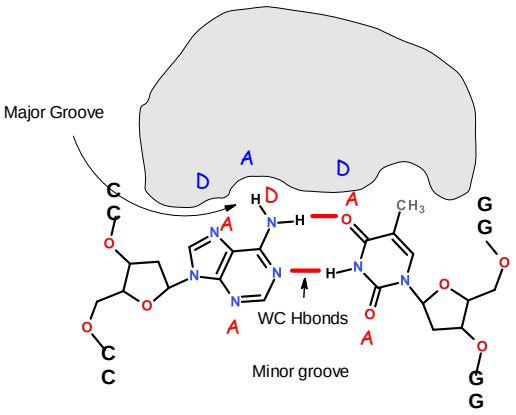
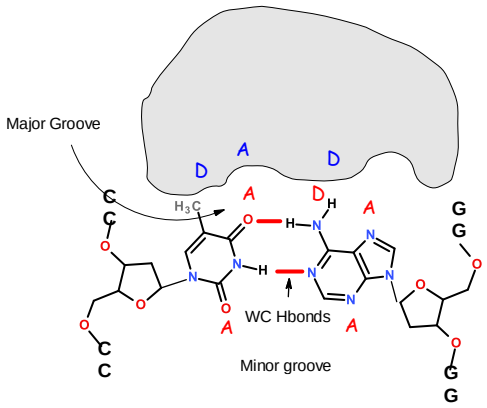
Note: This is an optional problem set that you can complete to replace a lower scoring problem set from earlier in the semester. If you do not submit it for grading, you are strongly encouraged to work through these problems in preparation for the final exam.



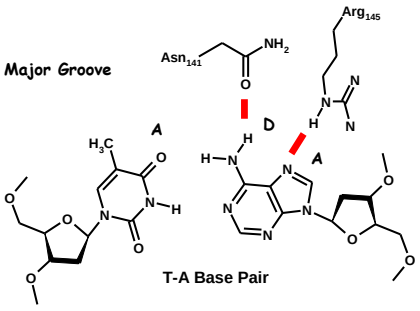


1. (10 pts, 15 min) An AT basepair is shown, contained within the sequence CCACC.
   i) Indicate the Watson-Crick hydrogen bonds (2 pts).
   ii) Identify all hydrogen bond donors and acceptors in both the major and minor groove on both (2 pts).
   iii) Label the major and minor groove (2 pts).
   iv) Assuming that a protein binds in the major groove, and contacts both the A and T bases, what will the DNA binding site on the protein look like, i.e. add the appropriate information to the gray area in the above diagram. Hydrogen bond donors and acceptors in the protein are shown. For the CCACC sequence (left diagram). Note that this protein will bind weakly to CCTCC (right diagram) because the hydrogen bond donors and acceptors are in the wrong position.

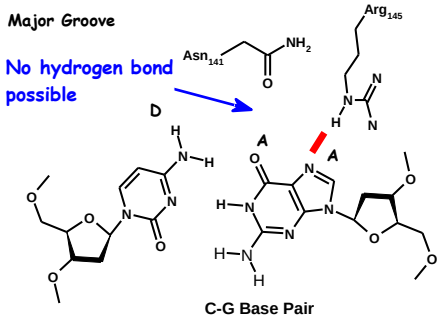
2. (8 pts, 20 min) The binding of proteins to short DNA fragments, such as restriction endonuclease sites, can be measured with equilibrium dialysis. The following set of experiments was performed using a **1 nM** solution of EcoRI inside the dialysis bag. The binding of two different double stranded DNA sequences was studied. The concentration of protein with DNA bound ([ML]) is shown below each DNA sequence.

Free [DNA]	DNA Bound to Protein $G_5A_6A_7T_8T_9C_{10}$ $C_{10}T_9T_8A_7A_6G_5$	DNA Bound to Protein $G_5A_6G_7T_8T_9C_{10}$ $C_{10}T_9C_8A_7A_6G_5$
10^{-9} M	0.50 nM	0.1 nM
10^{-8} M	0.90 nM	0.50 nM
10^{-7} M	0.99 nM	0.90 nM

GAATTC Recognition Sequence



GAGTTC Recognition Sequence



   i) Determine the dissociation constant, K_D, for the binding of each DNA to EcoRI (2 pts)

The fractional saturation is calculated as: Y=[ML]/Mtot. The "DNA bound to protein" is ML and M_{tot} is 1 nm. When (ML) = 0.5 nM, then Y = 0.5 and that DNA concentration is the K_D

K_D for the first sequence is 10^{-9} M

K_D for the second sequence is 10^{-8} M

   ii) Calculate the ΔG^o for the binding of Eco R1 to *each* DNA (2 pts) T=300K.

$\Delta G^o = -RT \ln (1/K_D)$

1st Sequence = $-(8.31)(300)\ln(10^{-9}) = -51.6$ kJ/mol

2nd Sequence = $-(8.31)(300)\ln(10^{-8}) = -45.9$ kJ/mol

iii) Explain, in terms of molecular interactions between Eco RI and the nucleotide bases, the difference in binding affinity of the two DNA sequences. Do the differences in the molecular interactions account for the ΔG^o difference that is measured in the binding experiment? (4 pts)

EcoR1 accepts a hydrogen bond from the amino group on A (e.g. the base is the donor). In the GAGTTC sequence the A has been replaced by a G. The G residue can only accept a hydrogen bond at that position, therefore the loss of binding energy corresponds to a loss of a hydrogen bond between the Asn and the base, worth about 5.7 kJ/mol (the difference in the ΔG^o values). Note that the Arg residue can still from a hydrogen bond with the G.

Note: The interaction of an Asn and Arg sidechains from EcoR1 with A_7 in the recognition sequence is shown to the right. You might find it useful to sketch the interactions between EcoR1 and the TA and CG basepairs in your answer. You should assume that the positions of these sidechains are the same in the wild-type and mutant protein.

3. (12 pts, 15 min) One of the Jmol pages associated with this problem set shows two DNA binding proteins - the methionine repressor and the lambda repressor. View this page and answer the following questions:

i) Both of these proteins control gene expression by preventing RNA polymerase from synthesizing mRNA. However, they bind to DNA by inserting different secondary structure element into a groove in the DNA. View the Jmol page and state the secondary structure that is used by each protein and determine whether the protein binds in the major groove or the minor groove (2 pts).

The methionine repressor has inserted a two-stranded β sheet into the DNA major groove while the lambda repressor has inserted an α-helix into the DNA major groove.

ii) Based on your answer to *part i),* why do you think this groove is preferentially used by most DNA binding proteins? (2 pts).

It is larger and can accommodate secondary structures.

iii) Click on the button marked "Lys" under the Jmol image of the methionine repressor. This will zoom in on the structure and highlight two Lys residues by placing a surface around their sidechain atoms. Briefly describe how each of these residues is interacting with the DNA and whether the interaction would be sequence specific or non-sequence specific (3 pts).

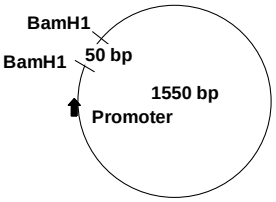
One lysine reside interacts with the phosphate in a non-sequence specific manner (each residue in DNA has a phosphate). The second lysine forms a non-Watson-Crick hydrogen bond with a base in the major groove, this interaction is sequence specific.

iv) Based on your answer to *part iii)*, would you expect the binding of DNA to the methionine repressor to be affected by salt or not? If so, how would increasing the salt affect the binding (5 pts).

The interaction between the lysine and the phosphate would be affected by salt.

The salt would decrease the affinity, or increase the K_D, by reducing the effective charges on the lysine and phosphate.

4. (8 pts, 15 min) You are interested in expressing a protein from the influenza virus in E. coli for vaccine development. You only need to express part of the protein in bacteria since it will be used for a vaccine. You decide to use PCR to amplify a segment of DNA that begins with TTG and ends with GCG; the DNA to be present in the final PCR product is shown in uppercase. You want to add restriction endonuclease sites at both ends of the PCR product to facilitate cloning into your



expression vector (shown below on the right). Note that your expression vector lacks a start codon and a stop codon, so these will have to be added during PCR.

```
          |-----------------------------400 bp -----------------------------|
ggcctaggtTTACGTTATGTTGCAGTATATAGGCGA-----------GGGTATAGCGCTAGCGATAGTTAATTGCGcgcgttagtagta
ccggatccaAATGCAATTCAACGTCATATATCCGCT-----------CCCATATCGCGATCGCTATCAATTAACGCgcgcaatcatcat
         LeuArgTyrValAlaValTyrArgArg..........GlyTyrSerAlaSerAspSer
```

Determine the sequence of both the right and left primers that you would use for the PCR reaction. Your primers should be 16 bases in length, including the additional nucleotide bases that comprise the restriction endonuclease sites, plus the start and stop codons (6 pts).

Considering the left primer only, what would the correct annealing temperature to use during the PCR reaction (2 pts) ? Use the following formula: T_M **= 81.5 +0.41(%GC)-675/N**

Example T_M calculation for the following primer: GGGAAACCCGTA

%GC=7/12=58.3%, N=12. Therefore, T_M = 81.5 + 0.41x58.3 - 675/12 = 49.1 C

The 5'-end of the primer should have the sequence for BamH1, the start codon (underlined), followed by the sequence that matches the upper strand of the DNA to be amplified. A potential 20 base primer is:

```
**   **   *    ** ←Location of G/C
GGATCCATGTTACGTT
12345678901234 56
          1
```

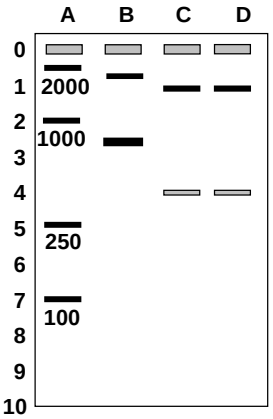
We need to calculate its melting temperature. When calculating T_M, I usually count the bases that anneal to the DNA after the restriction site has been incorporated into the template by the first few PCR cycles. However, you also have to keep in mind that the primer must anneal during the first cycle as well, using only the bases that are complementary to the template. Often both are calculated and the annealing temperature is changed after the first few cycles. The T_M calculation for the above 16 residue primer is 81.5 +0.41 x (100*7/16) - 675/16 = 57.5 The annealing temp should be ~52 C.

As for the right primer, its sequence would be (the sequence of the lower right strand + the complement to the stop codon (TAA) + the BamH1 site, it's Tm is 61C.

```
5'-GGATCCTTACGCAATT-3'
   1234567890123456
```

Note that this primer contains a stop codon (underlined), as required.

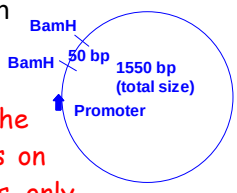
5. (12 pts, 20 min) Continuing on from question 5. You digest the plasmid and the PCR product with BamH1, mix the two digestion, and then add DNA ligase. After ligating your PCR product into the vector, you transform cells, isolate plasmid DNA from two bacterial clones and perform a restriction digest and separate the DNA fragments by size using gel electrophoresis. An image of the gel is shown on the right. Standards of the indicated size are in lane A (units are basepairs) the uncut plasmid from clone 1 is in lane B, and the BamH1 digest of that clone is shown in lane C. Lane D contains the BamH1 digest of the second clone.

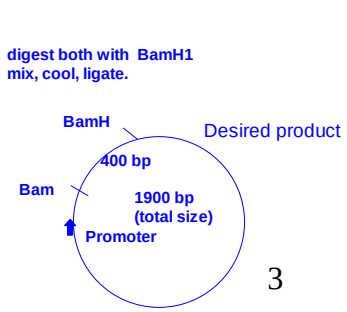


i) Why are there two bands in lane B (5 pts)?

One band is supercoiled (lower, more intense), the other (upper) is relaxed due to a nick that relieves the supercoils.

iii) The first clone does not produce any protein in bacteria, while the second clone does. Why (5 pts)? Hint, draw the possible ligation products.





The approach to generating the plasmid is shown on the right. Because the PCR product has the Bam sites on both ends the fragment can be inserted in two ways, only one of which will give the right mRNA sequence. In the other orientation the non-coding strand will be used to make the mRNA, giving the incorrect

sequence:

Take home: It is always best to have two different restriction sites so that the fragments can only be inserted one way.

6. (5 pts, 10 min) You want to express myoglobin from dinosaurs in bacteria to see how changes in the sequence of the myoglobin lead to efficient O_2 storage in bird muscle tissue. The first five amino acids of myoglobin are:

GlyTrpArgAsp…

Give the DNA sequence that you would use to code for the first amino acids in myoglobin. You should consult the following reference for E. coli codon usage when considering which codons to use. http://www.faculty.ucr.edu/~mmaduro/codonusage/codontable.htm (select the most frequent codon in class II, which is the codon usage in highly expressed genes).

The following DNA sequence could be used:

GlyTrpArgAsp

GGTTGGCGTGAC

7. (14 pts, 20 min) You have obtained the DNA that encodes a mutant HIV protease gene from a drug resistant patient. The sequencing reaction was performed on this DNA and the resultant trace is shown below. The wild-type sequence is also given to facilitate identifying the location of the mutation.

i) Determine the nucleotide change(s) in the mutant protease gene. Note, determine the reading frame by consulting the DNA and protein sequence of HIV protease (2 pts). The location of the change is around position 41 in the DNA sequence data (see top of next page, red circle). This corresponds to sequences around amino acid 82 in HIV protease (underlined on the HIV protease sequence on the next page).

```
          ProAspAsn
Mutant:   CCTGACAACAT
WT:       CCTGTCAACAT
          ProValAsn
```

ii) Give the sequence of the primer that was used to generate the sequencing data, assume the primer is 10 bases in length and the first dideoxy base that is added can be detected in the sequencing data(2 pts).

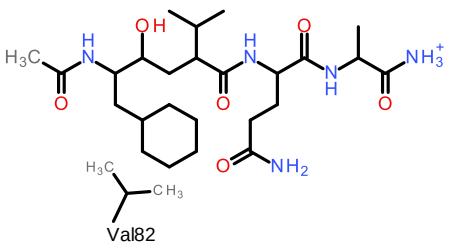
AGAAATCTGT - this is highlighted on the sequence.

The primer-template that would be elongated by polymerase during sequencing is:

```
          AGAAATCTGT
...GAGTATCTTTAGACACCTGTATTTCGATATCCATGT…
```

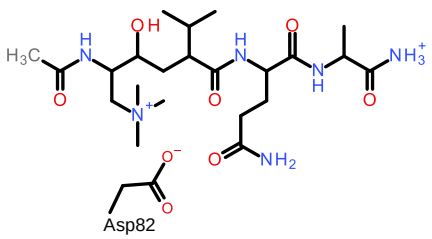
The first base to be added is a G, the second a G, the third a A, which gives the sequencing data.

iii) Determine the amino acid change induced by this mutation and comment on whether you expect this mutation to affect the substrate specificity or the catalytic efficiency of the protease (4 pts).



The codon for Val82 is changed to GAC, which is Aspartic acid. This would affect specificity, not catalytic ability (Asp at position 25 is responsible for catalysis).

iv) How might you modify the original, cyclohexane containing drug (shown on the right) such that it would bind, and effectively inhibit the activity of the virus containing the mutant protease?(6 pts)



Add a pos. charged group to the drug, to enhance the electrostatic interaction with Asp82 in the mutant protein, a quaternary amine is shown here.

Region of HIV DNA Coding for HIV protease.

5'-ggagccgatagacaaggaactgtatcctttaacttcCCTCAGATCACTCTTTGGCAA^{57}
ProGlnIleThrLeuTrpGln$_{7}$

^{58}CGACCCCTCGTCACAATAAAGATAGGGGGGCAACTAAAGGAAGCTCTATTAGAT**ACAGGA**^{117}
$_{8}$ArgProLeuValThrIleLysIleGlyGlyGlnLeuLysGluAlaLeuLeuAspThrGly$_{27}$

^{118}GCAGATG**A**TACAGTATTAGAAGAAATGAGTTTGCCAGGAAGATGGAAACCAAAAATGATA^{177}
$_{28}$AlaAspAspThrValLeuGluGluMetSerLeuProGlyArgTrpLysProLysMetIle$_{47}$

^{178}GGGGGAATTGGAGGTTTTATCAAAGTAAGACAGTATGATCAGATACTCAT**AGAAATCTGT**^{237}
$_{48}$GlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleLeuIleGluIleCys$_{67}$

^{238}GGACATAAAGCTATAGGTACAGTATTAGTAGGACCTACACCTGTCAACATAATTGGAAGA^{297}
$_{68}$GlyHisLysAlaIleGlyThrValLeuValGlyProThrPro<u>Val</u>AsnIleIleGlyArg$_{87}$

^{298}AATCTGTTGACTCAGATTGGTTGCACTTTAAATTTTcccattagccctattgagact^{354} -3'
$_{88}$AsnLeuLeuThrGlnIleGlyCysThrLeuAsnPhe