

7.03 Fall 2003
Problem Set #3 Solutions
Issued Friday, October 17, 2003

1. (a) We are analyzing mutagens that specifically induce G·C→A·T mutations in DNA. Therefore, we must determine the potential double stranded DNA sequences that will encode stop codons after going through this specific mutation.

We will start with 5'UAG3'. The double stranded DNA that corresponds to 5'UAG3' is:
3'ATC5' template strand
5'TAG3' coding strand

We need to figure out what specific double stranded DNA sequences could have undergone a G·C→A·T mutation to become the sequence above. To do this, just work backwards and change the AT base pairs in the above sequence into GC base pairs, one pair at a time. If you do this, you will get the following double stranded DNA sequences (the mutated bases are boldfaced):

3'**G**T**C**5' template strand
5'**C**A**G**3' coding strand (codes for Gln)

3'**A**C**C**5' template strand
5'**T**G**G**3' coding strand (codes for Trp)

These two sequences each underwent one G·C→A·T mutation to become DNA that encodes the 5'UAG3' stop codon. We have chosen to make only one mutation per three base pairs because the likelihood of a mutagen acting on two consecutive base pairs (or two base pairs out of three) is extremely small. We can neglect those extremely rare events since this problem is asking why something "generally" happens.

Now, repeat the process for the other two stop codons. The double stranded DNA that encodes 5'UGA3' is:

3'**A**C**T**5' template strand
5'**T**G**A**3' coding strand

And the DNA possible sequences that could have been mutated to become the above sequence are:

3'**G**C**T**5' template strand
5'**C**G**A**3' coding strand (codes for Arg)

3'**A**C**C**5' template strand
5'**T**G**G**3' coding strand (codes for Trp)

Lastly, repeat for 5'UAA3'. Its corresponding DNA is:

3'**A**T**T**5' template strand
5'**T**A**A**3' coding strand

And the DNA possible sequences that could have been mutated to become the above sequence are:

3'**GTT**5' template strand

5'**CAA**3' coding strand (codes for Gln)

3'**ACT**5' template strand

5'**TGA**3' coding strand (codes for UGA stop codon)

3'**ATC**5' template strand

5'**TAG**3' coding strand (codes for UAG stop codon)

Notice that for UGA and UAG, **the two candidate sequences for mutation both encode amino acids**, whereas for UAA, **only one of its candidate sequences encodes an amino acid**. The other two correspond to stop codons. When G·C→A·T mutagens are introduced, they mutate random base pairs along the DNA sequence. **Given that there are more targets for the creation of TGA and TAG nonsense mutations, these two mutations will occur more frequently than TAA mutations.**

(b) The amber TAG mutation has the following mRNA codon, tRNA anticodon, and corresponding DNA coding for the anticodon portion of the tRNA:

mRNA codon 5'UAG3'

tRNA anticodon 3'AUC5'

DNA encoding anticodon portion 5'TAG3' ← template strand for tRNA transcription
3'ATC5'

This DNA sequence is also the sequence that codes for the anticodon portion of amber suppressing tRNA alleles. In other words, we want the mutagen to change normal DNA sequences into this sequence, so that we can have amber suppressing alleles. To find which tRNA genes can be altered to become the DNA sequence above, we can work backwards like in part (a). Now, since we have a G·C↔A·T mutagen, we can change AT pairs into GC pairs, and vice versa. Again, we will change one base pair at a time. If we do this, we will get the following three DNA sequences (the mutated bases are in boldface):

DNA encoding anticodon portion 5'**CAG**3' ← tRNA template (for tRNA^{Gln})
3'**GTC**5'

DNA encoding anticodon portion 5'**TGG**3' ← tRNA template (for tRNA^{Trp})
3'**ACC**5'

DNA encoding anticodon portion 5'**TAA**3' ← tRNA template (for UAA stop tRNA)
3'**ATT**5'

If you follow the flow of genetic information, you will see that these three sequences code for the anticodon portion of tRNA^{Gln}, tRNA^{Trp}, and the tRNA that recognizes the UAA stop codon. **So the gln and trp tRNA genes can be mutated to become amber**

suppressors. The codons normally recognized by tRNA^{Gln} and tRNA^{Trp} are 5'CAG3' and 5'UGG3', respectively.

We gave full credit to those who ended the solution here. But since the problem asked which genes could "in principle" be altered to become amber suppressors, we can consider cases where two or three base pairs from each codon were altered by the mutagen. In that case, we would arrive at the following DNA sequences (with the corresponding mRNA codons also listed):

DNA encoding anticodon portion 5'CGA3' ← tRNA template (for tRNA^{Arg})
 3'GCT5'
 mRNA codon 5'CGA3'

DNA encoding anticodon portion 5'CGG3' ← tRNA template (for tRNA^{Arg})
 3'GCC5'
 mRNA codon 5'CGG3'

DNA encoding anticodon portion 5'CAA3' ← tRNA template (for tRNA^{Gln})
 3'GTT5'
 mRNA codon 5'CAA3'

DNA encoding anticodon portion 5'TGA3' ← tRNA template (for UGA stop tRNA)
 3'ACT5'
 mRNA codon 5'UGA3'

So assuming the mutagen can change two or three base pairs in the same triplet, we arrive at additional tRNA genes that can be mutated to become amber suppressors. These are **two tRNA^{Arg} genes and another tRNA^{Gln}**.

2. (a) ⇨ = IS, ► = OriT

Hfr #1: A ⇨►⇨B ⇨ C ⇨ D A transferred early

Hfr #2: A ⇨B ⇨◄◄⇨ C ⇨ D C (and then D) transferred early

Hfr #3: A ⇨B ⇨ C ⇨►⇨D C (then B then A) transferred early

(b) label the order of IS from A ⇨ ... ⇨D
 1 4

Hfr #	Recombination Between IS #s	Recombinatory Result	Product	Early
1	1 & 2	⇨►	F	---
1	1 & 3	A ⇨B ⇨◄◄⇨ C ⇨ D	Hfr	C, D
1	1 & 4	⇨►⇨B ⇨ C	F'	C, B

1	2 & 3 (or 3&4*)	A ⇨▶⇨B ⇨ C ⇨ D	Hfr	A
1	2 & 4	A ⇨▶⇨ D	Hfr	A
2	1 & 2 or 3 & 4	A ⇨B ⇨◀⇨ C ⇨ D	Hfr	C, D
2	1 & 3	A ⇨▶⇨B⇨ C ⇨ D	Hfr	A
2	(1&4*)	⇨B ⇨◀⇨ C	F'	C, B
2	2 & 3	◀⇨	F	---
2	2 & 4	A ⇨B ⇨ C ⇨▶⇨ D	Hfr	C, B, A
3	2 & 3 (or 1&2*)	A ⇨B ⇨ C ⇨▶⇨D	Hfr	C, B, A
3	1 & 3	A ⇨▶⇨D	Hfr	A
3	1 & 4	B ⇨ C ⇨▶⇨	F'	C, B
3	2 & 4	A ⇨B ⇨◀⇨ C ⇨D	Hfr	C, D
3	3 & 4	▶⇨	F	---

* denotes recombinatory events between two IS sequences NOT flanking the OriT
(The p-set asked only for events between sequences that flank the OriT. Therefore these are extraneous answers)

3. (a) The donor strain is Lac⁺ and the recipient strain is Lac⁻. Therefore, in the Lac⁺ Kan^r transductants, *lacI*⁺ was cotransduced with Tn5. So the distance between Tn5 and the *lacI*⁻ mutation is:

$$(18/100) \cdot 100\% = 18\%$$

(b) Since none of the 100 Kan^r transductants were Lac⁺, we can conclude that Tn5 was never co-transduced with *lac2*⁺. This indicates that the distance between *lac2*⁻ and Tn5 is at least one phage head (10⁵ bp). We know from part (a) that Tn5 and *lacI*⁻ are within one phage head since their cotransduction frequency was 18%. But since we do not know the relative order of the three markers, we cannot say whether the two *lac* mutations are within one phage head. If Tn5 were between *lacI*⁻ and *lac2*⁻, then the distance between the two mutations would be more than one phage head. However, if Tn5 were not the middle marker, we cannot say whether *lacI*⁻ and *lac2*⁻ are one phage head apart. **So the only conclusion we can draw from this data is that *lacI*⁻ and *lac2*⁻ are very far from each other. Therefore, they are unlinked, and are mutations in two different genes.**

(c) The best way to solve this type of problem is to draw out the two crosses. We will consider only two possible orders, instead of three, because the order where Tn5 is in the middle is impractical for three-factor cotransduction experiments. It is likely that Tn5 insertion between the two markers of interest (which are close together in this case) will occur in a coding region, causing an additional mutation that would skew the experimental data.

There are two reciprocal crosses and two possible orders. So we will have to draw four diagrams:

(1- and 1+ will denote the mutant and WT loci of *lac1*, respectively. Similarly, C- and C+ will denote the mutant and WT loci of *lac^c*)

	Order #1 (Tn5, <i>lac1</i> , <i>lac^c</i>)	Order #2 (Tn5, <i>lac^c</i> , <i>lac1</i>)																																								
Cross #1 (Tn5, 1- crossed to C-)	<table style="width: 100%; border-collapse: collapse;"> <tr> <td style="text-align: center;">Tn5</td> <td style="text-align: center;">1-</td> <td style="text-align: center;">C+</td> <td></td> </tr> <tr> <td colspan="4" style="text-align: center;">~~~~~</td> </tr> <tr> <td style="text-align: center;">X</td> <td style="text-align: center;">X</td> <td style="text-align: center;">X</td> <td style="text-align: center;">X</td> </tr> <tr> <td colspan="4" style="text-align: center;">-----</td> </tr> <tr> <td></td> <td style="text-align: center;">1+</td> <td style="text-align: center;">C-</td> <td></td> </tr> </table>	Tn5	1-	C+		~~~~~				X	X	X	X	-----					1+	C-		<table style="width: 100%; border-collapse: collapse;"> <tr> <td style="text-align: center;">Tn5</td> <td style="text-align: center;">C+</td> <td style="text-align: center;">1-</td> <td></td> </tr> <tr> <td colspan="4" style="text-align: center;">~~~~~</td> </tr> <tr> <td style="text-align: center;">X</td> <td></td> <td style="text-align: center;">X</td> <td></td> </tr> <tr> <td colspan="4" style="text-align: center;">-----</td> </tr> <tr> <td></td> <td style="text-align: center;">C-</td> <td style="text-align: center;">1+</td> <td></td> </tr> </table>	Tn5	C+	1-		~~~~~				X		X		-----					C-	1+	
Tn5	1-	C+																																								
~~~~~																																										
X	X	X	X																																							
-----																																										
	1+	C-																																								
Tn5	C+	1-																																								
~~~~~																																										
X		X																																								

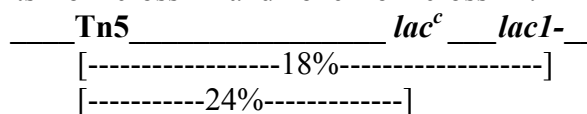
	C-	1+																																								
Cross #2 (Tn5, C- crossed to 1-)	<table style="width: 100%; border-collapse: collapse;"> <tr> <td style="text-align: center;">Tn5</td> <td style="text-align: center;">1+</td> <td style="text-align: center;">C-</td> <td></td> </tr> <tr> <td colspan="4" style="text-align: center;">~~~~~</td> </tr> <tr> <td style="text-align: center;">X</td> <td></td> <td style="text-align: center;">X</td> <td></td> </tr> <tr> <td colspan="4" style="text-align: center;">-----</td> </tr> <tr> <td></td> <td style="text-align: center;">1-</td> <td style="text-align: center;">C+</td> <td></td> </tr> </table>	Tn5	1+	C-		~~~~~				X		X		-----					1-	C+		<table style="width: 100%; border-collapse: collapse;"> <tr> <td style="text-align: center;">Tn5</td> <td style="text-align: center;">C-</td> <td style="text-align: center;">1+</td> <td></td> </tr> <tr> <td colspan="4" style="text-align: center;">~~~~~</td> </tr> <tr> <td style="text-align: center;">X</td> <td style="text-align: center;">X</td> <td style="text-align: center;">X</td> <td style="text-align: center;">X</td> </tr> <tr> <td colspan="4" style="text-align: center;">-----</td> </tr> <tr> <td></td> <td style="text-align: center;">C+</td> <td style="text-align: center;">1-</td> <td></td> </tr> </table>	Tn5	C-	1+		~~~~~				X	X	X	X	-----					C+	1-	
Tn5	1+	C-																																								
~~~~~																																										
X		X																																								
-----																																										
	1-	C+																																								
Tn5	C-	1+																																								
~~~~~																																										
X	X	X	X																																							

	C+	1-																																								

The data given will allow us to determine which of the two possible orders is correct. As in any three factor cross, we determine order by looking for the rarest class. In this case, the rarest class shows normal B-gal expression. The genotype of this class is Tn5, 1+, C+.

From the data, we see that cross #1 produced four such transductants while cross #2 produced none. This is the key observation that allows us to determine order.

We look back to our diagrams. The X's indicate the crossovers that needed to occur to give the Tn5, 1+, C+ genotype (which gives normal B-gal expression). Assuming order #1 is correct, we would get more normal B-gal transductants from cross #2, since only two crossovers are required as opposed to four. Assuming order #2 is correct, we would get more normal B-gal transductants from cross #1. This is what the data shows. There are four normal B-gal transductants from cross #1 and none from cross #2. **Therefore, order #2 is the correct order:**



The Tn5 to *lac1-* distance is 18% as calculated in part (a).

The Tn5 to *lac^c* distance is found in the data for the first reciprocal cross. The 76/100 Kan^r b-gal constitutive transductants from this cross received Tn5 but did not receive C+. Therefore, P(no cotransduction between Tn5 and *lac^c*) = 76%

We can see that: $P(\text{cotransd. btw Tn5 and } lac^c) = 1 - P(\text{no cotransd. btw Tn5 and } lac^c) = 100\% - 76\% = 24\%$

The distance between *lac^c* and *lacI* is very small. This is indicated by the fact that there are very few transductants that exhibit normal b-gal expression (crossing over between *lac^c* and *lacI* needs to occur to produce transductants with normal b-gal expression in both crosses). The exact distance between *lac^c* and *lacI* cannot be calculated since we only selected for Kan^r transductants. Therefore, all data is relative to the Tn5 marker.

(d) The F' plasmid carries Tn5 and *lac3-* while the chromosome has a *lacI-*. When selecting for Kan^r, you are selecting for cell that have successfully taken in the F plasmid, which confers resistance to kanamycin.

If there merodiploids express b-gal normally, then you can conclude that the ***lacI-* and *lac3-* mutations lie in different genes (they complement each other).**

If the merodiploid were Lac⁻, then you can draw one of two conclusions, which are indistinguishable without further experimentation: **(1) the two mutations lie in the same gene, or (2) one (or both) of the mutations is dominant to wild type.**

4. A Kan^R, Tet^R transductant results when recombination of the Tn5 occurs in the genome in a place not overlapping with the insertion site of Tn10. This happens 9920/10000 times, or 99.2% of the time. Similarly, A Kan^R, Tet^S transductant results when recombination of Tn5 occurs in the genome in a place overlapping with the insertion site of Tn10, thus disrupting the Tet^R gene. This happens 80/10000 times, or 0.8% of the time.

We are told that the average size of a recombined fragment is 55kbp, this means that a phage head can fit 55 kbp of DNA, which is larger than either Tn10 or Tn5. We can find the size of the genome by seeing that 55kbp is proportional to the size of the genome just as 0.8% is proportional to 100% (0.8% recombination tells us the phage is holding 0.8% of the genome). However, before we can use 55kbp to find the genome size, we must subtract the 5kbp of Tn5. We subtract off 5kbp because there is no Tn5 in the genome, so there is no possibility of recombination within this region. Therefore, it is not counted in our measurement of recombination.

Therefore: $80/10,000 = 0.008 =$ the probability that the 50 kbp phage DNA with Tn5 recombined into a spot overlapping Tn10.

$0.008/50 = 1/(\text{total genome length}) \rightarrow$ from this we get a **total genome length of 6250kbp.**

note: A common mistake would be to subtract the size of the Tn10 off of the final answer. This is not correct, because it was never included in the 6250kbp in the first place, as there is no Tn10 in the 55kbp of DNA and there was never a chance of recombination.