UNIVERSITY OF CALIFORNIA, BERKELEY
CHEM C130/MCB C100A
MIDTERM EXAMINATION #1
SEPTEMBER 19, 2012

INSTRUCTORS: John Kuriyan and David Wemmer

THE TIME LIMIT FOR THIS EXAMINATION: 1 HOUR 50 MINUTES

SIGNATURE:

_____

Please **SIGN** your name on the line above in **INDELIBLE INK**.

YOUR NAME:

_____

PLEASE PRINT your name (**IN INDELIBLE INK**) on the line above (& on the top right hand corner of every page).

PLEASE CIRCLE the name of your GSI:

Katherine N. Alfieri Varsha Desai

Michael Lawson Avi Samelson

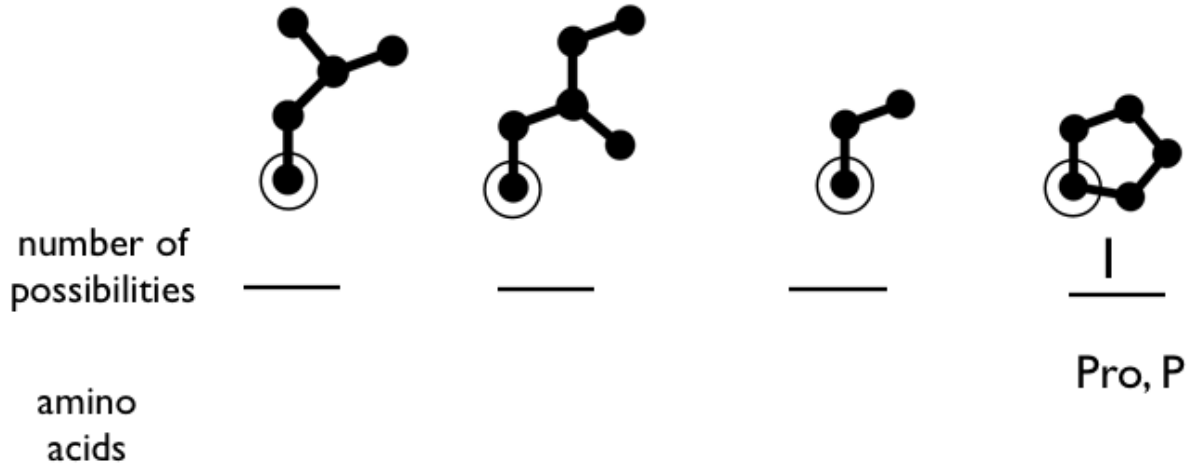
Tim Wendorff

PLEASE WRITE all of your answers **AS LEGIBLY AS POSSIBLE**.

SCORING. The exam consists of 5 questions totaling 100 pts as broken down in this table:

Question	Part A	Part B	Part C	Part D	Your Total	Max Score
1.	9	5	6			20
2.	4	6	10	----------		20
3.	8	12	----------	----------		20
4.	6	4	8	2		20
5	----------	----------	----------	----------		20
				TOTAL →		100

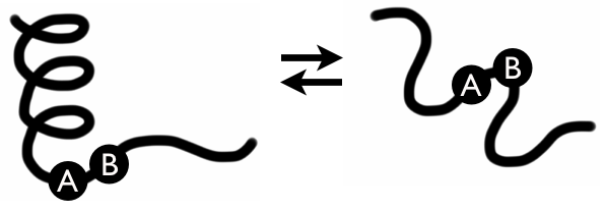
Q1. (20 points total)
Q1A (9 points) The seven schematic bonding diagrams below show amino acid Cα atoms and sidechains. The Cα atom of each amino acid is circled. No hydrogen atoms are shown, and carbons, oxygens and nitrogens are not distinguished.



(i) (3 points) For each bonding diagram, indicate the number of different amino acids that are consistent with it. For example, the last diagram is only consistent with proline, and so the corresponding entry is 1. If the amino acid is NOT a standard genetically encoded amino acid, write zero or 0.

(ii) (6 points) Write the single letter codes and the three letter codes for each of the amino acids that are consistent with each bonding diagram. For example, (Pro, P) is written below the last amino acid. If the amino acid is NOT a standard genetically encoded amino acid, write X.

Q1B. (5 points) A monomeric protein has only one histidine residue, at the position marked "A" in the schematic diagram below. The protein can exist in folded and unfolded (random coil) forms, as indicated in the diagram.



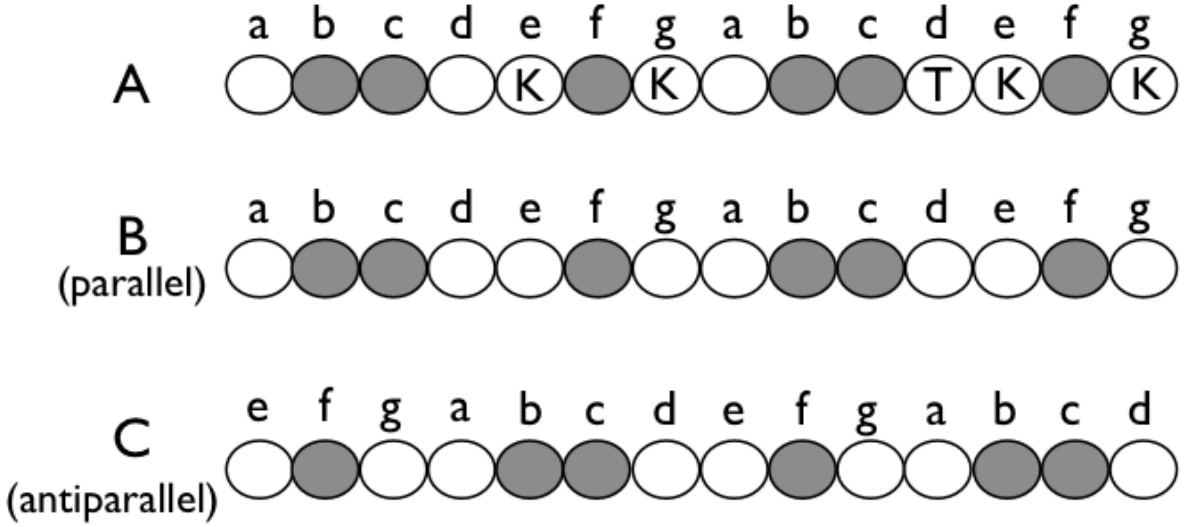
(i) (2 points) At neutral pH (7.0), the random coil form is observed to be dominant. Explain why the helical form is not stable.

(ii) (2 points) The pH of the solution is decreased to 5.5. At this pH, the protein is seen to be in the helical form. Based on this information, is the histidine likely to be located at the N-terminal end or the C-terminal end of the helix? Briefly justify your answer.

(iii) (1 point) A mutation is introduced in the protein at the position marked B. The sidechain at B can interact with the histidine sidechain. For the mutant protein, the helical form is *unstable* at pH 5.5. What is a likely choice for the new amino acid residue? Explain your reasoning.

Q1C (6 points)

The diagram below shows the sequences of three proteins, A, B and C. A forms a **parallel** coiled coil dimer with B. A also forms an **antiparallel** coiled coil dimer with C.
The residues in the white ovals in diagram are either L, T, E or K. In the diagram below, the a and d positions are at the interface with the partner coiled coil.

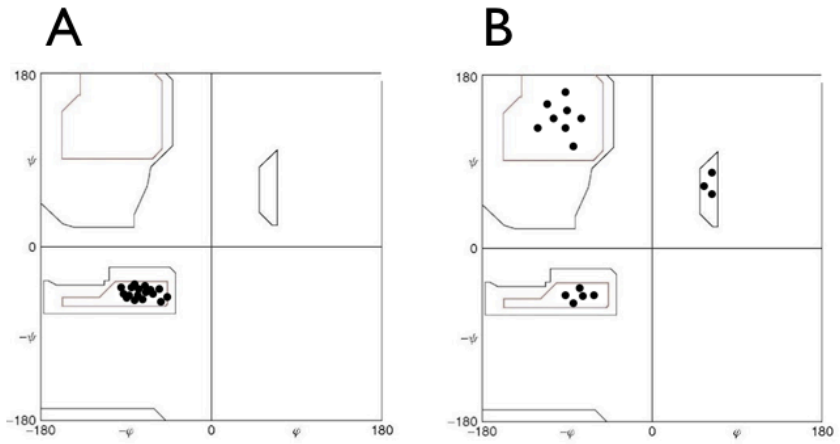


(i) For sequence A, fill in the white ovals with L, T, E or K, consistent with the ability of this protein to form a coiled coils with B and C.

(ii) For sequence B, fill in the white ovals, consistent with the ability of this protein to form a parallel coiled-coil with A.

(iii) For sequence C, fill in the white ovals, consistent with the ability of this protein to form an antiparallel coiled-coil with A.

Q2 (20 points total)
Q2A (4 points)
A small peptide containing ~15 residues undergoes a conformational change as it is transferred from water to octanol. Shown below are Ramachandran diagrams for the peptide in the two solvents.



(i) (2 points) Which Ramachandran diagram (A or B) corresponds to the peptide in octanol? Explain the reasoning behind your answer.

(ii) (2 points) As the peptide is transferred from the solvent corresponding to A to the solvent corresponding to B, the Φ and Ψ values for many residues have to change from values in the lower left hand quadrant to values in the upper quadrants. This would require the Φ and Ψ values to pass through regions of the diagram that are “forbidden” during the transition. Explain why the Ramachandran diagram fails to account for the possibility of such conformational transitions.

Q2B (6 points) Recall the *Nature* paper on RecA that was discussed in your section.

(i) (2 points) Summarize briefly what the RecA protein does.

(2 points) Describe briefly some of the important points that were learnt about the mechanism by studying the structure of the RecA-DNA complex.

(ii) (2 points) In this study, several RecA proteins were linked together by using artificial linkers containing 14 residues to connect the C-terminal end of one RecA molecule to another. The structure reveals that the distance between the last residue in one RecA molecule and the first residue in the next one is 28 Å. If the scientists had, instead, used a flexible linker containing 10 residues, would it have been long enough to have spanned this distance without strain? Explain your answer.

Q2C (10 points) Give a brief explanation of the following observations (i.e. explain briefly why these things occur, within the space provided):

(i) (2 points) Hoogsteen interactions in base pairs are seen more frequently in naturally occurring RNA structures than in naturally occurring DNA.

(ii) (2 points) DNA sequences that encode proteins are less conserved than the proteins that they encode.

(iii) (2 points) One chemical difference between DNA and RNA gives rise to most of the preference in the predominant conformation assumed by these.

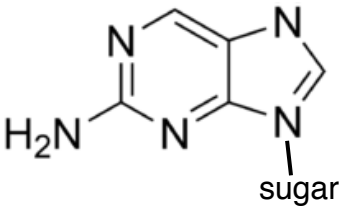
(iv) (2 points) Wobble base pairs have been shown to lead to DNA duplexes that are of very similar stability to those that contain only Watson Crick base pairs, however wobble pairs are essentially never found in natural DNA.

(v) (2 points) Give two significant conformational differences between B-DNA and Z-DNA.

#1:

#2:

Q3 (20 points total)
Q3A. (8 points)
(i) (4 points) The base 2-aminopurine can be chemically incorporated into DNA. It is useful because it is fluorescent while natural bases are not.



Under typical solution conditions (room temperature, moderate salt concentration and neutral pH) it can pair with a natural base in a double stranded DNA that yields a structure with the same stability as natural base pairs.

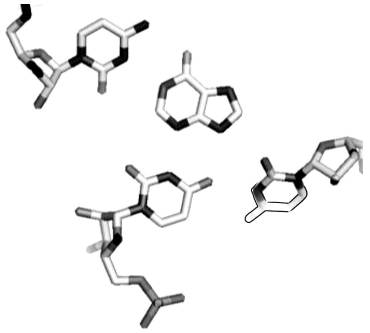
What natural base do you predict would be the best base pair?
Draw the structure of the base pair which you predict would be formed.

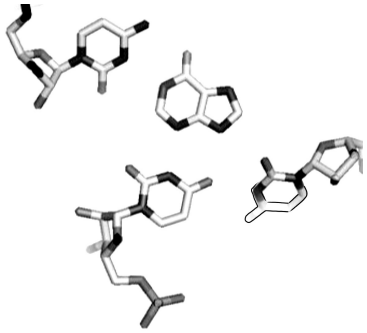
(ii) (2 points) At low pH it is found that a second, different pairing can occur with almost the same stability. What is the likely alternative base for pairing (explain your choice)?

(iii) (2 points). A sequence alignment can be done for DNA in the same way that it is done for proteins. Do you think that a 3D-1D profile comparison for DNA would work to recognize distantly related DNA sequences that same way that it does for protein sequences? (Explain briefly.)

Q3B (12 points) Some RNA molecules bind to specific compounds, for example nucleic acid bases, amino acids and vitamins. Below is part of the structure of an RNA bound to a **non-natural** aromatic base.

In this view three natural bases from the RNA can be seen as well as the non-natural base.





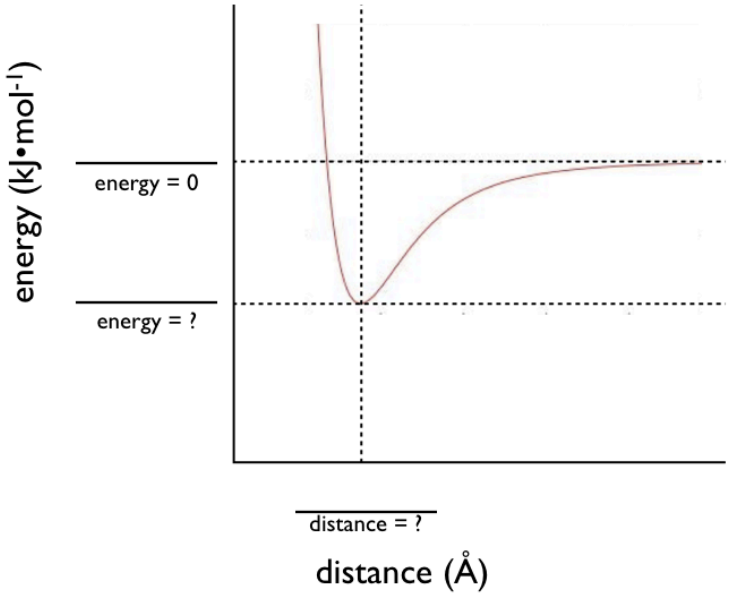
(i) (4 points) On the diagram on the **left**, draw in hydrogens on both the non-natural base and the surrounding RNA bases.

(ii) (3 points) Identify the three natural bases by putting the one letter code for the base next to the ring, in the diagram on the left.

(iii) (3 points) The RNA molecule also binds to the guanine base. In the diagram on the **right**, add (draw in) appropriate atoms to the non-natural base to show what guanine probably looks like bound to this RNA. Draw all the hydrogen bonds between the guanine and the other bases.

(iv) (2 points) Do you expect the non-natural base or guanine to bind more strongly to this RNA?

Q4 (20 points total)
Q4A (6 points) The diagram below shows a graph of the van der Waals energy for two carbon atoms that interact noncovalently.



(i) (2 points) In the space provided, fill in the energy value and the distance corresponding to the dotted lines in the graph.

(ii) (2 points) Draw two horizontal lines corresponding to 0 +/- "thermal energy" (*RT*) at 10 K (ten degrees Kelvin). To calculate the value of the thermal energy, use the fact that $RT \simeq 2.5$ kJ•mol^{-1} at 300 K.

(iii) (2 points) Based on your diagram, do you expect random collisions to disrupt noncovalent interactions between carbon atoms at 10 K? Explain your answer.

Q4B (4 points) A protein contains six cysteine residues. The active and correctly folded form of the protein contains only two disulfide bonds (that is, two cysteines are unpaired in the correct structure). An Anfinsen-type experiment is carried out as follows:

The correctly folded protein is unfolded in urea and reducing agent, and is then allowed to refold by simultaneously removing urea and adding oxygen. 100% of the initial activity is recovered in this way.

(i) (2 points) The protein is unfolded as before, but now oxygen is added in the presence of urea. After some time, the urea is removed. Assuming that all cysteine residues can form disulfide bonds equally well with all other cysteines, what is the expected recovery of activity?

(ii) (2 points) In a modified form of the experiment, we begin by adding an alkylating agent to the correctly folded and fully active protein. The alkylating agent reacts with any unpaired cysteines and permanently blocks the sulfhydryl group from further reaction. The alkylating agent is then removed and the experiment is carried forward as described in (i). What is the expected recovery of activity?

Q4C (6 points) Shown below is a portion of the BLOSUM amino acid substitution matrix.

	C	K	V	I	W
C	9				
K	-3	5			
V	-1	-2	4		
I	-1	-3	3	4	
W	-2	-3	-3	-3	11

(i) (2 points) Explain why the diagonal scores for C and W are higher than for the other residues.

(ii) (4 points)Two sequence alignment shown below:

Alignment 1

Sequence A	**S**	**R**	**I**	**V**
Sequence B	**S**	**R**	**I**	**I**

Alignment 2

Sequence C	**S**	**R**	**I**	**C**
Sequence D	**S**	**R**	**I**	**K**

Based on the definition of the Blosum score, calculate the relative likelihood that the A-B pair corresponds to evolutionarily related proteins versus the C-D pair. That is, calculate the following likelihood ratio:

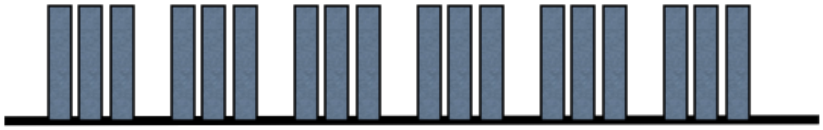
$$\frac{\text{likelihood that the A-B alignment is of related proteins}}{\text{likelihood that the C-D alignment is of related proteins}}$$

Q4C (iii) (2 points) Shown below is an aligned sequence block from a set of related proteins. Only K, R and N residues are shown. The blank squares represent other amino acid residues.

K		R				K			K		
		K		K		N			R		R
K				K					N		
		K							R		
K		N				R			N		

Based on this aligned block, which substitution will have a larger BLOSUM score, the K-N substitution or the R-N substitution? Answer the question, and explain your answer, without doing any calculations.

Q4D. (2 points) Returning to the *Nature* paper, recall that single-stranded DNA is stretched out when it binds to the RecA protein. But the DNA is not stretched out uniformly. Instead, groups of 3 nucleotides retain B-form conformation, while the phosphate backbone between them is extended, as shown schematically below:



The protein uses less energy to distort the DNA in this way than would be required to uniformly stretch DNA to the same final length. What is the principal energetic contribution that resists a more uniform distortion of the DNA? Explain the origin of this energetic contribution.

Q5. Multiple choice and True/False questions. Circle the correct option (or circle either TRUE or FALSE). **+2 points for each correct answer, –2 points for each wrong answer.** To get the maximum score you do not need to answer all the questions, so be careful not to answer questions incorrectly. Unanswered questions do not change the score. **Maximum points: 20. Minimum points: 0.**

(i) Which of the following statements is **least applicable** to the description of hydrogen bonding in protein folding (circle the worst option):

- the hydrogen bonding interaction causes the formation of α helices and β sheets.

- hydrogen bonding ensures the specificity of the folded structure.

- the formation of regular hydrogen bonds in α helices and β sheets compensates for the loss of interactions with water.

- By maintaining the donor and acceptor atoms at distances of about 4 Å, the formation of hydrogen bonds prevents nonspecific collapse of the protein chain.

(ii) Transcription factors are proteins that recognize specific sequences in DNA. They do so by interacting with the major groove. Of the following statements, circle the one that best explains this phenomenon:

- Insertion of α helices into the minor groove causes van der Waals collisions.

- different base pairs do not provide a unique set of interactions in the minor groove.

- α helices have a structure that is more consistent with the major groove.

- transcription factors interact with the phosphate backbone of DNA, and these are more readily accessible in the major groove.

(iii) DNA is taken from an organism that has been shown to have 60% G-C base pairs and 40% A-T base pairs.

If bases occurred randomly in sequence what would be the probability of finding the specific sequence of bases TAATATA within a specific block of seven base pairs? Circle the best answer:

- 0.6^7
- 0.4^7
- 0.3^7
- 0.2^7

(iv) In bacterial DNA it is found that this sequence (TAATATA) and others with very few differences are found much more frequently in segments of DNA that do not encode proteins than expected by chance, but are located immediately before the start of a gene. Consider the following reasons for why this sequence is over-represented in the bacterial DNA. Circle the statement that is **least likely** to be true.
- Because these regions do not encode proteins, the interactions between the two DNA strands does not need to be strong.

- In order for RNA polymerase to transcribe the gene, it has to open double helical DNA, and this sequence is easier to open than those containing C-G pairs.
- Transcription factors that recruit RNA polymerase bind to these sequences.
- The regions upstream of genes are control elements, and conservation of DNA sequence in these regions reflects the fact that a common machinery is used to transcribe all genes.

(v) Which of the following sequences is most likely to form Z-DNA?

- ATCGATCGATCG -CCCGGGCCCGGG

- CGCGCGCGCGC -CACACACACACAC

(vi) A membrane protein contains a channel within which water molecules form a hydrogen-bonded array from one side of the membrane to the other. The protein is embedded in a membrane, and both sides of the membrane are at the same pH. The protein will allow net proton flow from one side to the other.

TRUE / FALSE

(vii) A helical segment in an protein contains 15 consecutive hydrophobic residues, flanked by polar residues at both ends. This helix can, on its own, span the membrane in a cell.

TRUE / FALSE

(viii) Loops connecting two secondary structural elements are usually at the surface of a protein rather than in the interior because (circle the best answer):

- Active sites, where loops play important roles, are usually at the surface of the protein, where substrate molecules can bind.
- Loops usually have backbone groups that do not form hydrogen bonds with the rest of the protein, and these elements need to be exposed to water.

(ix) The BLOSUM score for substituting Aspartate by Asparagine is +1. Thus, this is a conservative substitution, and is favored in evolution over random chance. Replacement of an internal aspartic acid residue in bacteriorhodopsin by asparagine is therefore expected to maintain the function of the membrane protein.

TRUE / FALSE

(x) Different globin proteins that bind oxygen have related folds, despite having very low sequence identity. Circle the answer that BEST describes the reason for this.

- There is only one kind of fold that allows a protein to bind to oxygen, and so the fold is conserved.
- Only one kind of protein fold can wrap around the large heme group, and so the fold is conserved.
- The fold is conserved because all these proteins arose from a common ancestor.
- There are very limited ways in which alpha-helices can pack against each other, mutations that disturbed the original packing were lost to evolution because they destabilized the protein.

(xi) The Anfinsen experiment is conceptually very important in structural biology because: (circle the best answer)
- it shows that digestive enzymes such as ribonuclease can be repeatedly folded and unfolded.
- it demonstrates the importance of disulfide bonds for protein stability.
- it shows the importance of protein chaperones, since most proteins cannot easily be folded in a test tube.
- it was the first demonstration that all the information required for a protein to fold is contained in its sequence.

(xii) In principle, you can imagine constructing a membrane protein with hydrophobic sidechains on the outside and hydrophilic (polar) sidechains on the inside of the protein. But most membrane proteins (except for some channels) have hydrophobic sidechains on the interior as well. Circle the best explanation for this.
- It is difficult to get hydrophilic groups into the membrane.
- Hydrophilic groups in the interior have to be paired with other hydrophilic groups in very specific arrangements, and this would make the protein very susceptible to inactivation by mutation.
- Hydrophilic groups are more flexible, and they would not pack properly to form a rigid interior.
- Hydrophilic groups usually have water molecules bound to them, which are difficult to remove.

(xiii) The presence of a transmembrane helix in a protein sequence is detected by graphing the value of the hydrophobic index, summed or averaged over several adjacent residues, centered on each residues in turn. Over how many residues is the mean or aggregate value calculated? Choose the best answer:

(a) 10 (b) 20 (c) 30 (d) 40

(xiv) Which of the following sequence motifs is likely to be found in an amphipathic β strand?

NQSLASVQEKL ILVAFSDEKK LKVRFSIQV

(xv) A chemist synthesizes a protein out of D-amino acids. What kinds of secondary structural elements are likely to be found in such proteins? Circle the best answer.

(a) right-handed α helices (b) left-handed α helices
(c) no α helices (d) both right-handed and left-handed α helices