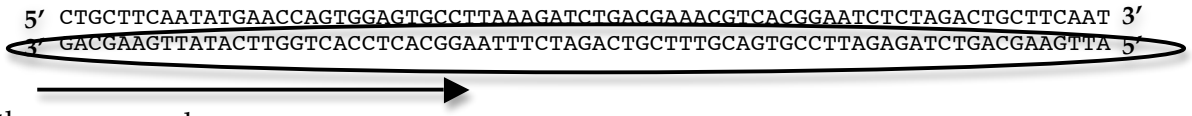
Solution key - 2010 7.012 Problem Set 4

Question 1

The following is a *partial* sequence of a double stranded **bacterial DNA** that encodes a short peptide. *Please note that the promoter for this gene is not shown.*

Wild-type DNA sequence:



a) For the sequence above,

- Circle the template strand for transcription.
- Label the 5′ and the 3′ ends of each strand.
- Indicate the direction of transcription by an arrow.

b) Give the sequence of

- The first 10 nucleotides of the mRNA transcript and label its 5′ and 3′ ends. *Note: You may assume that transcription starts from the first base of the sequence above.*

  5′CUGCUUCAAU3′

- The peptide produced from this mRNA transcript and label its N and C ends. *Please note: A codon chart is provided on the last page of this problem set.*

  N-met-asn-gln-trp-ser-ala-leu-lys-ile-C

c) Give the base sequence of the **anti-codon** that inserts the fourth amino acid into the peptide and label its 5′ and the 3′ ends.

3′ACC 5′

d) The following are two mutant versions of the wild-type DNA sequence that is shown above. The mutated base pair in both versions is **bold and underlined.**

Mutant 1:

CTGCTTCAATATGAAC**T**AGTGGAGTGCCTTAAAGATCTGACGAAACGTCACGGAATCTCTAGACTGCTTCAAT
GACGAAGTTATACTTG**A**TCACCTCACGGAATTTCTAGACTGCTTTGCAGTGCCTTAGAGATCTGACGAAGTTA

Mutant 2:

CTGCTTCAATATGAA**T**CAGTGGAGTGCCTTAAAGATCTGACGAAACGTCACGGAATCTCTAGACTGCTTCAAT
GACGAAGTTATACTT**A**GTCACCTCACGGAATTTCTAGACTGCTTTGCAGTGCCTTAGAGATCTGACGAAGTTA

For each mutant version,

i) Write the sequence of the peptide that is produced. Label N and C termini.

Mutant 1: *N-met-asn-C*

Mutant 2: *N-met-asn-gln-trp-ser-ala-leu-lys-ile-C*

ii) Identify the type of point mutation. *Choose from silent/ missense/ nonsense/ frameshift.*

Mutant 1: *This is a nonsense mutation since a premature stop codon is inserted.*

Mutant 2: *This is a silent mutation, since the peptide sequence is not changed by this mutation.*

Question 1 continued

e) Would the substitution of a base that is a part of the 4th codon in the given wild-type DNA sequence always change the resulting peptide sequence? **Explain** your answer.

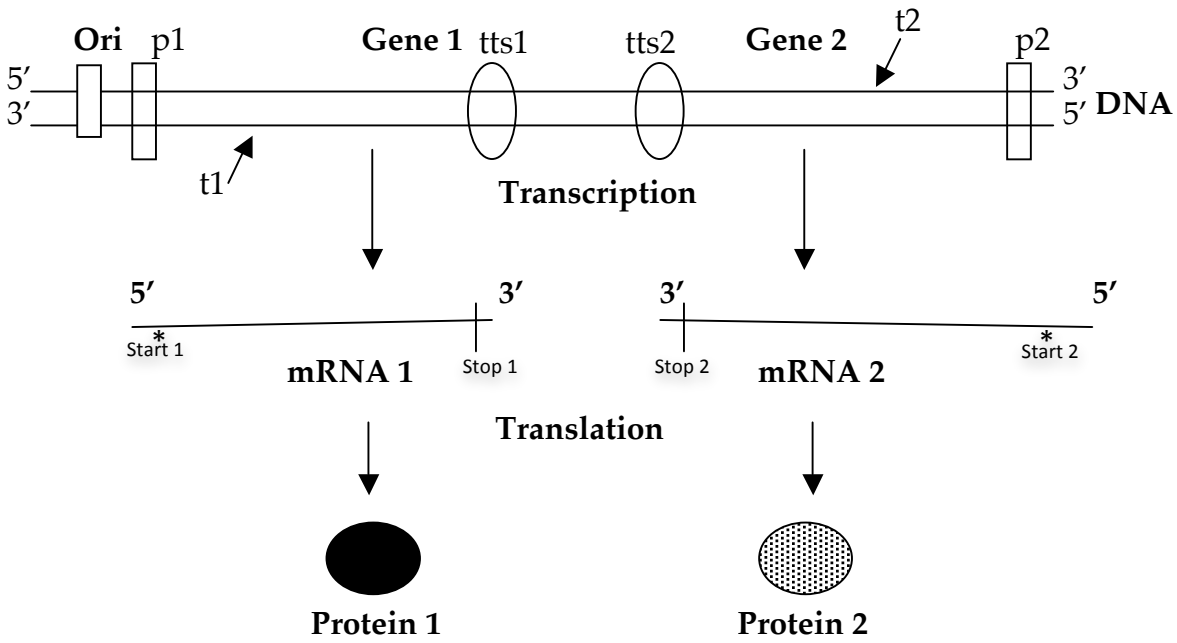
There is only one codon for amino acid tryptophan. Therefore any substitution in this codon will change the amino acid and the resulting protein sequence.

f) Would the substitution of a base that is a part of the 3rd codon in the given wild-type DNA sequence always change the resulting peptide sequence? **Explain** your answer.

There are two codons for asn – 5'AAU 3' (codon 1) and 5'AAC 3' (codon 2). If the 3rd base of codon 2 is changed to "T" you will see the insertion of asn at the same position in the peptide. However if you change the third base of codon 1 to any other base, or change the first or second base pair the amino acid encoded will change

Question 2

a) For each of these genes indicate the location of the following components **on the appropriate molecules** (DNA / RNA) on the diagram. *(As an example, the origin of replication (ori) is shown as a boxed region on the double stranded DNA).*



b) If a mutation **inactivates** the promoter for **Gene 1**, do you expect **protein 1** to be produced (Yes/No)? **Explain**.

No, if RNA polymerase cannot be recruited to the promoter, the mRNA for gene 1 cannot be transcribed, so no protein can be produced.

c) If a mutation inactivates the promoter for **Gene 1** do you expect **protein 2** to be produced (Yes/No)? **Explain**.

Each eukaryotic gene has its own promoter. Therefore inactivation of the promoter for gene 1 should not influence the transcription of gene 2.

As an aside, if you assume that product of gene1 regulates the expression of gene 2 then the inactivation of the promoter for gene 1 will inhibit the production of protein 2.

If you argue that gene 1 and gene 2 are prokaryotic genes and are a part of the same operon, then they will both have the same promoter which, if inactivated, will inhibit the production of protein 1 and protein 2.

Question 2, continued

d) You have identified and characterized a Gene 1 as having three exons and two introns; exon 1(400bp), exon2 (250bp), exon 3 (550bp), intron 1 (2.5kb) and intron 2 (4.5kb).

- Would you classify this as a prokaryotic or eukaryotic gene? **Explain**.
  It is a eukaryotic gene since prokaryotic genes typically lack the introns.

- Explain why this gene may encode multiple **mature mRNA**.
  This gene may encode multiple mature mRNA transcripts due to the alternative splicing of the introns.

Question 3

You discover a novel secreted protein "X" which may have huge commercial potential. You decide to engineer a mouse cell line that will secrete a large amount of Protein X, so that you can purify protein X from the medium. *(Note: A cell line is a single type of cell, which continuously grows in culture).*

a) List four components in the host eukaryotic cell line, which are **absolutely** required for protein synthesis and secretion.
Many answers can be accepted here including: Signal recognition peptide (SRP), SRP receptor on Endoplasmic reticulum (ER), translocon, ER, golgi, ribosomes, vesicles,charged, tRNAs, mRNA etc.

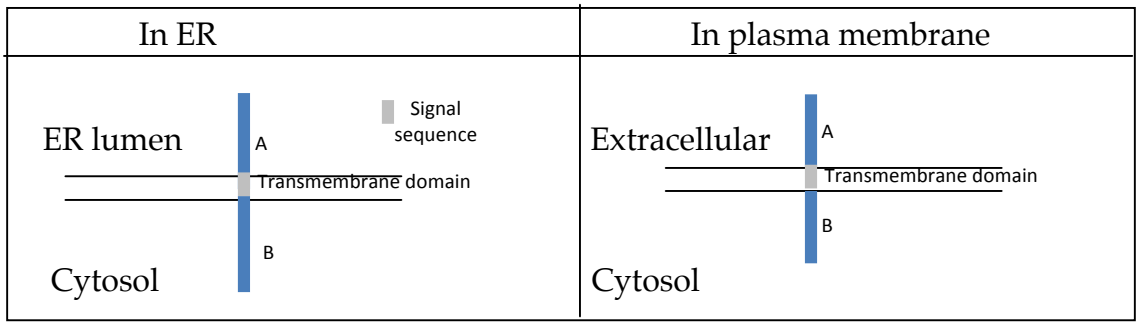
b) You introduce the gene for Protein X into the cell line, you find that protein X is produced but not secreted. (*You may assume that your eukaryotic cell line is normal and contains all the required components for protein production).*

i. What modifications would you make to the X gene that would allow Protein X to be secreted?
ii. Where in gene X (5′ or 3′ end) would you insert the modification?
You would include DNA that encodes a signal sequence such that the signal peptide is on the produced on the N terminus of the polypetide. Thus you will add DNA base pairs immediately after the DNA encoding the start codon. This signal sequence will allow SRP to bind and protein synthesis will temporarily halt. The signal sequence–SRP complex will then be relocated to the ER and will be recognized by the SRP receptor (translocon). The signal sequence will then pass through the ER membrane, the SRP will then detach, the signal sequence will be cleaved and protein synthesis will resume. Once synthesized in the ER lumen, the protein will be packaged in vesicles that will be transported to cell membrane. The vesicles will then fuse with the plasma membrane and protein inside them will be secreted in the surrounding environment.

c) Assuming your manipulation in part (b) of this problem is successful, would the length of the modified protein secreted by the host cell line be the same, larger or smaller than the unmodified protein? Explain your answer.
The nascent protein made from the modified gene would be longer by the amount of the signal sequence. However, signal sequences are removed in the ER so the secreted version of the protein would be the same.

d) Interestingly, you find that this protein, when secreted, can bind to the following plasma membrane receptor protein, which are located on the surface of target. Here A and B are the hydrophilic domains.

Draw this receptor as it is inserted into the endoplasmic reticulum (ER) and the plasma membrane. Label the N and the C termini and include all the domains shown in its schematic.



Question 4

You are interested in purifying and characterizing a specific plant protein (PKA) that acts as a potent anti-depressant. You start by fusing the DNA encoding GFP (green fluorescent protein) to the DNA encoding the C terminus of PKA gene. The following is the partial **cDNA sequence** encoding the C terminus of the PKA protein. The sequence encoding the stop codon is shown in bold. The bars above the sequence show restriction enzyme recognition sites.

```
          Z                    Y
5'-TCAAGAGGATCCCCGCGGTACCGAATTCCATGTTA TAG CAAGCTCGGAATTAACCCTCAC-3'
3'-AGTTCTCCTAGGGGCGCCATGGCTTAAGGTACAAT ATC GTTCGAGCCTTAATTGGGAGTG-5'
```

The following is the partial cDNA sequence encoding the N terminus of GFP. The codons are underlined. The bars above the sequence show the restriction enzyme recognition sites.

```
                        Z       Y
5'-TCTAGAGGTACCGGGATCCGAATTCCC ATG CCA AGC GGC-3'
3'-AGATCTCCATGGCCCTAGGCTTAAGGG TAC GGT TCG CCG-5'
```

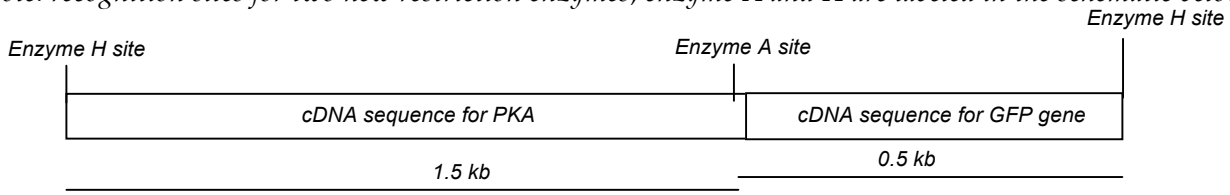
The recognition sequences and the cleavage sites (indicated by /) for each enzyme are given below.

Enzyme Z	***Enzyme Y***
5' G/GATC C 3'	**5' G/AATT C 3'**
3' C CTAG/G 5'	**3' C TTAA/G 5'**

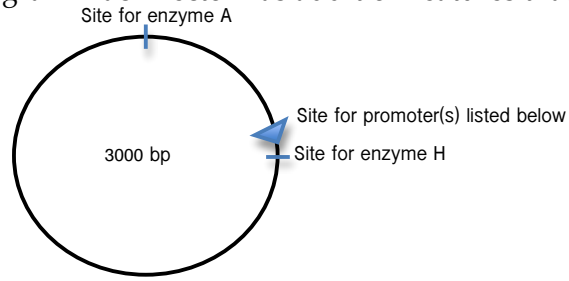
a) Choose the restriction enzyme that you will use to cut the two genes before ligating them to make a fusion gene. **Explain** why you chose this restriction enzyme.
You would cut with restriction enzyme Z, because it is the onlyone that maintains the reading frame needed to express GFP.

Question 4, continued

You successfully create a DNA fragment that encodes the PKA-GFP fusion protein as shown below. *(Note: recognition sites for two new restriction enzymes, enzyme A and H are labeled in the schematic below).*



b) You plan to clone this PKA-GFP gene into a vector that will allow you to amplify and express PKA - GFP fusion gene in bacterial as well as mammalian cells. You have the choice of 4 vectors each with the general features shown in the diagram. Each vector has addition features that are listed below.



Vector 1 contains: 1) ampicillin resistence gene, 2) bacterial origin of replication and 3) bacterial promoter

Vector 2 contains: 1) ampicillin resistence gene, 2) bacterial origin of replication, and 3) mammalian origin of replication, 4) bacterial promoter and 5) a mammalian promoter

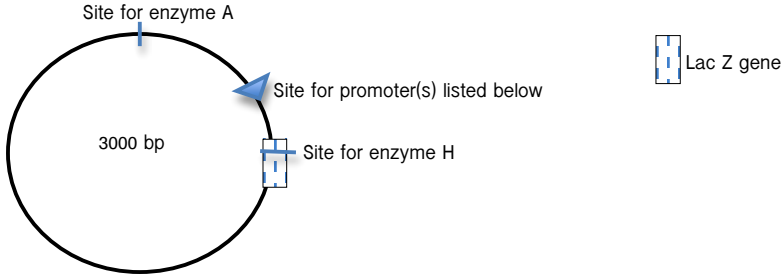
Vector 3 contains: 1) ampicillin resistence gene, 2) bacterial origin of replication, and 3) mammalian origin of replication, 4) a mammalian promoter

Which of the above vectors would allow you to clone and express the fusion gene in both bacterial cells and mammalian cells? **Explain** why you selected this vector.
Only vector 2 will allow expression of the fusion gene in both bacterial cells and mammalian cells because only vector 2 has both types of ORIs and both types of promoters.

c) Based on the plasmid that you selected, what should be the phenotype of the bacterial cells **prior to transformation**?
The cells prior to transformation must be sensitive to ampicillin.

d) Following bacterial transformation, you want to identify the bacterial cells that received a plasmid. Onto what type of media would you plate your transformation mix?
You would plate the cells on media containing ampicillin.

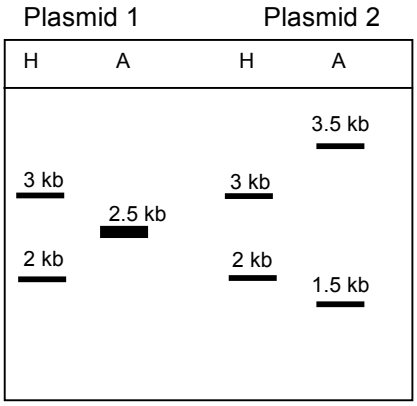
e) The media chosen above does not allow you to distinguish between cells carrying recombinant plasmids and cells carrying non-recombinant plasmids. You design a new vector that also includes the complete lacZ gene. The lacZ gene encodes an enzyme that breaks down the substrate X-gal into a new substance that is blue. When colonies expressing this gene are grown on media containing X-gal, the cells of the colony are bright blue instead of white.



Explain how you would use this vector to distinguish between cells carrying recombinant plasmids and cells carrying non-recombinant plasmids. Include how you would clone the fusion gene into this vector, what type of bacterial cells you would transform with the recombinant vector, what type of media you would use for plating your transformation mix, what type of cells would grow on this type of media (*Untransformed / transformed with non-recombinant plasmid / transformed with recombinant plasmid)*, and how you would distinguish between cells carrying recombinant plasmids and cells carrying non-recombinant plasmids.

1) Cut both the vector DNA and the insert DNA with enzyme H.
2) Ligate the vector and insert DNA
3) Transform ampicillin sensitive cells with the ligation mix and plate cells on media with ampicillin and X-gal.
4) Any cell that received a plasmid will grow. Cells that received the original plasmid will be blue, those that received a recombinant plasmid will be white.

e) You purify the plasmid DNA from two colonies (Colony A containing recombinant plasmid 1 and Colony B containing recombinant plasmid 2) from one of these plates. You perform agarose gel electrophoresis on digested plasmid DNA and the pattern is shown below. *(Note: the distance between Enzyme H and Enzyme A recognition sites on the plasmid vector is 1kb).*



*

i. Which plasmid would you use to express the PKA-GFP fusion protein in mammalian cells? **Explain** your choice.
Only plasmid 1, when inserted into mammalian cells will express PKA-GFP. In this plasmid, the orientation of the fusion gene with respect to the promoter is correct.

ii. How would you detect the **localization and expression** of the PKA fusion protein **in mammalian cells**?
You could follow the localization of this protein by looking for the regions that show fluorescence due to the GFP.