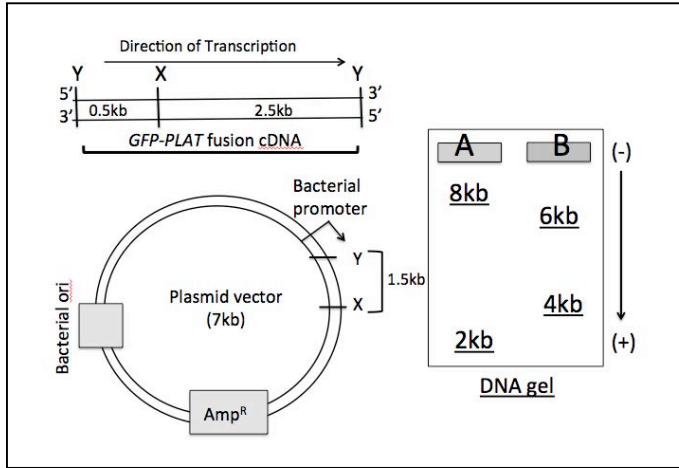

b) Which cell population will express the *GFP-PLAT* fusion cDNA: **Colony A** or **Colony B** and **why?**
The recombinant Plasmid in Colony B has the GFP-PLAT fusion gene oriented correctly along the direction of the promoter. So this colony will express the fusion gene and will fluoresce green.

(1pt, 0.5 for colony and 0.5 for the explanation)

Question 4 (3pts)

An increase in cholesterol and low-density lipoprotein (LDL) levels is associated with heart attacks and stroke. Cholesterol and cholesterol esters are packaged into LDL and transported throughout the circulation. When the LDL packaged cholesterol complex reaches liver cells, it binds to the Cell surface LDL-receptors (LDL-R). The LDL bound LDL-R complex is then endocytosed and further processed.
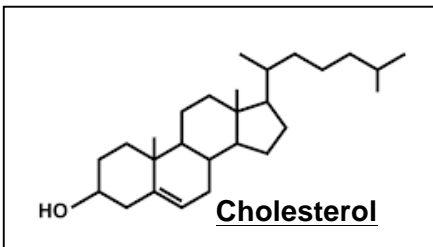
a) Why must cholesterol and cholesterol esters be packaged in LDL particles for transport?
*Cholesterol and cholesterol esters are hydrophobic and therefore cannot diffuse on their own in the circulation. Instead they are packaged into the LDL protein and transported. Also, the LDL-R bound to cholesterol can be endocytosed by the liver cells and further processed. (0.25pts)*

b) Give **a location** in the cell where cholesterol can be found.
The cell membrane and membranes of different organelles (No points)

4

Question 4 continued

Familial hypercholesterolemia (FH) is an inherited genetic disorder that leads to aggressive and premature cardiac diseases such as a heart attack or stroke. The following is a pedigree for FH that is caused by a mutation in the *LDL-R* gene. ***Note:*** *All affected individuals are shaded. Also listed are the SNPs for some individuals.*

1
2
(C, C)

3   4   5   6   7   8
(G, A) (A, C) (G, C)

9   10   11   12   13   14
(G, C)
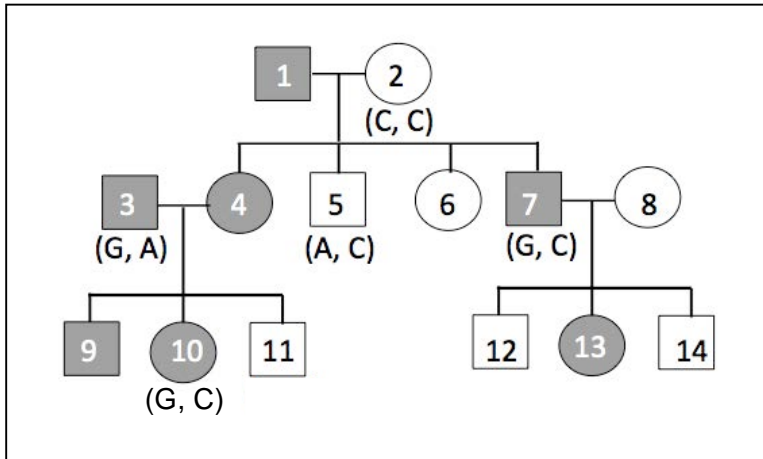
c) Assuming complete linkage between the *LDL-R* gene and the given SNP, list the **mode of inheritence** of FH.
Autosomal dominant (0.5pts)

d) Give the SNP profile(s) for each of the following individuals and **circle** the SNP that is associated with FH where applicable.

1: *G, A*   4: *G, C*   11: *A, C*

"G" SNP in 1 and 4 is associated with FH (1pt, 0.25pts for SNPs of each individual and 0.25 for identifying the disease related SNP)

e) You were introduced to **DNA and mRNA** expression microarrays. *(0.5pt, 0.25 for each part)*

   i. Which microarray (**DNA / mRNA expression microarrays**) would you use for SNP analysis and **why**? *You would use DNA microarray since the SNP may be present in the regulatory regions (such as enhancer or promoter regions), coding (exons) or noncoding (introns) regions.*

   ii. Which microarray (**DNA / mRNA expression microarrays**) would you use to analyze the rate of transcription of *LDL-R* gene? *You would use an mRNA expression microarrays to look at the variation in the expression profile of the LDL-R gene in the LDL-R gene expressing cells.*

f) You alleviate the symptoms of this disease in Individual #10 by using **somatic cell gene therapy**. Following treatment, this patient has a child with a normal healthy male. Would this treatment cause any change in the probability of the child having the disease? **Why or why not?**
#10 has FH and her genotype is "Aa". If she has a child with a normal healthy male genotype: aa) the probability of their child having FH would be ½. This probability will remain since this therapy will not change the genetic makeup of the gametes produced by individual #10. (0.25pt for explanation)

g) You fellow classmate wants to use the **C**lustered **R**egularly **I**nterspaced **S**hort **P**alindromic **R**epeats (CRISPR) and CRISPR associated endonuclease (Cas9) technologies to create an FH mouse model.

   i. Which gene sequence is recognized by the guide RNA? *This will recognize part of your gene of interest. It will specifically recognize 20 nucleotides upstream of the PAM sequence (NGG) at the 3' end of the DNA strand that you want to cut. (0.25pts)*

   ii. Propose one mechanism by which the Cas9-exonuclease generates mutations in the target gene to produce an FH mouse model. *Cas9 cuts the DNA leading to double stranded DNA breaks. During the repair process, the ends of the DNA at the breakpoints are processed and random base pairs maybe inserted or deleted resulting in a frame-shift or premature stop codon or an altered nonfunctional gene product. Because the gene will cease to function, the mouse model will be similar to FH where the LDL-R is either not expressed on non-functional. (0.25pts any reasonable explanation should be accepted)*

Question 5 (4pts)

You are studying a disease that shows an **autosomal recessive mode of inheritance** and is associated with a mutation in the *LRP* gene, which encodes LDL-R related protein (LRP). You clone a 3.0kb linear double stranded DNA fragment that includes the allele of the *LRP* gene that is associated with the wild-type phenotype (Fragment A) into a plasmid. You also clone the 5.0kb linear double stranded DNA fragment that includes the allele of the *LRP* gene that is associated with the disease phenotype (Fragment B).

You know that Fragment B differs from Fragment A by a single base pair insertion mutation.

The DNA sequence that flanks (at the two ends) Fragment A and Fragment B is given below.

```
5'- CACGTACCA _____ TTACGGACT-3'
3'- GTGCATGGT _____ AATGCCTGA-5'
            └_____┘
                      5kb
```
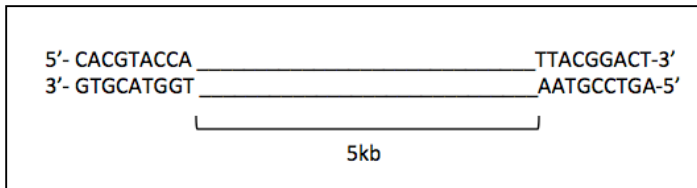
a) You use the polymerase chain reaction (PCR) to amplify Fragments A and B. Design the primers (each 5 nucleotides long) that you would use to amplify both strands of the 5.0kb long DNA fragments and label their 5' and the 3' ends.

  i. **Forward primer** to make the top strand: 5'- ***CACGTA***-3' *(0.25pts)*

  ii. **Reverse primer** to make the bottom strand: 5'- ***AGTCCG***-3' *(0.25pt)*

b) You PCR amplify Fragment A and Fragment B and digest them separately with a series of restriction enzymes. You resolve the restriction digested DNA fragments on two separate DNA gels and obtain the profiles below. ***Note:*** *The size of each digested DNA fragment is written in kilobases (kb).*

Based on the restriction pattern shown below, draw the restriction map of Fragments A & B. Indicate the restriction sites that are present and the distance (in kb) between the sites. *(0.5pts, 0.25 each)*

Restriction map of Fragment A from a healthy donor

Restriction map of Fragment B from a patient
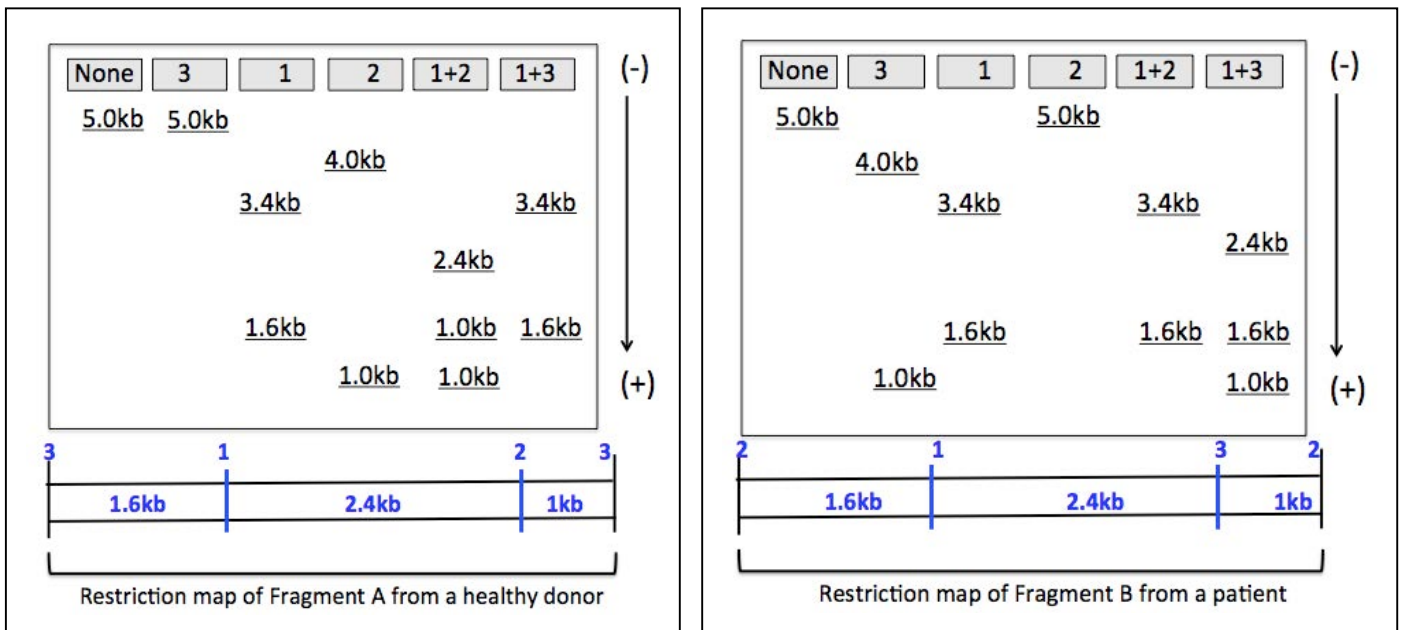
Question 5 continued
Using the fluorescence di- deoxy sequencing method, you sequence both Fragment A and Fragment B by using the **template DNA strand for sequencing**. The fprofiles below show the base sequence corresponding to the first 15 bases in the coding/ mRNA like strands of Fragment A (in schematic A) and Fragment B (in schematic B).

Note: *The sequence specifically recognized by restriction enzymes 1, 2 and 3 is given below and the site at which they cut is shown by a slash (/).*

DNA sequencing gels
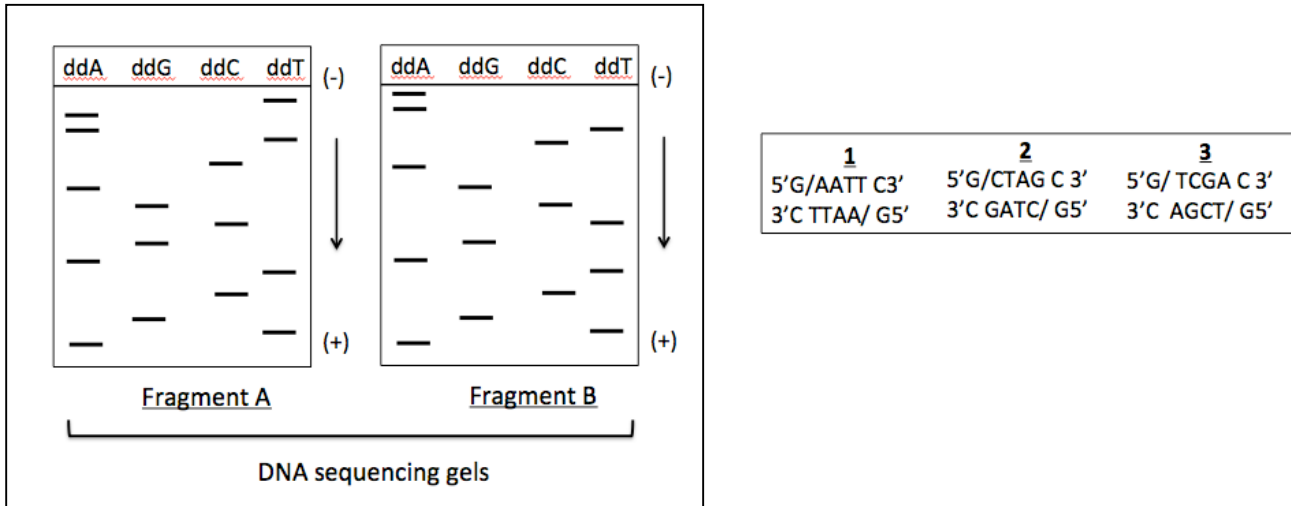
c) Write the sequence of the first 15 base pairs of the double helical DNA Fragments A and B and label their 5' and 3' ends. *(1pt, 0.5 each)*

- **Fragment A:** 5'ATGCTAGCGACTAAT3' *The site for Enzyme 2 is shaded*
  3'TACGATCGCTGATTA5'

- **Fragment B:** 5'ATGCTAG*T*CGACTAA3' *The site for Enzyme 3 is shaded*
  3'TACGATC*A*GCTGATT5'

d) Circle the recognition site for the relevant restriction enzymes (*1/2/3*) on Fragment A and fragment B in part (c) above. *(NO points)*

e) Write the amino acid sequence encoded by the coding sequence shown in schematic A and schematic B for Fragment A and Fragment B. Label the N and the C ends. ***Note**: A codon chart is provided on the last page of this problem set.*

- **Amino acid sequence for Fragment A: N-*met-leu-ala-thr-asn* –C** *(0.5pts)*
- **Amino acid sequence for Fragment B: N- *met-leu-val-asp-C (stop codon)*-C** *(0.5pts)*

f) On sequence that you gave in part (c), box the base pair in Fragment B that accounts for the point mutation that results in the disease phenotype. *It is the base pair shown in red (0.5pt)*

g) Identify the mutation that you circled in part (f) as **nonsense/ missense/ silent/ frame shift**. *It is a frame-shift that generates a premature stop codon (0.5pt)*

Question 6 (3pts)

It has been observed that the incidence of cardiovascular disorders in pre-menopausal females is significantly less than the age matched males. This is attributed to cardio-protective effects of estrogen hormone.

Estrogen hormone based signaling involves several receptors which can be classified as INTRACELLULAR receptors (n-ERs) and CELL MEMBRANE BOUND receptors (m-ERs). The pathway and its steps are outlined below.
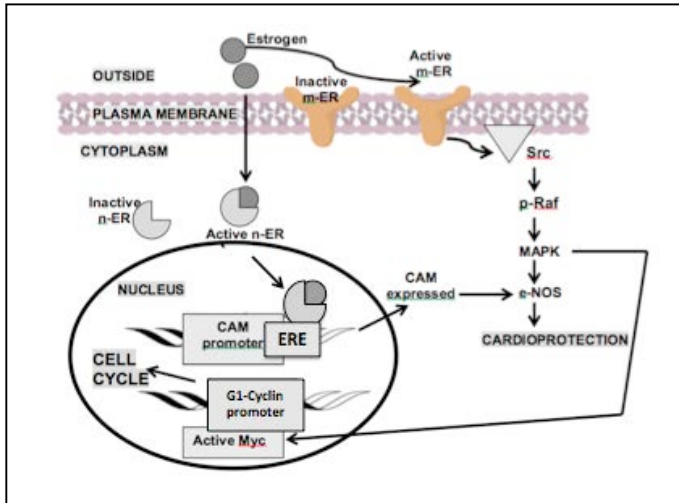
Step 1: Estrogen mediates its cardio-protective effects by binding to and activating the membrane estrogen receptor (m-ER).

Step 2: The activated m-ER activates Src kinase, which activates the Raf-MAPK kinase cascade. This results in the phosphorylation and activation of the eNOS enzyme, which provides cardio-protection.

Step 3: Active MAPK also activates Myc, which promotes cell cycle progression by triggering G1-Cyclin expression.

Step 4: Estrogen also binds to and activates the nuclear estrogen receptor (n-ER). The activated n–ER translocates to the nucleus and binds to a specific sequence called estrogen response element (ERE) on the DNA and acts as a transcription factor.

Step 5: The n-ER – ERE binding upregulates the transcription of CAM gene, which further provides cardio-protection.

Consider the following **mutations** in different components of the estrogen-signaling pathway.
(3pts: 1 pt for each row, 0.5 for the first three cells of the row and 0.5 for the next three cells)

- **Mutant #1:** The n-ER undergoes a loss-of-function mutation.

- **Mutant #2:** eNOS shows a deletion of the MAPK binding site.

- **Mutant #3:** MAPK is constitutively active.

- **Mutant #4:** SRC undergoes a gain-of function mutation

- **Mutant #5:** Promoter region of myc gene is heavily methylated

Complete the table for each of the following <u>**homozygous mutations**</u> in the <u>**presence**</u> of **estrogen**.

Mutants	m-ER activated (*Yes/ No*)?	CAM gene expressed (*Yes/ No*)?	Cardio-protection compared with wild-type (*same/ more/ less/ none*)?	G1-cyclin expressed (*Yes/ No*)?	Cell division (*Yes/ No*)?
#1 & #2	Yes	No	None	Yes	Yes
#3 & #4	Yes	Yes	More	Yes	Yes
#4 & #5	Yes	Yes	More	No	No

8

Question 7 (1pt)
The m-ER protein is shown as part of the schematic below. The start codon on the m-ER mRNA is shown.
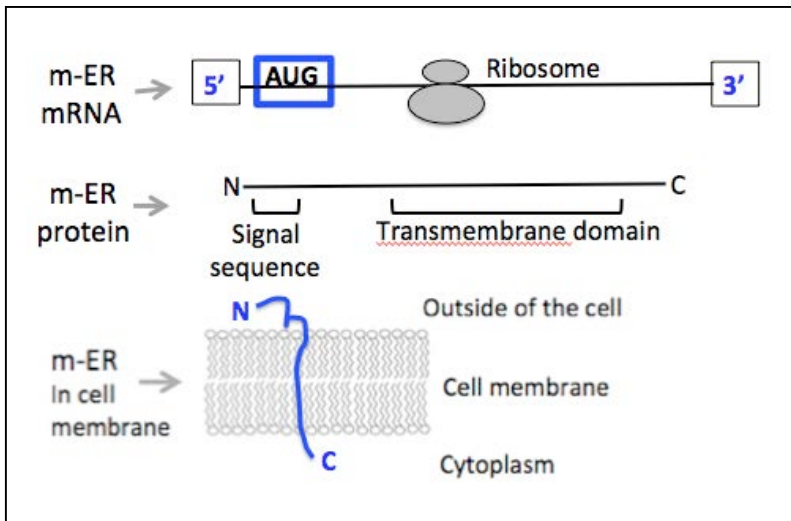
a) Label the 5' and 3' ends of the mRNA that codes for m-ER. *(0.5pts)*

b) On the **m-ER mRNA**, show the approximate location of the codons corresponding to the signal sequence by drawing a box. *(0.5pts)*

c) Which part of the **m-ER protein** is cleaved off before it is trafficked to the cell membrane?
The signal sequence (No points, not graded)

d) List the parts of the m-ER protein that is **ABSENT** in the n-ER protein.
The signal sequence (n-ER has a nuclear localization sequence instead) and the transmembrane domain(No points, question not graded)

e) Draw the **orientation of m-ER protein** in the cell membrane and label its amino (-NH$_2$) terminus and carboxyl (-COOH) terminus. *(no points, question not graded)*

7.013 Introductory Biology
Spring 2018