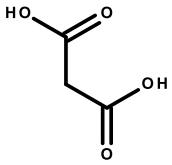
This exam has **13 pages** and is out of **150 points**. You should allot 1 min/point. On questions with choices all of your attempts will be graded and you will receive the best grade. Use the space provided, or the back of the preceding page. A codon table and codon usage information is given on the 14th page.

**1.** (6 pts) Maleic acid contains two carboxylic acid groups, one with a pKa of 2.0 and a second with a pKa of 4.0. The fully protonated form of maleic acid is shown on the right. Briefly describe how you would prepare 0.25 L of a 0.1 M buffer at pH=2.0, assuming that you are starting with the disodium salt of the acid. You should give the moles of sodium malate and HCl that you would need to make this buffer.




Since the pH is the same as the pKa, the fraction protonated = fraction deprotonated = 0.5.
Since this is a diprotic acid, and you are beginning with the fully deprotonated species (disodium salt), you would have to add 1.5 equivalents of HCl. One equivalent to fully protonate the first group and then ½ equivalent to protonate the second $-COO^-$ groupl.

Moles of HCl are 1.5 eq X 0.25 L X 0.1 moles/L = 0.0375 moles (-1 pt if just ½ equivalent was added)
This would be added to 0.25L x 0.1 M = 0.025 moles of the deprotonated acid.

**2.** (3 pts) All weak acids have "buffer" regions near any of their pKa values. Briefly explain why the pH of the solution is resistant to change in these regions.

In this region, any added base (or acid) is used to deprotonated (or protonate) the weak acid, thus the concentration of hydrogen ions does not change much, so the pH does not change much.

**3.** (10 pts) Most proteins generally consist of secondary structural elements.
i) Name the **two** common secondary structures. (2 pt)
ii) Describe the overall structure of **one** of these, including the position of sidechains (3 pts).
iii) What are the principle interactions that stabilizes both of these structures? (2 pts).
iv) How are the interactions that you discussed in part iii) depicted on a Ramachandran plot? (3 pts).

i) and ii)
Alpha helix = hydrogen bonds parallel to helix axis, sidechains project out.
Beta sheet = hydrogen bonds perpendicular to stand directions, sidechains project above and below the sheet.
iii) They are stabilized by mainchain hydrogen bonds and favorable van der Waals interactions.
iv) The contour lines on the Ramachandran plot show regions of favorable van der Waals interactions (1/2 credit for just sketching a Ramachandran plot)



Points:_____

**4.** (10 pts) Draw the chemical structure of a tri-peptide, i.e. three amino acids linked together. The sidechain of the first amino acid is a methyl group (-$CH_3$), the second is just a hydrogen atom, and the third is an isopropyl group ($CH_3$-CH-$CH_3$) (4 pts).

i) Label a peptide bond in your drawing and indicated whether it is drawn in the cis or trans form. (2 pts)

Trans form

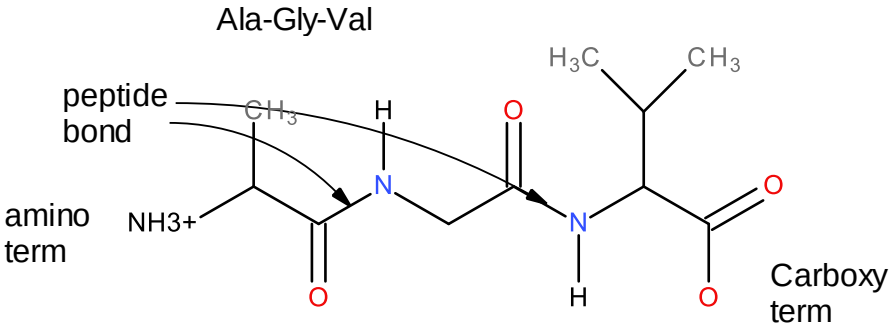
ii) Which form of the peptide bond is more stable, cis or trans, and why? (2 pt)

Trans – reduces unfavorable van der Waals between mainchain atoms.

iii) Label the amino and carboxy terminus of the protein (1 pt).

See diagram

iv) Give the names of the amino acids that you have drawn and write out the sequence of the protein (1 pt)



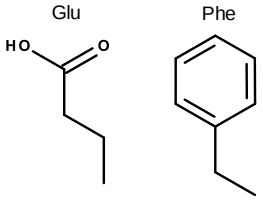

**5.** (4 pts) The following amino acid sequence is found in a soluble globular protein:

**Glu-Phe-Glu-Phe-Glu-Phe**

i) What is the likely secondary structure of this sequence? Why? (2 pts)

ii) Where do you expect the Phe residues to face, the interior of the protein, or to the solvent? Why? (1 pt)

iii) If this was an integral membrane protein, where would you most likely expect to find the Phe residues? Why? (1 pts)




i) B-sheet, because of the alternating polar – non-polar nature of the sidechains, so that the non-polar side chains would project into the interior of the protein.

ii) The non-polar Phe would face the interior, due to the hydrophobic effect.

iii) Facing the non-polar lipids, i.e. on the exterior of the integral membrane protein.

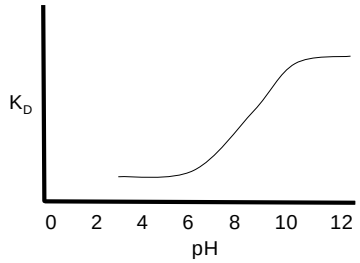


Points:_____

**6.** (7 pts) A DNA binding protein binds to DNA in a non-specific manner. The protein contains lysine (-$NH_2$) residues which are close to the DNA.

i) What group on the DNA would the lysine residue most likely interact with? (1 pt)

ii) Assume that you measured the dissociation constant ($K_D$) as a function of pH. Plot the $K_D$ as a function of pH over the range of 0 to 12. You can assume the pKa value of Lys is 9. Please justify your answer (6 pts).



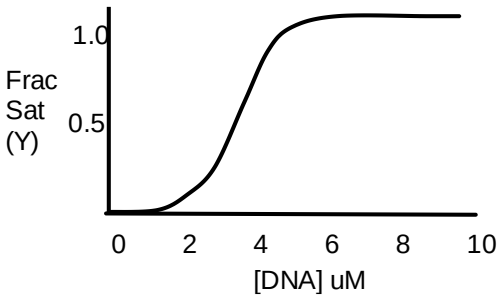

i) Phosphate groups on the DNA (because the binding is non-specific) by an electrostatic interaction between the + charge on the lysine and the negative charge on the phosphate.

ii) The binding would be strong (low Kd) when the lysine is protonated, the Kd would increase as the lysine was deprotonated, as shown on the right. Note that the curve is only drawn to pH=3, at this pH the phosphate backbone on the DNA would begin to become protonated, causing a loss of the electrostatic interaction (it was not necessary to discuss phosphate ionization).

**7.** (6 pts) The binding curve for the same protein in the previous question is shown on the right, for a single pH value. The protein has a molecular weight of 100 kDa protein binds one or more short DNA oligonucleotide (say 10 bases, 6 kDa). Please answer the following questions:

i) What is the $K_D$ for binding? (1 pt)

ii) Based on the shape of the binding curve, how many DNA molecules are likely binding to the protein, one or more? *Justify your answer*. (2 pts)




iii) What additional plot might you do to confirm your hypothesis to part ii? How would this plot be used to confirm your hypothesis? (2 pts)

iv) How could you determine the binding constant using size exclusion chromatography? (1 pt)

i) 3 uM, where Y = 0.5

ii) It is cooperative (positive), so more than one DNA must be binding.

iii) Do a Hill plot, the slope of the line when it crosses the axis should be something other than one.

iv) Size exclusion chromatography can be used to separate the protein from the DNA protein complex since they differ in size, giving free protein and bound protein: Y=bound/total.



Points:_____

**8.** (6 pts) Please do **one** of the following choices. Be sure to discuss **all** the important enthalpic **and** entropic considerations in your answers.

**Choice A:** A mutation in a protein converts a buried phenylalanine residue to a threonine residue (Threonine sidechain is –C-C-OH). How will this affect the stability of the protein?

**Choice B:** A mutation in a DNA sequence converts a G-C base pair to an A-T base pair. How will this affect the stability of the double stranded (ds) form of DNA? Will it make it more stable or less stable? Why?

**Choice A:**

Large non-polar residue is replaced by a smaller polar residue.

Less stable because of loss of van der Waals (enthalpic) and a reduced hydrophobic effect (entropic)

**Choice B:**

Less stable:

two Watson crick H-bonds instead of three

weaker van der Waals between the stacked bases.

(both enthalpic)

**9.** (8 pts) Please answer the following questions on enzymatic activity. You can use an example from class to illustrate your answer, but it is not necessary to give specific details about any particular enzyme.

i) Why are enzymes specific for particular substrates, what is the relationship between $k_{off}$, $k_{CAT}$ for good and bad substrates? (1 pt).

ii) How do enzymes increase the rate of catalysis? Provide a general principle that holds for all enzymes (6 pts).

iii) Why are transport proteins (e.g. $K^+$ channel) considered (at least by me) to be enzymes, even though they do not change the chemical structure of the substrate? (1 pt)

i) Good substrates have many interactions with the active site, leading to **small off-rates**. Therefore, the substrate is bound long enough for the chemical transformation to occur, i.e. $k_{off}<k_{cat.}$ Full credit for relating many interactions to a slow off-rate.

ii) They reduce the energy of the transition state, which increases its concentration and therefore the rate (3 pts).

This occurs with all enzymes because functional groups are pre-ordered, so there is no decrease in energy as the substrate-enzyme complex goes to the transition state. (3 pts)

iii) They catalyze the reaction of moving the ion across the membrane, reducing the activation energy, e.g. by favorable interaction between the K+ ion and the C=O groups in the selectivity channel.

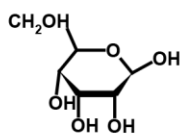


Points:_____

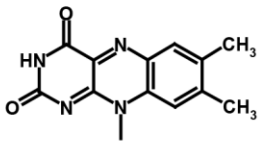
**10.** (8 pts, 1/2 pt each) On the right are a series of 15 biochemical structures (A-O), on the left is a list of names or descriptions. Indicate the correct match by writing the letter next to the description or name. Note that a structure *should* only be used once and one item should not have a structure associated with it.

| Description | Match (A-O) |
|---|---|
| 1. Product of glycolysis | I |
| 2. Product of anaerobic metabolism in humans. | K |
| 3. Product of anaerobic metabolism in Yeast. | G |
| 4. Fatty acid | J |
| 5. Triglyceride | L |
| 6. Phospholipid | D |
| 7. Cholesterol | - |
| 8. Six carbon aldose. | A |
| 9.Saccharide found in bacterial cell walls | H |
| 10. Disaccharide | F |
| 11. Electron carrier in the TCA cycle and fatty acid oxidation. | B |
| 12. Electron carrier in electron transport chain. | N |
| 13. Final electron acceptor in most species. | O |
| 14. Nucleotide normally found in DNA | E |
| 15. Nucleotide normally found in RNA | C |
| 16. Nucleotide that is used in DNA sequencing. | M |

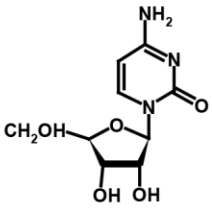
A



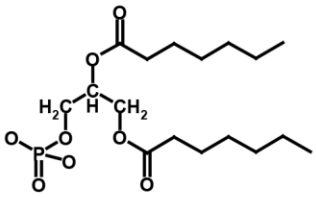
B



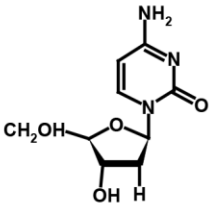
C



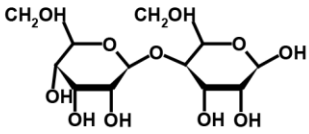
D



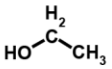
E



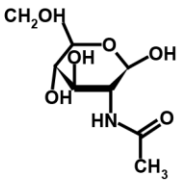
F



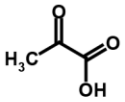
G



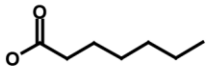
H



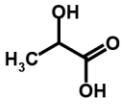
I



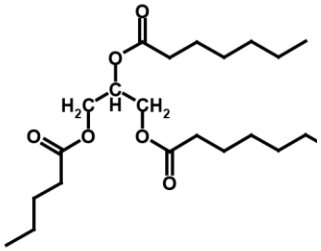
J



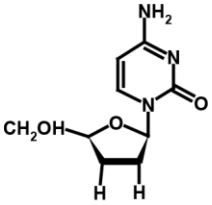
K



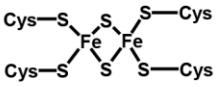
L



M



N



O

$O_2$

**11**. (1 pt) You are given a sample of either protein or a nucleic acid. What simple method could you use to determine whether it is protein or nucleic acid [Hint: You could also measure the concentration with this method]?

Protein and DNA have different absorption maxima, DNA = 260nm, protein =280nm. Simply take the absorption spectra.



Points:_____

**12**. (6 pts) Please do **both** parts of this question:

i) Briefly describe the most important characteristics of allosteric systems, including activators and inhibitors (4 pts).

ii) Illustrate your answer with **one** of the following topics from the course: a) Oxygen delivery, b) altitude adjustment, c) enzyme inhibitors, d) metabolic regulation (glycogen or glycolysis), or e) regulation of DNA transcription by the lac repressor (2 pt).

i)

- Enzyme is in two forms – relaxed (active) or tense (inactive), these two forms are in **equilibrium** with each other.
- Activators and inhibitors binding away from the active site, changing the shape of the enzyme.
- Inhibitors shift the equilibrium to the tense form/decreasing the activity (or binding affinity)
- Activators shift the equilibrium to the relaxed form/increasing the activity (or binding affinity)

a) oxygen transport – oxygen is an allosteric activator; the binding of one oxygen increases the binding of subsequent oxygens – allowing full saturation in the lungs and effective release to the tissues.

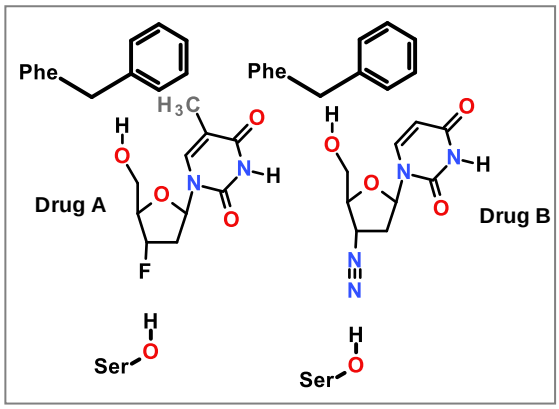
b) altitude adjustment – bisphosphoglycerate is an allosteric inhibitor, it is increased at high altitudes. It shifts the binding curve to the right (lower affinity) but changes its shape so an equivalent amount of oxygen is delivered at high altitude.

c) Mixed type inhibitors bind at a different location than the active site and cause an allosteric change in the active site that inhibits the enzymes.

d) ATP is an allosteric inhibitor of PFK. F26P, AMP, ADP activate PFK. Protein phosphorylation is a form of allosteric regulation – used to regulate glycogen synthesis and degradation.

e) Lac repressor is released from DNA by IPTG, IPTG causes a change for the relaxed DNA binding form to the tense non-DNA binding form.

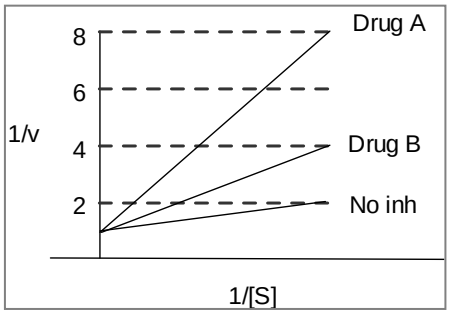
**13.** (8 pts) The two drugs on the right are used to inhibit the growth of viruses by inhibiting the DNA polymerase that the virus uses to replicate its genetic material. The Ser and Phe are amino acids from the polymerase that interact with the drugs.




i) Are these drugs based on purine or pyrimidine bases (circle correct answer)? (1 pt).

ii) Are these drugs competitive inhibitors or mixed type? Justify your answer (2 pt).

iii) The ability of the drugs to inhibit polymerization was measured using steady-state enzyme kinetics with a constant amount of inhibitor. The resultant double reciprocal plots are shown on the right. Which inhibitor is more effective? Drug A or Drug B? Justify your answer with reference to **both** the double reciprocal data **and** the interaction between the drug and the polymerase (4 pts)




i) pyrimidine bases.

ii) Competitive – they look like the substrate (nucleotides) (1 pt) and the enzyme kinetic data shows no change in Vmax (same intercept on y-axis) (1 pt)

iii) Drug A is the better inhibitor:

Slope of plot is larger, therefore a is larger, therefore $K_I$ is SMALLER. No need to calculate $K_I$, which you cannot do because the inhibitor concentration is not given. (2 pts)

Drug A has a more favorable H-bond between the Ser and the electronegative fluorine (1 pt)

Drug A has a more favorable vdw & hydrophobic effect between Phe and -$CH_3$ group. (1 pt)



Points:_____

**Bonus 1**: These drugs are actually pro-drugs, in that they need to be converted to another compound by cellular enzymes before they are effective. What modifications to these drugs would likely occur to make them bind effectively to a polymerase? What type of enzyme would perform these modifications? (1 pt)

Conversion to nucleoside triphosphates by kinases.

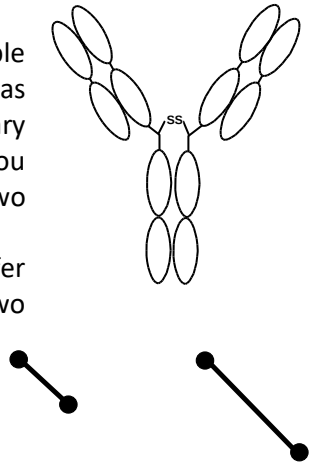
**Bonus 2**: Although these drugs bind to the DNA polymerase and reduce its activity, they also affect polymer growth by acting as suicide inhibitors. How does this occur? (1 pt)

They lack a 3' OH group, and therefore once incorporated the growing DNA chain would be terminated, in the same fashion as would occur with ddNTPs that are used in DNA sequencing.

**14.** (5 pts) Please do **one** of the following choices.

**Choice A:** The quaternary structure of the immunoglobulin was determined by simple techniques 20 years before the X-ray structure of an immunoglobulin was determined. What techniques were initially used to determine the quaternary structure? Briefly describe the techniques, the data you would obtain, and how you would use this data to substantiate the structure shown on the right. (Note, the two heavy chains are linked by a disulfide bond).

**Choice B:** The diagram on the right shows two simple diatomic molecules that differ only in their bond lengths. Explain why the scattered X-rays from these two molecules would be different such that their structures could be determined by X-ray diffraction.




**Choice A:**

- The overall molecular weight would be obtained by size exclusion chromatography.
- The subunit molecular weights would be obtained by SDS-PAGE.
- One experiment would be done without BME, which would give the molecular weights of any chains linked by disulfide bonds, in this case the two heavy chains would be linked together, but the two light chains would be separate (not the case for most immunoglobulins, the light is also disulfide bonded to the heavy).
- SDS PAGE with BME (beta mercaptoethanol) would give the isolated chains, in a ratio of 1:1 of light and heavy chains.
- In order to agree with the overall molecular weight from gel filtration, there would have to be two light and two heavy chains.

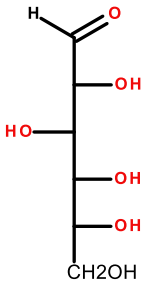
**Choice B:**

- The relative position of the atoms affects the interference between the X-rays scattered from each atom.
- The difference in interference leads to a difference in intensities.
- The difference in intensities can be used to determine the relative positions.



Points:_____

**15**. (5 pts) The structure of glucose is shown on the right.



glucose

i) Draw β-glucose in its ***furanose*** form using the reduced Haworth representation. (1 pt)

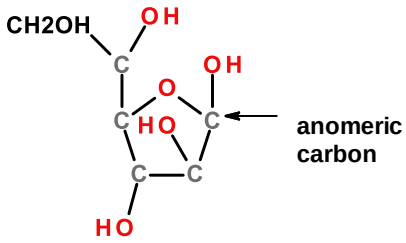
ii) Indicate the location of the new chiral centers on the ring of glucose, what is this new center called? (1 pt)

iii) Sketch, or describe, the chemical structure of any **one** of the following carbohydrates (2 pts)

a) lactose b) sucrose

c) glycogen d) cellulose

iv) Which of the above (a-d) could be used as an energy source? Indicate **all** possibilities (1 pt).




i) See diagram.

ii) The C1 is becomes chiral and is called the anomeric carbon. Note that the beta-form is drawn, the -OH on C1 could also be down (alpha)

iii) lactose = galactose+glucose, sucrose=glucose+fructose, glycogen = glucose α(1-4) + α(1-6) branches, cellulose = glucose β(1-4) linkages, not branched.

iv) Preferred answer: All but cellulose can be used by humans for energy.
Accepted answer: All are possible energy sources, cellulose can only be used by organisms that have cellulases.

**16**. (4 pts) Yeast can produce ethanol obtained from glucose. What biochemical pathways are involved in the production of ethanol? What growth conditions would maximize the production of ethanol?

- Glucose enters glycolysis and is converted to pyruvate.
- Under low oxygen conditions the pyruvate is converted to ethanol (via acetylaldehyde) to regenerate NAD+ as an electron acceptor in glycolysis.



Points:_____

**17.** (6 pts) Please do **one** of the following questions.

**Choice A:** Pretend you just finished the Pittsburgh marathon. As a consequence, your glycogen levels and ATP levels in the liver are quite low. Discuss the process, with the major focus on regulation in your answer, by which your glycogen levels and ATP levels are restored as you eat lots of carbohydrates after the race.

**Choice B:** Pretend that you didn't run the Pittsburgh marathon, but lounged around all morning eating pancakes (with maple syrup of course). Consequently, the ATP levels in your liver cells are high. You are walking to campus and a ferocious dog, with rather large teeth, begins to chase you. What hormone is released and how does this hormone affect your ability to escape from the dog? You should discuss how this hormone will affect the regulation of metabolic pathways that produce glucose.

**Choice A:**

- Blood glucose levels would become high
- Causing release of insulin
- Insulin would cause dephosphorylation of enzymes in cell.
- Glycogen synthase is active and glucose would be stored in glycogen
- F26P levels will rise, activating PFK in glycolysis.
- Oxidation of glucose will restore ATP levels.

**Choice B:**

- Epinephrine would be released from your central nervous system
- Binding to its receptor would signal enzyme phosphorylation due to the activation of kinases.
- This would activate glycogen phosphorylase, causing glucose to be released from glycogen.
- High levels of ATP would allow synthesis of glucose from pyruvate by gluconeogenesis (F26P levels are low so bisphosphatase in gluconeogenesis can be active)

**18.** (5 pts) Please do **one** of the following choices.

**Choice A**: Describe the basic reaction mechanism for a typical DNA polymerase. Discuss why the Gibbs energy for the overall reaction is negative and also comment on the fidelity of the reaction, or why the polymerase is more likely to incorporate the correct base (**Note**: do not discuss removal of an incorrect base, see choice B).

**Choice B**: Discuss the mechanism by which some DNA polymerases remove incorrectly incorporated bases. What is the consequence of lack of this function in HIV reverse transcriptase and how does it affect the treatment of HIV?

**Choice A:**

- Require a template with an annealed primer, with a 3'OH
- dNTPs are brought into the active site,
- matching based on Watson-Crick Hbonds, plus purine-pyrimidine size
- 3'OH attacks phosphate, forming new phosphodiester bond, releasing pyrophosphate
- Pyrophosphate hydrolyzed to inorganic phosphate – large neg free energy drives polymerization by indirect coupling.

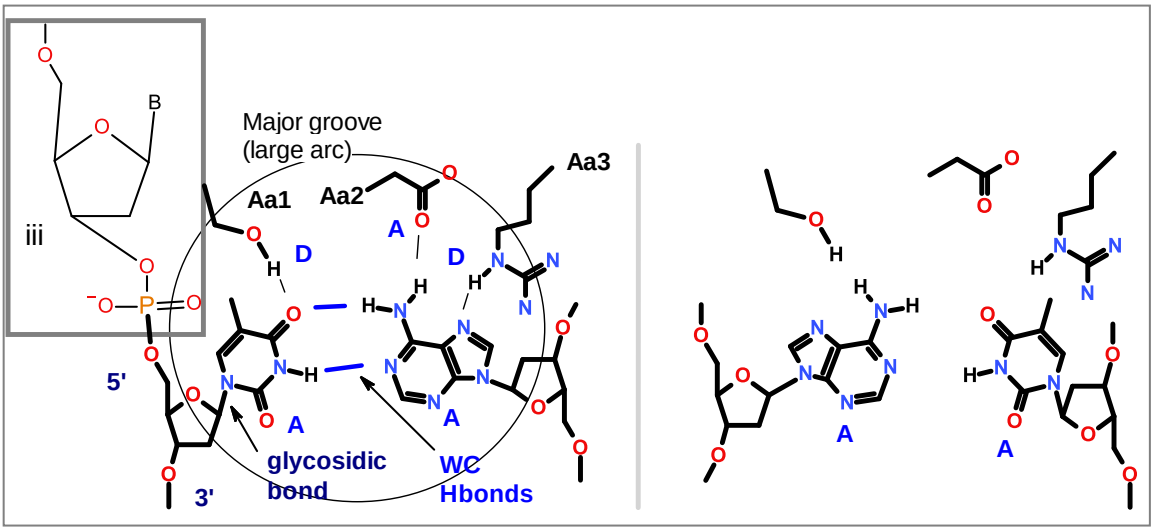
**Choice B:**

- The mismatched base is removed by hydrolysis, by 3' – 5' exonuclease activity.
- If this activity is not present in the polymerase, then mutations will be introduced into the DNA sequence
- These changes can cause changes in the aminoacid sequence of HIV enzymes, such that the drug that inhibits the enzyme can no longer bind.



Points:_____

**19**. (10 pts) A protein binds to a specific sequence of double stranded nucleic acid. Part of the interaction between the protein and the nucleic acid is shown on the left side of the diagram. The amino acid sidechains from the protein are labeled Aa1, Aa2, and Aa3. The reversal of the two bases is shown on the right part of the diagram, along with a duplication of the protein shown in the left panel. The right panel will be useful for parts vii and viii.




i) Label the 5' and 3' carbons of left-most base (1 pt).

ii) Identify the glycosidic bond on the left-most base (1 pt).

iii) Place the appropriate missing atoms in the box labeled "iii" that would be required to connect this T residue to the previous residue. Include all atoms on the previous residue, and indicate that base with "B"(1 pt).

iv) Indicate the "Watson-Crick" hydrogen bonds on the left-most base pair (1 pt).

v) Indicate H-bond donors (D) and acceptors (A) *on the protein* that could potentially interact with the DNA bases (1 pt)

There are three potential H-bonds between the protein and the DNA. The order of these on the protein is: Donor Acceptor Donor

vi) Is this protein binding in the major groove or the minor groove? How did you determine this? (1 pt).

The major groove is indicated by the large arc of the circle that is drawn through the C1' carbons on the ribose.

vii) How would the binding affinity change if the protein bound to the reversed basepair (shown on the right)? You should assume that the structure of the protein does not change. (2 pts)

The affinity would be weaker, because the hydrogen bond donors and acceptors no longer line up (e.g. donor paired with a donor).

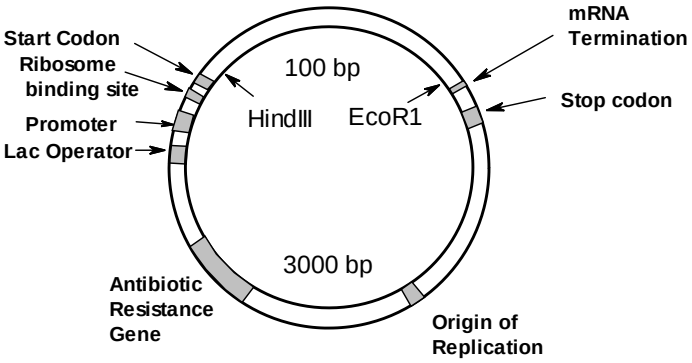
viii) If a protein (not necessarily the one shown in the diagram above) used a similar type of interaction in the other groove, i.e. if the amino acid sidechains entered from the lower part of the diagram, would your answer to part vii change? Why? (2 pts)

The affinity would be the same for either TA or AT because either basepair shows the same pattern of hydrogen bond acceptors in the minor groove, two acceptors on the DNA, in approximately the same location.



Points:_____

**20.** (5 pts) The diagram to the right shows an expression vector with several essential DNA features labeled. **Two** of the labeled DNA sequences are in the wrong order. Identify **both**, and give their correct position. Briefly justify your answer with a description of the role of that DNA sequence in the production of mRNA and/or protein and why its location is incorrect.




The **lac operator** sequence and the promoter are switched. The promoter should come before the operator so that the binding of the lac repressor to the lac operator can interfere with RNA Pol. and mRNA production.

The **stop codon** and the mRNA termination signals are reversed, as it is, the mRNA would have no stop codon so the protein would not be terminated correctly.

**21.** (2 pts) Please do **one** of the following:

**Choice A:** Why is there an "antibiotic resistance gene" in all plasmids? What is its role?

**Choice B:** Why is there an "origin of replication" in all plasmids? What is its role?

**Choice A:** The antibiotic resistance gene makes bacterial that contain the plasmid resistant to antibiotics – so that only bacteria with the plasmid can grow in the presence of the antibiotic, thus preventing the growth of bacteria without the plasmid. The bacteria without the plasmid will not produce the protein of interest.

**Choice B:** When the bacterial cell divides, it replicates its own chromosomal DNA so that the new daughter cells each get a copy of the genetic material. The plasmid must also be replicated and the origin of replication will cause the host cell replication machinery to copy the plasmid as well so that it can be passed on to the daughter cells.

**22.** (6 pts) Please do **one** of the following choices:

**Choice A:** Discuss the roles of the ribosome binding site, the start codon, and the stop codon on the overall process of protein synthesis.

**Choice B:** Briefly discuss the elongation step in protein synthesis. Be sure to address the role of the tRNA binding sites in the process.

**Choice A:**

Ribosome binding site: binds to the rRNA in the 30S small subunit and positions the start codon at the right position on the ribosome, in the P tRNA site.

Start codon: Binds the first tRNA, a special tRNA containing the modified amino acid fMet.

Stop codon: Recruits the protein termination factor to the A site, causing hydrolysis of the protein from the last tRNA.

**Choice B:**

The peptide, attached to the last tRNA used, is in the P site.

The charge tRNA for the next codon enters the A site.

Peptide bond formation occurs, the amino group of the amino acid in the A site attacks the carboxyl end of the growing peptide, transferring the protein to the A site.

The ribosome shifts, moving the peptide-tRNA to the P site.



Points:_____

**23.** (8 pts) The following DNA contains a protein coding sequence of HIV reverse transcriptase. You would like to produce this protein in bacteria so that you could study drug resistant strains of the HIV virus. The sequence, along with the protein translation, is given in bold. The entire coding region of the enzyme is 500 nucleotide bases in length, this is contained within the 10,000 bases of the entire HIV genome.

```
              |------------------500 bases-----------------|
---AGCTGCTCATGCTCCCCACATGCAATTCCTCCCCA......GTGAGGGGGAAATTAACCGCCGGCG---
---TCGACGAGTACGAGGGGGTGTACCTTAAGGAGGGGT......CACTCCCCCTTTAATTGGCGGCCGC---
              MetLeuProThrCysAsnSerSerPro      ValArgGlyLysStop
```

The expression vector (complete diagram shown question 20) contains a HindIII site just after the start codon and a EcoR1 site adjacent to the stop codon that are contained in the vector, i.e. the vector sequence is:

```
                                      HindIII                    EcoR1
--TTAGTAGGGCACCTCAATGAAGCTT---100 bases---GAATTCTTAA---
```

(i.e. the start and stop codons are already in the expression vector/plasmid).

i) Give the sequence of **both** the left and right primers that would generate the desired PCR product so that you could insert the PCR product into the vector. The total length of your primers 12 bases (3 pts).

ii) Briefly explain *why* you wanted to add HindIII and EcoR1 sites to your PCR product (3 pts)

iii) Calculate the $T_M$ for the **left** primer (1 pt). $T_M = 81.5 + 0.41*(\%GC) - 625/N$

iv) Based on this $T_M$ what annealing temperature would you use for PCR? Why? (1 pt)

i) The left primer sequence will begin with the required restriction site (HindIII), followed by the sequence of the top strand at the left boundary. The right primer will begin with the EcoR1 site, followed by the sequence of the bottom strand at the right boundary.

Left primer = A A G C TT C T C C C C [Note start codon is omitted]

Right primer = G A A T T C T T T C C C [Note: stop codon is omitted]

ii) So that the PCR product can be cut with these enzymes to generate single stranded sticky ends. Treatment of the plasmid will generate the same sticky ends, allowing the PCR product to bind to the plasmid to insert the HIV reverse transcriptase gene into the plasmid.

iii) Tm = 81.5 +0.41(100 x (7/12)) – 625/12 = 53.3

iv) The annealing temperature should be 5 degrees below this, or 48.3.

**24**. (2 pts) Please do **one** of the following choices. Regardless of your choice, briefly describe how the process works to purify the protein (i) or cause its export out of the cell (ii).

**Choice A:** You are trying to purify the mutant reverse trancriptase and cannot obtain pure protein using size exclusion or ion exchange chromatography. How could you modify the expression vector to allow affinity chromatography?

**Choice B:** High levels of reverse transcriptase that are produced from the plasmid are toxic to the cell. How would you modify the expression vector to cause the reverse transcriptase to be exported out of the cell?

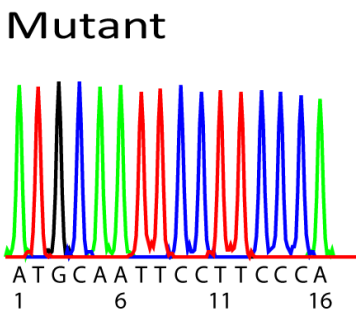
**Choice A:** Add codons for 6 His residues to the C-terminus of the protein. This will cause the protein to stick to Ni ions on chromatography resin beads.

**Choice B:** Add codons for the leader sequence to the amino terminus of the protein, this will cause the protein to be exported out of the cell by the export machinery (signal recognition particle binds to the leader, bringing the ribosome to the translocating machinery).



Points:_____

**25.** (5 pts) The wild-type and mutant reverse transcriptases are sequenced using the primer: ATGCTCCCCAC**.** Identify the amino acid change associated with the mutation. The codon table is provided on the last page.




```
                    Cys Asn Ser Ser Pro
Wild-type sequence: A TGC AAT TCC TCC CCA
Mutant sequence:    A TGC AAT TCC TTC CCA
                    Cys Asn Ser Phe Pro
```

The primer (underlined below) indicates that the sequencing should start with the highlighted "A". This "A" is the last base of a Thr codon, so the TGC is the first full codon and every three bases after that corresponds to a codon. **The change is from Ser to Phe.**

The sequence of the reverse transcriptase gene is provided here for your convenience (primer sequence underlined):

**AGCTGCTCATGCTCCCCACATGCAATTCCTCCCCA.....GTGAGGGGGAAATTA**ACCGCCGGCG
TCGACGAG**TACGAGGGGGTGTACCTTAAGGAGGGGT.....CACTCCCCCTTTAAT**TGGCGGCCGC
**MetLeuProThrCysAsnSerSerPro ValArgGlyLysStop**

**Bonus 3:** Briefly explain how the third "peak", corresponding to a G, was generated

This band is generated due to the presence of ddGTP in the reaction, which is a chain terminator, this band represents the addition of ddGTP as the third base to the growing strand.

**26.** (4 pts) You would like to make a human peptide hormone in yeast cell by using an expression plasmid that has the necessary signals for expression of proteins in eukaryotic cells, such as yeast. The peptide hormone will be used as a drug to treat individuals lacking this hormone. The sequence of the hormone is:

Met-Ala-Gly-Phe-Trp-Ala

The DNA that codes for this hormone is not available and thus you have to chemically synthesize the DNA instead of performing PCR. Assume that you are using the same expression vector as in Q20, which contains the start and stop codons, separated by the ~~EcoR1 and BamH1~~ **HindIII and Eco R1** sites. Give the DNA sequence that you would request from the DNA synthesis company. Codon usage frequencies for yeast are given on the last page. Write your answer below:

You need to add the restriction site for HindIII, followed by the codons for Ala-Gly-Phe-Trp-Ala, then the site for EcoR1. You want to select the codon that is found in the more highly expressed proteins (the typo in the question may have confused some students, keep this in mind when grading.)

```
AAGCTT GCG GGT TTC TGG GCT GAATTC
       Ala-Gly-Phe-Trp-Ala
```

**Bonus 4.** In what way is the ribosome like an apple? Similar in shape, stem represents the growing polypeptide coming out of the exit tunnel.

**Bonus 5.** Why is it necessary to produce T7 RNA polymerase when using the T7 expression system?

The T7 expression system replaced the normal promoter with the T7 promoter which can only be recognized by T7 RNA polymerase.



Points:_____

| 5' Base | Middle Base | | | | 3' |
|---|---|---|---|---|---|
| | **T** | **C** | **A** | **G** | |
| **T** | Phe | Ser | Tyr | Cys | **T** |
| | Phe | Ser | Tyr | Cys | **C** |
| | Leu | Ser | **Term** | **Term** | **A** |
| | Leu | Ser | **Term** | Trp | **G** |
| **C** | Leu | Pro | His | Arg | **T** |
| | Leu | Pro | His | Arg | **C** |
| | Leu | Pro | Gln | Arg | **A** |
| | Leu | Pro | Gln | Arg | **G** |
| **A** | Ile | Thr | Asn | Ser | **T** |
| | Ile | Thr | Asn | Ser | **C** |
| | Ile | Thr | Lys | Arg | **A** |
| | **Met** | Thr | Lys | Arg | **G** |
| **G** | Val | Ala | Asp | Gly | **T** |
| | Val | Ala | Asp | Gly | **C** |
| | Val | Ala | Glu | Gly | **A** |
| | Val | Ala | Glu | Gly | **G** |

Yeast Codon Usage - Middle column gives the percentage in highly expressed genes.

| | | | | |
|---|---|---|---|---|
| Ala | GCT | 14.52 | 27.54 | 22.86 |
| | GCC | 27.62 | 16.14 | 23.67 |
| | GCA | 19.63 | 24.01 | 31.27 |
| | GCG | 38.23 | 32.30 | 22.19 |

| | | | | |
|---|---|---|---|---|
| Gly | GGT | 32.91 | 50.84 | 31.79 |
| | GGC | 43.17 | 42.83 | 24.51 |
| | GGA | 9.19 | 1.97 | 24.75 |
| | GGG | 14.74 | 4.36 | 18.95 |

| | | | | |
|---|---|---|---|---|
| Phe | TTT | 55.09 | 29.08 | 67.14 |
| | TTC | 44.91 | 70.92 | 32.86 |



Points:_____