

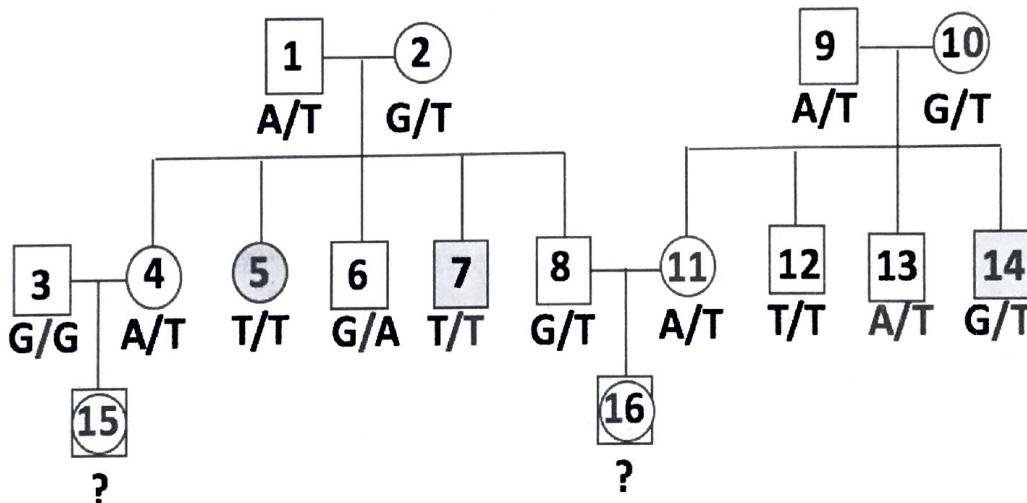
Name Matthew Feng Rec. Section R17 TA Sappington**7.012 Fall 2018: Problem Set 5**

Due: Mon 11/5/2018

The solutions to these problems must be submitted electronically to your TA through the 7.012 Stellar site. All submissions must be received before 9:50 AM on November 5, 2018. Check your file to ensure it was successfully submitted. Only the material that is received prior to the deadline will be graded, no additional material will be accepted after the deadline.

Question 1 (4 points)

Shown below is a pedigree for two families which have the same autosomal disease. Individuals affected by the disease are shaded in grey (males as squares and females as circles). For the disease gene, Gene Q, 'Q' represents the normal allele and 'q' represents the disease allele. The three alleles of a nearby SNP in two families (A, G and C) are shown for each individual. Assume no recombination has occurred between the disease gene and the SNP in the family of individuals 1 and 2 (and their descendants) and in the family of individuals 9 and 10 (and their descendants). Assume that individual 3 does not have any disease-associated alleles.



A.) Which SNP allele(s) is/are tightly linked with the disease allele in the following individuals?

(i) individual 1: T
individual 2: T

(ii) individual 9: T
individual 10: G

Question 1, continued

B.) What is the genotype of the following individuals at the disease locus?

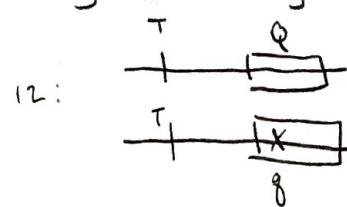
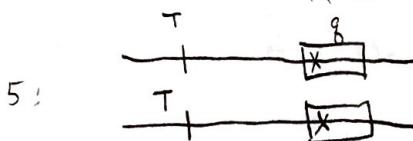
individual	Genotype at the Q locus
4	Qq
5	qq
6	QQ
8	Qq
13	QQ

C.) What is the probability that individual 15 is a carrier for the disease? 1/2

D.) What is the probability that individual 16 is affected by the disease? Ø

E.) Individuals 5 and 12 have the same SNP alleles but individual 5 is affected and individual 12 is not. How can you explain the difference between the SNP genotype and disease phenotypes between individuals 5 and 12? Draw the chromosomes to explain.

Within a single family, the T allele will likely be inherited with the q allele; however, population wide, the genotype T can have appeared in many other ways.



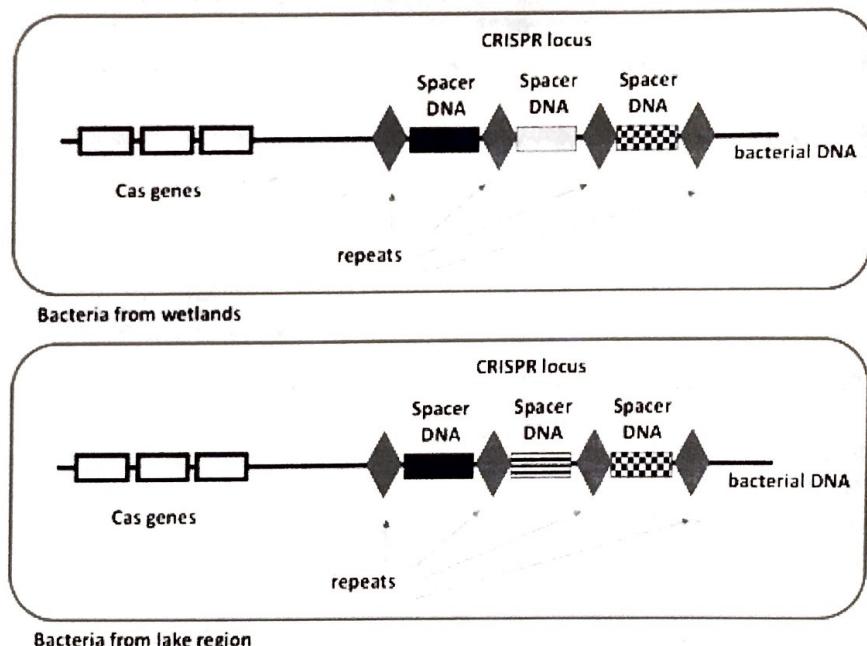
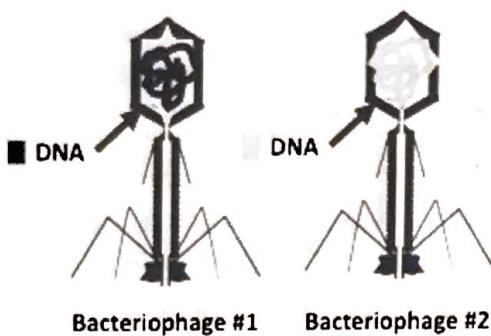
F.) You are studying a family that has a history of migraine headaches. You have reason to believe that there is a single gene that causes the migraines. Which method would be best to find the gene responsible for this trait? Choose from (i) cloning by functional complementation (ii) cloning by protein expression or (iii) positional cloning. Explain your reasons as to why you chose the method and why you did not choose the other two methods.

Positional cloning; we are dealing w/ humans.

Cloning by functional complementation would be lengthy & unethical, & we do not know the protein for protein expression.

Question 2 (3 points)

You study a newly found bacteria. With help from Prof. Lander, you sequence the genome of the bacteria species isolated from two separate locations: from wetlands and from a lake. You identify Cas genes and the CRISPR locus in the bacterial genome. During your studies you also identify two types of bacteriophages and sequence their genomes as well. You are surprised to find out that you can find some bacteriophage DNA (shown in black and grey in the schematic below) in the genome sequence of bacteria.



A.) Why does the bacterial genome contain bacteriophage DNA?

The bacterial genome includes bacteriophage DNA to recognize & cut up bacteriophage DNA if found elsewhere in the cell.

Question 2, continued

- B.) Name one way in which the immunity provided by CRISPR system differs from immunity provided by restriction enzymes in bacteria?

Restriction enzymes target a predetermined sequence; CRISPR system uses guide RNA to cut DNA.

- C.) Assuming different phage DNA sequences are depicted in different colors (black and grey) within the phages and bacterial genomes, make a prediction about the habitat of each phage.

Phage 1: Both wetlands & lake

Phage 2: Wetlands.

- D.) Indicate which bacteria are resistant to which phage. Explain your reasoning.

Wetlands bacteria is resistant to both phages, as their DNA is in the CRISPR locus; lake bacteria

- E.) You sequence bacteria from a nearby bog. The genome sequence region is similar to bacteria from wetlands, except the Cas genes have been deleted. Which phage would bacteria from the bog be resistant to? Explain your answer.

Neither; Cas genes are required to create the protein that actually cut the DNA.

Question 3 (5 points)

Choose from methods listed below to answer question 2A:

- traditional gene knockout method by ES cells (1)
- CRISPR/Cas9 (2)
- introduction of transgene under strong promoter (3)
- RNAi (4)
- conditional gene knockout (5)

- A.) Which method(s) could you use in mice if you wanted to study the effect of

(i) the absence of a gene: 1, 2, 4, 5

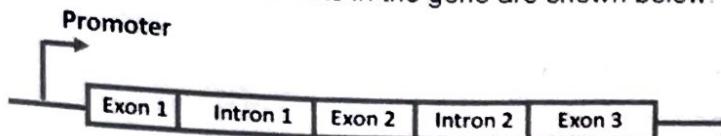
(ii) overexpression of a gene: 2, 3

(iii) disrupting a gene in the liver: 5

(iv) disrupting the expression of a gene without cutting the genome: 4

Question 3, continued

Suppose you decide to study the function of a gene in mice by traditional gene knockout method. The DNA elements in the gene are shown below:



Your helpful lab mate would like to help you by designing a few constructs for you to transfet into ES cells. Please note that in the construct shown below in (iii), the antibiotic selection marker is integrated into Intron 1.

- i.

Intron1	Antibiotic Selection	Negative Selection	Intron2
---------	----------------------	--------------------	---------
- ii.

Antibiotic Selection	Exon1	Negative Selection
----------------------	-------	--------------------
- iii.

Intr-	Antibiotic Selection	-on1	Negative Selection
-------	----------------------	------	--------------------
- iv.

Intron1	Antibiotic Selection	Intron2	Negative Selection
---------	----------------------	---------	--------------------

B.) For each of the constructs indicate whether they would be successful in knocking out the gene of interest in ES cells. If the construct would not work, indicate why.

- I : Not successful — negative selection is inside homologous region
- II : Not successful — negative selection in homologous region
- III : Not successful — would knockout introns \Rightarrow no use.
- IV : Successful

C.) After you choose the correct construct, you transfer your construct into ES cells obtained from mice with **black coat color**. Once you get the ES cells that contain your construct in the correct genomic locus, you inject these ES into the blastocyst of mice. (Please note that white coat color is recessive to black color.)

(i) In which coat color mice, would you be able to detect chimeric mice: black or white?

White

(ii) After you get chimeric mice, which coat color mice would you cross your mice with the gene knockout: black or white? Why do you need to perform this step?

White; we do this to obtain knockout gene mice
with every cell having one copy of the
knocked out gene.

Question 3, continued

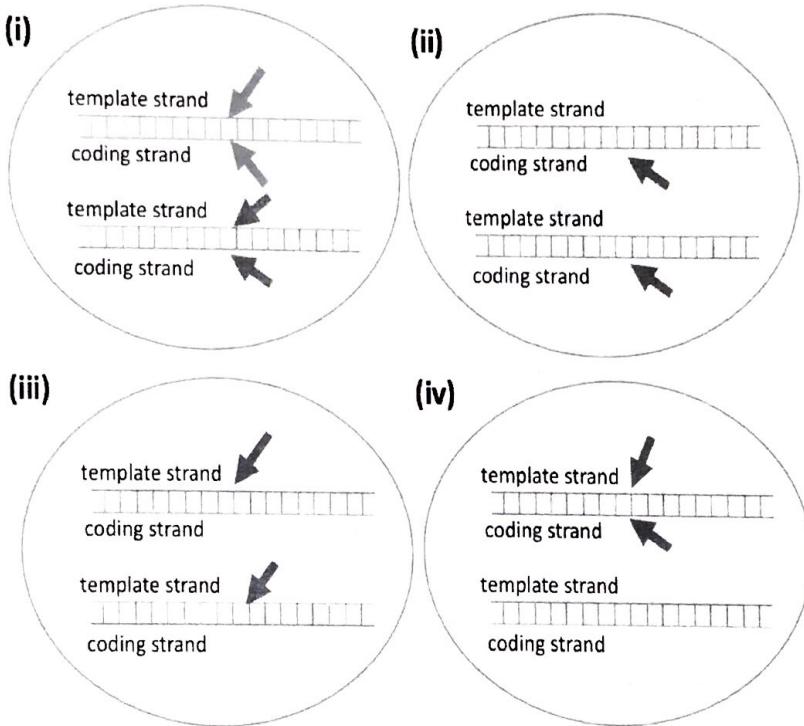
D.) You decide you instead would like to use CRISPR/Cas9 method to target your gene of interest. Which DNA elements would you need to introduce into ES cells to achieve your goal with CRISPR/Cas9 method? Circle all that apply from below:

- (i) Single guide RNA (ii) Telomerase (iii) Cas9 gene (iv) repair template (v) RNA polymerase

E) If you wanted to use CRISPR/Cas9 technology to restore a point mutation in your gene to the wild-type sequence instead, which DNA element would you need to introduce into ES cells in addition to those you chose in part D?

We need to add the repair template so that the ES cells use it to repair the cut, inserting our desired DNA.

Two homologous chromosomes in ES cells are depicted below with the portion of your gene of interest that is targeted with the CRISPR/Cas9 method. Proposed cleavage sites of Cas9 on each strand is shown with arrows.



F.) Which schematic (i-iv) correctly portrays the action of Cas9?

I

Question 4 (6 points)

You have found the human gene for the perfect pitch (PP) protein, which can give one the perfect pitch in baseball and in music. You would like to produce lots of the PP protein and sell it. The best way is to express the protein in *E. coli*. You find a suitable expression vector and need a strong promoter that can produce lots of protein. Your roommate says that you should use a constitutive promoter to ensure that you are expressing your protein at all times to get the highest yield of your protein.

A.) Do you agree with your roommate? If not, what type of promoter should you use? Explain your answer.

Yes. Constitutive promoter will allow PP protein to be continuously produced in *E. coli*.

B.) You choose the correct promoter for the best yield. You paste your gene into the expression vector in front of your promoter. However, despite following the protocols correctly, you get a very low protein yield. Your roommate says that your protein is arginine rich and that might be a problem. Find below a chart for arginine codon usage between humans and *E. coli*.

arginine codon	human usage (%)	<i>E. coli</i> usage (%)
AGA	20	7
AGG	20	4
CGA	11	7
CGC	19	36
CGG	21	11
CGU	8	36

Consider the chart above and explain why you might have a low protein yield and suggest how you can improve your protein yield. In your answer, give specific examples from the chart.

The anticodons that *E. coli* has in abundance are not found (the codons) in the human DNA very often; we can improve protein yield by adding tRNA that binds AGA, AGG, and CGG. C.) Now you have great production of your protein in *E. coli*. However, you realize that you do not have a purification tag on your protein. Your roommate says no problem. Your protein binds DNA so you should be able to use a purification column with some DNA bound to it and purify your protein by binding it to the DNA on the column. Your lab mate disagrees and insists that you should use a purification tag like Histidine tag. Which friend's method will give you purer and higher yield of your protein? Explain your answer.

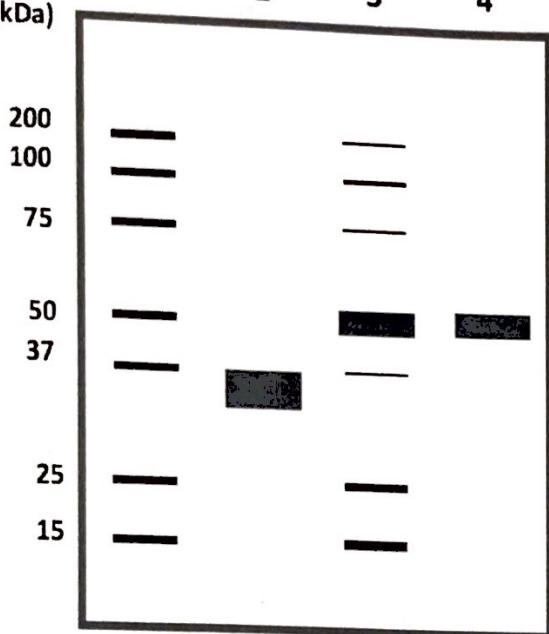
Roommate; separation can be done w/o the histidine tag, which may affect affinity.

Question 4, continued

D.) You need lots of pure protein for your protein structure work. Three of your friends decide to help out with purification of your protein. They each purify your protein and then run the purified protein preps together on a protein gel (shown below on lanes 2-4). Your protein of interest is 50 kDa. Molecular weight markers were run on lane 1. Which protein prep (lanes 2-4) will you use for your protein structure studies? Explain why.

Molecular Weight**Markers**

(kDa)



Lane 4 — it contains only
~50 kDa proteins.

E.) You would like to get a crystal structure of your protein to find out how close a histidine residue is to bound substrate. You think a histidine might act as a catalytic base, but are not sure if the histidine is close enough to substrate to perform this job. Your crystals diffract to 6 Å resolution. Is 6 Å resolution high enough to determine precisely where the histidine is located? Briefly explain. Can you trust a 6 Å resolution structure to be accurate?

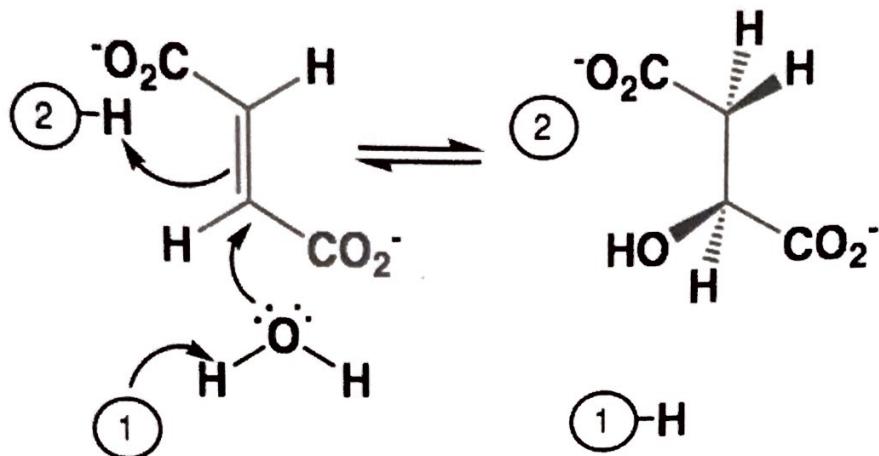
No — 6 Å resolution only is sharp enough for secondary structure, such as α -helices.

F.) Why are people saying that there is a "resolution revolution" going on with cryo-electron microscopy? Briefly explain.

Cryo-electron microscopy is shrinking the diffraction resolution, so that we can image smaller and smaller proteins.

Question 5 (2 points)

Below is a mechanistic proposal for an enzyme. The numbers 1 and 2 correspond to amino acids.



A.) The amino acid at residue 1 is shown to be acting as a _____ (select acid catalyst, base catalyst, or covalent catalyst) in the forward direction.

B.) The amino acid at residue 2 is shown to be acting as a _____ (select acid catalyst, base catalyst, covalent catalyst) in the forward direction.

C.) The structure shows that there are multiple amino acids close enough to substrate (on left) to be "residue-1." These amino acids include histidine, serine, threonine, lysine and asparagine. Which of these residues, if any, is likely to perform the task of activating the water molecule that is shown above? Briefly explain.

Asparagine ; it is polar but uncharged

D.) The residues listed in Part C (histidine, serine, threonine, lysine, and asparagine) may also be important in forming electrostatic interactions with the carboxylate group(s) of the substrate (on left). Which residues, if any, could form favorable electrostatic interactions with a carboxylic group? Briefly explain.

Lysine has least no. of lone pairs and is positively charged.

E.) You want to test if the serine mentioned above is catalytically important. The codon used for serine in this case is UCU. What base would you alter so that the new amino acid is Ala?

First U to G , GCU.