PLASMID/MOLECULAR BIOLOGY LAB

Name_____ Section_____

Part 1. Finding the major coding regions in the plasmid.

Penicillin resistance in bacteria occurs in many species. The resistance is due to production of an enzyme, beta-lactamase, from a plasmid encoded gene. The enzyme is secreted from the cell, and hydrolyzes the antibiotic penicillin.

1. Locate the start of the penicillin resistance gene. The first codon (ATG) is at base pair 143. Locate the stop codon (TAA) for penicillin resistance gene at position 1001. Highlight this entire protein coding region with a yellow highlighter. Write " Pen^r " in the right margin next to the highlighted region

2. Locate the start codon (ATG) of the coding sequence for beta-galactosidase enzyme at position 2424. This enzyme is part of the lactose operon and is used in this plasmid as a selectable marker. Locate the stop codon at the end of this enzyme coding sequence at position 2793 (TAA). Highlight the entire protein coding region with yellow also. Write " Lac " in the right margin next to the highlighted region.

3. Locate the start of the large RNA molecule (called "**RNA II**") that provides the primer to start DNA replication. Recall that DNA polymerase requires a primer to begin DNA synthesis. Instead of primase making a primer this RNA is transcribed and used as a primer. The first nucleotide of the primer sequence is at position 1701. Locate the end of RNA II at position 2253. Highlight the whole RNA II sequence. Write "RNA II" in the right margin.

4. Locate the RNA that controls the assembly of the RNA II with the plasmid. The sequence of this RNA is entirely **WITHIN** the sequence of RNA II but runs in the opposite direction because it is coded for by the complimentary strand. It's called **RNA I**. It is 108 bases long and ends at 1703. So where is the start (the 5'-phosphate end)? Draw a box around this and label it RNA I.

5. Locate the BspH 1 restriction enzyme site (TCATGA) that starts at position 1589. Locate the Dra 1 restriction enzyme site (TTTAAA) at 1100 to 1105. The region between these sites contains the necessary information for proteins from a bacterial virus called M13 to bind the DNA and initiate synthesis of a single-stranded version of DNA sequences that can be inserted here. Highlight the region between these sites in yellow. Write " M13 origin" in the right margin.

6. Locate the special sequence that acts as a terminator for the mRNA of the beta-galactosidase gene. Its starts at 2797 and extends to 2818. Put a box around, label it "terminator" and draw it below showing the actual RNA sequence as it would form base paired stem and a small unpaired loop. You will have to mentally transcribe from the DNA sequence given to the RNA sequence. Just change the T's to U's!

Part 2
Cracking the code.

7. Lightly mark off each of the first 10 codons for the amino acids of the penicillin resistance protein (use little parentheses.) Start with the ATG at 143 but of course in the RNA it would be AUG.

8. **Translate** these first 10 codons of the penicillin resistance gene using the genetic code. You will of course have to translate to RNA from the DNA code given. Again, just change T's to U's. Write down the first ten codons below, with the proper amino acid below
Like this:

AUG -
met -

9 How many amino acids are in the penicillin resistance protein? _____
You don't have to count them. Since it takes three nucleotides to code for one amino acid just subtract the base pair number of the first codon
from the base pair number of the last codon and divide by 3. If you don't get a whole number you are using the wrong number to represent the stop codon location. Figure it out.

10. Locate the start codon for the beta-galctosidase enzyme. Like all other bacterial proteins, it starts with the amino acid methionine and the code 'ATG'.
But.... in the mRNA, it reads A__G
Mark off the first 10 codons and translate them like you did for the pen^r gene.
Write them here:

AUG -
met -

11. How many amino acids are there in this small protein? _____

PART 3.
Control of transcription and translation.

12. Locate the promotor for the penicillin resistance gene. The consensus sequences are (the -35) TTGACA (starting at bp100) and (the -10) TATAAT (starting at 120). The RNA polymerase molecule covers both of these sites and the DNA between them so draw a nice oval showing where it would bind. The first nucleotide of the transcript RNA of the message is at position 131. Put a little arrow below it to show the direction of transcription.

13. Locate the leader sequence/**binding site for the ribosomes** that will translate the penicillin gene. The bacterial ribosome binding site sequence is AGGAAG. Obviously, this site must be BEFORE the first codon. Draw a nice circle around it and label it "ribosome binding site".

14. Locate the promoter consensus sequences (given in #12 above) for the beta-galactosidase gene. It starts at 2292 and ends at 2317. Draw an oval around these showing where RNA polymerase binds.

15. Locate the start of transcription of "lac" at 2325. Use an arrow and line to show the mRNA direction of transcription. Write "lac mRNA"

16. Locate the leader sequence/ribosome binding site for lac at 2413. Draw a circle around it and label it, "ribosome binding site."

17. Locate the operator/binding site of the lacI repressor protein from 2381 to 2410. circle the whole binding region and label it, "Operator". This sequence is an inverted repeat with the same sequence appearing on BOTH strands of the DNA read in a 5' to 3' direction. You only have the top strand sequence however so you will find the compliment of the repeat. The double stranded version is shown below. Write in the complimentary strand under the appropriate location in the sequence given and use arrows to show the inverted repeat sequences.

TGTGG**AATTGTGA**GCGGATCACAATTTCAC
ACACCTTAACACTCGCCT**AGTGTTAA**AGTG

```
1     GTTAACTACG TCAGGTGGCA CTTTTCGGGG AAATGTGCGC GGAACCCCTA
51    TTTGTTTATT TTTCTAAATA CATTCAAATA TGTATCCGCT CATTTGACAT
101   TGACACTTAT AAATGCTTCT ATAATATTGA AAAAGGAAGA GTATGAGTAT
151   TCAACATTTC CGTGTCGCCC TTATTCCCTT TTTTGCGGCA TTTTGCCTTC
201   CTGTTTTTGC TCACCCAGAA ACGCTGGTGA AAGTAAAAGA TGCTGAAGAT
251   CAGTTGGGTG CACGAGTGGG TTACATCGAA CTGGATCTCA ACAGCGGTAA
301   GATCCTTGAG AGTTTTCGCC CCGAAGAACG TTCTCCAATG ATGAGCACTT
351   TTAAAGTTCT GCTATGTGGC GCGGTATTAT CCCGTGTTGA CGCCGGGCAA
401   GAGCAACTCG GTCGCCGCAT ACACTATTCT CAGAATGACT TGGTTGAGTA
451   CTCACCAGTC ACAGAAAAGC ATCTTACGGA TGGCATGACA GTAAGAGAAT
501   TATGCAGTGC TGCCATAACC ATGAGTGATA ACACTGCGGC CAACTTACTT
551   CTGACAACGA TCGGAGGACC GAAGGAGCTA ACCGCTTTTT TGCACAACAT
601   GGGGGATCAT GTAACTCGCC TTGATCGTTG GGAACCGGAG CTGAATGAAG
651   CCATACCAAA CGACGAGCGT GACACCACGA TGCCTGTAGC AATGGCAACA
701   ACGTTGCGCA AACTATTAAC TGGCGAACTA CTTACTCTAG CTTCCCGGCA
751   ACAATTAATA GACTGGATGG AGGCGGATAA AGTTGCAGGA CCACTTCTGC
801   GCTCGGCCCT TCCGGCTGGC TGGTTTATTG CTGATAAATC TGGAGCCGGT
851   GAGCGTGGGT CTCGCGGTAT CATTGCAGCA CTGGGGCCAG ATGGTAAGCC
901   CTCCCGTATC GTAGTTATCT ACACGACGGG GAGTCAGGCA ACTATGGATG
951   AACGAAATAG ACAGATCGCT GAGATAGGTG CCTCACTGAT TAAGCATTGG
1001  TAACTGTCAG ACCAAGTTTA CTCATATATA CTTTAGATTG ATTTACCCCG
1051  GTTGATAATC AGAAAAGCCC CAAAAACAGG AAGATTGTAT AAGCAAATAT
1101  TTAAATTGTA AACGTTAATA TTTTGTTAAA ATTCGCGTTA AATTTTTGTT
1151  AAATCAGCTC ATTTTTTAAC CAATAGGCCG AAATCGGCAA AATCCCTTAT
1201  AAATCAAAAG AATAGCCCGA GATAGGGTTG AGTGTTGTTC CAGTTTGGAA
1251  CAAGAGTCCA CTATTAAAGA ACGTGGACTC CAACGTCAAA GGGCGAAAAA
1301  CCGTCTATCA GGGCGATGGC CCACTACGTG AACCATCACC CAAATCAAGT
1351  TTTTTGGGGT CGAGGTGCCG TAAAGCACTA AATCGGAACC CTAAAGGGAG
1401  CCCCCGATTT AGAGCTTGAC GGGGAAAGCG AACGTGGCGA GAAAGGAAGG
1451  GAAGAAAGCG AAAGGAGCGG GCGCTAGGGC GCTGGCAAGT GTAGCGGTCA
1501  CGCTGCGCGT AACCACCACA CCCGCCGCGC TTAATGCGCC GCTACAGGGC
1551  GCGTAAAAGG ATCTAGGTGA AGATCCTTTT TGATAATCTC ATGACCAAAA
```

1601 TCCCTTAACG TGAGTTTTCG TTCCACTGAG CGTCAGACCC CGTAGAAAAG

1651 ATCAAAGGAT CTTCTTGAGA TCCTTTTTTT CTGCGCGTAA TCTGCTGCTT

1701 GCAAACAAAA AAACCACCGC TACCAGCGGT GGTTTGTTTG CCGGATCAAG

1751 AGCTACCAAC TCTTTTTCCG AAGGTAACTG GCTTCAGCAG AGCGCAGATA

1801 CCAAATACTG TTCTTCTAGT GTAGCCGTAG TTAGGCCACC ACTTCAAGAA

1851 CTCTGTAGCA CCGCCTACAT ACCTCGCTCT GCTAATCCTG TTACCAGTGG

1901 CTGCTGCCAG TGGCGATAAG TCGTGTCTTA CCGGGTTGGA CTCAAGACGA

1951 TAGTTACCGG ATAAGGCGCA GCGGTCGGGC TGAACGGGGG GTTCGTGCAC

2001 ACAGCCCAGC TTGGAGCGAA CGACCTACAC CGAACTGAGA TACCTACAGC

2051 GTGAGCTATG AGAAAGCGCC ACGCTTCCCG AAGGGAGAAA GGCGGACAGG

2101 TATCCGGTAA GCGGCAGGGT CGGAACAGGA GAGCGCACGA GGGAGCTTCC

2151 AGGGGGAAAC GCCTGGTATC TTTATAGTCC TGTCGGGTTT CGCCACCTCT

2201 GACTTGAGCG TCGATTTTTG TGATGCTCGT CAGGGGGGCG GAGCCTATGG

2251 AAAAACGCCA GCAACGCGGC CTTTTTACGG TTCCTGGCCT TTTGACAGCC

2301 TTTTGCTCAC ATATAATGTG AGTTAGCTCA CTCATTAGGC ACCCCAGGCT

2351 TTACACTTTA TGCTTCCGGC TCGTATGTTG TGTGGAATTG TGAGCGGATA

2401 ACAATTTCAC ACAGGAAACA GCTATGACCA TGATTACGCC AAGCTACGTA

2451 ATACGACTCA CTAGTGGGCA GATCTTCGAA TGCATCGCGC GCACCGTACG

2501 TCTCGAGGAA TTCCTGCAGG ATATCTGGAT CCACGAAGCT TCCCATGGTG

2551 ACGTCACCGG TTCTAGATAC CTAGGTGAGC TCTGGTACCC TCTAGTCAAG

2601 GCCTTAAGTG AGTCGTATTA CGGACTGGCC GTCGTTTTAC AACGTCGTGA

2651 CTGGGAAAAC CCTGGCGTTA CCCAACTTAA TCGCCTTGCA GCACATCCCC

2701 CTTTCGCCAG CTGGCGTAAT AGCGAAGAGG CCCGCACCGA TCGCCCTTCC

2751 CAACAGTTGC GCAGCCTGAA TGGCGAATGG CGCTTCGCTT GGTAATAAAG

2801 CCCGCTTCGG CGGGCTTTTT TTT