UNIVERSITY OF CALIFORNIA, BERKELEY
CHEM C130/MCB C100A
MIDTERM EXAMINATION #1
SEPTEMBER 24, 2014

INSTRUCTORS: John Kuriyan and David Savage

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:

Pradeep Bandaru Caroline Cypranowska

Julian Hassinger Madeleine Jensen

Laura Nocka

PLEASE WRITE all of your answers **AS LEGIBLY AS POSSIBLE**.
Note that any exam submitted for a regrade should be written in indelible ink.
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.	10	6	4	----------		20
2.	4	8	2	6		20
3.	5	6	9	----------		20
4.	11	5	4	----------		20
5	----------	----------	----------	----------		20
				TOTAL		100

Q1. (20 points)
(A) (10 points) Atoms of argon are non-polar, and the energetics of noncovalent interaction between argon atoms is comparable to that of carbon atoms. The van der Waals radius of Argon is 1.8 Å.

In two separate experiments, a scientist introduces argon atoms into liquid methane and into liquid water, and makes measurements at room temperature. She monitors whether the argon atoms are moving independently of each other, or forming clumps (that is, clusters of argon atoms packed close together).

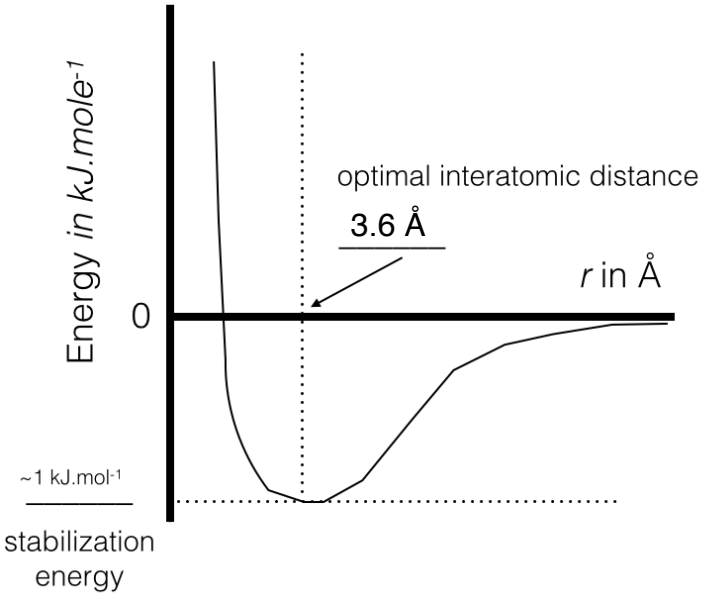
(i) (2 points) Do you expect the argon atoms to form clumps in methane? Circle the best choice and justify your answer. YES / NO
Answer: No. Since argon is non-polar, and methane is also non-polar, the argon atoms would not form stable interactions with each other at room temperature.

(ii) (2 points) Do you expect the argon atoms to form clumps in water? Circle the best choice and justify your answer. YES / NO
Yes. Because argon is non-polar, it will clump together through the hydrophobic effect in water.

(iii) (2 points) The scientist prepares a sample of pure argon and cools it to 20 K. She monitors the sample for clumping of Argon atoms. Do you expect her to find clumps of argon? Circle the best choice and justify your answer. YES / NO

At 20K, the value of RT is 0.17 $kJ.mol^{-1}$. Thus, at 20K the stabilization energy of the van der Waals interactions between argon atoms will be lower than RT and the atoms will tend to stick together.

Q1A, continued.
(iv) (4 points) On the graph below, sketch the energy as a function of distance for two argon atoms interacting with each other. The stabilization energy is indicated by a horizontal dotted line, and the optimal interatomic distance by a vertical dotted line. Fill in the values of the optimal distance and the stabilization energy in the blank lines provided.



Q1B (6 points)

A small protein consists of 37 residues. The sequence of the protein is as follows:

```
....|....|....|....|....|....|....|..
xRxxExxxKxxDxxxRxxExxxKxxDxxxRxxExxxR
```

In this diagram, x represents any neutral polar or hydrophobic residue. Every 5th residue in the sequence is indicated by a vertical line above the sequence.

When dissolved in water at neutral pH, the peptide forms a stable and monomeric α helix.

(i) (2 points) How many turns does the α helix have?

Since there are 3.6 residues per turn in an alpha helix, this helix will have ~10 turns.

(ii) (2 points) A mutation is introduced into the protein so that the lysine at position 9 is replaced by glutamate. The mutant protein no longer forms an α helix in water. Explain why.

The i, i+3, i+4 pattern of oppositely charged residues indicate that there are ion pairs between these residues. Changing Lys 9 to Glu would create three negative charges in a row, which could destabilize the helix through electrostatic repulsion.

(iii) (2 points) The mutant protein (with glutamate at position 9) is transferred to liquid hexane (C_6H_{14}). It now forms a very stable α helix. Explain why.

In liquid hexane, a nonpolar solvent, the backbone groups can no longer form hydrogen bonds with water. Hence, the unfolded form of the helix will be very unstable. This has the effect of stabilizing the helical form, because then all the backbone amide and carbonyl groups can form hydrogen bonds (except at the very end). This will compensate for the unfavorable glutamate residue.

Q1C. (4 points) A protein of molecular weight 30 kDa is roughly spherical in shape, and has a diameter of 30 Å.

(i) (2 points) A β strand spans the diameter of the protein (that is, it goes from one point on the surface of the protein to a point diametrically opposite). How many residues would you expect to be in this β strand? Justify your answer.

The average Cα - Cα distance in an extended β strand is ~3.5 Å. So, to span 30 Å, the number of residues required will be: 30/3.5 = 8.6 or roughly 9 to 10 residues.

(ii) (2 points) Would you expect this β strand to be amphipathic? Justify your answer.

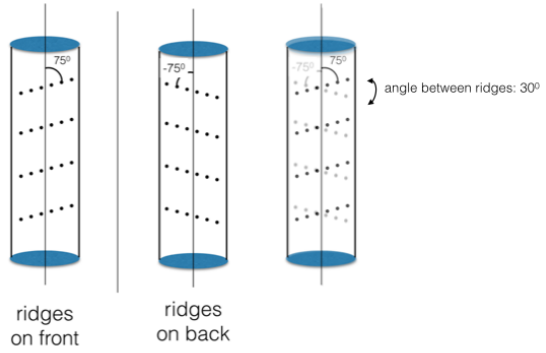
No, this strand would not be amphipathic, since it would traverse the interior of the protein. The sidechains would have a complex pattern of hydrophobic/hydrophilic and not the simple alternating pattern expected for an amphipathic strand.

Q2. (20 points)
(A) (4 points) Chemists can synthesize artificial proteins in which the amino acids are modified so that the backbone geometries are slightly different from that of natural amino acids.

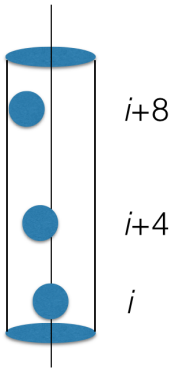
(i) (2 points) Using a set of modified amino acids, a chemist makes an α helix so that the ridges are arranged at an angle of 75^0 to the vertical axis. The diagram below shows two views of the helix, one showing the ridges on the front and one the ridges on the back. When two such helices pack against each other using these ridges and the grooves between them, what is the angle between the helices? Circle the best answer:

(a) 75^0 (b) 15^0 (c) **$\underline{30^0}$** (d) 90^0 (e) 60^0

Explain your choice by drawing a diagram illustrating the logic behind your answer.



(ii) (2 points) Using another set of modified amino acids, the chemist makes a right-handed α helix that has 4.1 residues per turn. By placing leucine residues at appropriate positions within the helix, he finds that he can create helices that form dimeric coiled coils. Would these coiled coils be right-handed or left-handed supercoils? How many residues would be in the repeating motif of such a coiled-coil? Explain your reasoning.



The coiled-coil is likely to be left-handed. There would be 8 residues per repeat (4 would also be a reasonable answer). As shown in the diagram below, the residues at *i*+4 and *i*+8 are slightly to the left of the vertical line running through the i^{th} residue. So, a left handed deformation of a helix interacting with this one (not shown). would bring the residues into alignment.

Q2 (B) (8 points) The diagram below shows a part of a protein, with seven amino acid residues in this region of the protein. Only the Cα atoms are shown for the peptide backbone, and no hydrogen atoms or double bonds are shown. Heteroatoms (e.g., nitrogens or oxygens) are not highlighted.

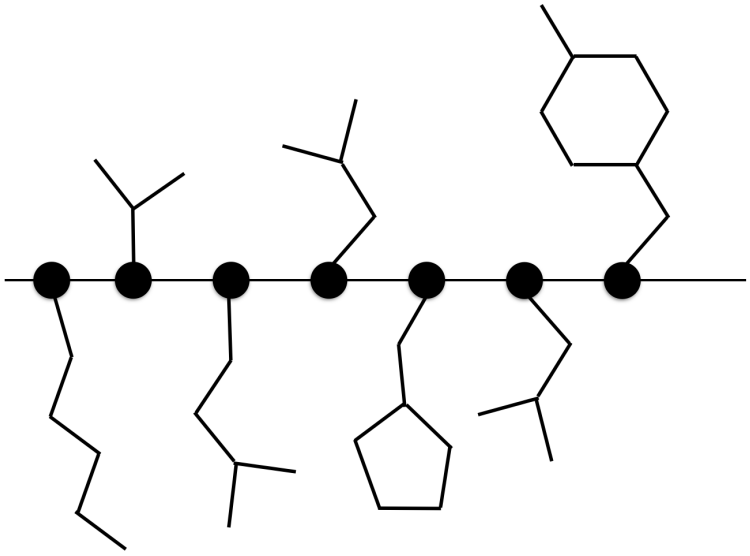
- This region forms an amphipathic β strand that packs against the rest of the protein.
- One side of the amphipathic strand contains 3 ion pairs.
- The other side of the strand only has neutral residues.

(i) (7 points) Identify each of the amino acid residues using both the three letter and one letter codes, on the diagram below.

(ii) (1 point) Given that there are 3 ion pairs on one side of the strand, what is the likely pH range under which the measurements are made? (Circle the best answer, and explain your answer)

(a) less than 5.0 (b) 5.0 to 7.5 (c) greater than 7.5

A _____ B _____ C _____



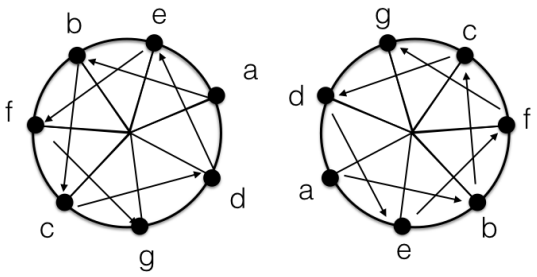
D _____ E _____ F _____ G _____

Answer: A = Val, V ; B = Leu, L ; C = Tyr, Y ; D = Lys, K ; E = Glu, E ; F = His (His^+), H ; G = Asp, D

To be sure of getting 3 ion pairs, the pH should be 5.0, where the His is more likely to be charged.

Choice (a) was accepted if it was justified with the correct reasoning. For example, the pH could not be too much lower than 5 otherwise we would start protonating the acidic bases.

Q2C. (2 points) In a coiled-coil structure, the deformation of the α helix causes the helices to have exactly 3.5 residues per turn, leading to every 7th residue being exactly aligned on the helical axis. The diagram below shows two helices that form a parallel coiled-coil. The N-termini are below the plane of the paper, and the C-termini are above. The residues in the heptad repeat are identified by the letters *a*, *b*, *c*, *d*, *e*, *f* and *g*. The *a* residues are identified on both helices. Label the other residues on each helix, and draw an arrow connecting consecutive residues (that is, draw an arrow connecting a to b, etc.).



Q2D (6 points) A mixture of 4 proteins (A,B,C and D) are present in solution. 3 of these proteins are known to bind to nucleotides (e.g. ATP) while the third one does not. The properties of the proteins are summarized in the table below:

Protein	pI (isoelectric point)	Molecular weight	Nucleotide binding?
A	7.4	90,000	Yes
B	3.2	16,000	Yes
C	7.8	22,000	No
D	7.9	21,000	Yes

(i) (2 points) The mixture of proteins is passed over an ion exchange column, leading to the separation of one protein from the others. Which one is it, and why?

Protein B has a pI value that is well separated from the others, so it will separate from the others on an ion exchange column.

(ii) (2 points) The mixture of proteins is passed over a gel filtration column. Explain which protein elutes first from the column, and why.

Protein A has the largest molecular weight and it will elute first.

(iii) (2 points) A scientist wishes to prepare a sample containing only proteins B and D. Describe a two column procedure that will result in a separation of a mixture of these two proteins from A and C.

Gel filtration column, remove protein A. Then affinity purification, using a column in which the resin is bound to ATP or an ATP analog, which will remove protein C.

Q3 (20 points)

Q3A (5 points)
In the analysis of ribonuclease refolding carried out by Anfinsen, the following two experiments were done:

Experiment 1:

folded protein $\xrightarrow{\text{add urea, reducing agent}}$ unfolded protein $\xrightarrow{\text{remove urea, add oxidizing agent}}$ measure activity.

Experiment 2:

folded protein $\xrightarrow{\text{add urea, reducing agent}}$ unfolded protein $\xrightarrow[\text{add oxygen}]{\text{remove reducing agent}}$ wait $\xrightarrow{\text{remove urea}}$ measure activity.

A variant of the Anfinsen analysis is carried out using a protein that has 7 cysteine residues (instead of the 8 in ribonuclease). In the correctly folded protein, there are 3 disulfide bonds and one unpaired cysteine residue. When Experiment 1 is carried out, 100% of the enzyme activity is recovered after the the unfolding and folding steps. Assuming that the protein is completely unfolded by urea, what percentage of the activity do you expect to observe at the end of the second experiment? Assume that the same amounts of protein are used in the two experiments. Show all the steps of how you work out your answer.

Answer: The number of pairings of 7 residues is 7! = 5040

The ordering of the pairs is not important, so we have to divide by 3!

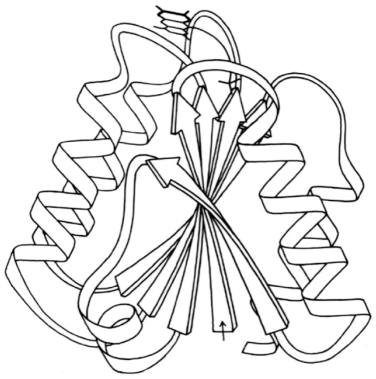
5040/6 = 840

Within each pair, the order does not matter, so we have to divide by $2^3 = 8$

840/8 = 105

Only of these combinations corresponds to the correct set of disulfide pairing, so the expected recovery of activity will be 1/105 = ~1%

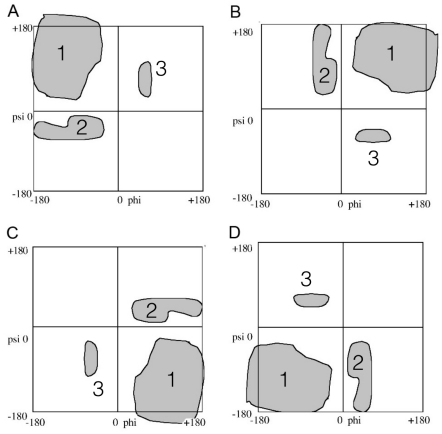
Q3B (6 points) A spacecraft exploring mars sends back samples to earth in which proteins are detected. People are concerned about whether the proteins are truly of extraterrestrial origin, or simply contaminants from earth. The structure of one of the proteins is determined and is shown below:



(i) (2 points) Based on this diagram, is this protein likely to be of extraterrestrial origin or from earth? Explain your answer.
All proteins of earth are made of L-amino acids, and these have right-handed α helices (the only exceptions are peptides in certain bacterial cell walls, which are not folded into compact structures). Since the α helices in this protein are left-handed, we conclude the protein is not from earth.

(ii) (2 points) Shown below are four Ramachandran diagrams. Which one corresponds to the protein shown above? A B **C** or D (circle the correct answer). Briefly explain your choice.

Answer: Proteins made of D-amino acids are mirror images of proteins made with L-amino acids. Since everything is inverted, the rotational angles of the backbone are also inverted, and since C is an inversion of the normal Ramachandran diagram that is the one likely to represent the alien protein.



(iii) (2 points) For natural proteins on earth, what types of secondary structure correspond to the three marked regions in the Ramachandran diagram?
Region 1: __β strand_____ Region 2: ___α helix_____ Region 3: __left-handed helix

Q3C (9 points)
Shown below is a selected portion of the BLOSUM-62 matrix.

	A	N	L	F	W	D	E
A	4	-2	-1	-2	-3	-2	-1
N		6	-3	-3	-4	1	0
L			4	0	-2	-4	-3
F				6	1	-3	-3
W					11	-4	-3
D						6	2
E							5

The entries in this matrix are the BLOSUM score, $S_{i,j}$, where:

$$S_{i,j} = 2 \times \log_2 L_{i,j}$$

(i) (4 points) In the BLOSUM-62 matrix shown above, $S_{LL} = 4$ and $S_{WW} = 11$. The BLOSUM-62 matrix is calculated by excluding from the sequence block multiple instances of sequences that are more than 62% identical to each other. Only one instance of such sequences are included. For example, mammalian proteins are usually nearly identical to each other in sequence, and only one example would be included in the BLOSUM-62 matrix.

Consider a BLOSUM matrix calculated by include *all* examples of related proteins, without this filter. Which of the following statements best describes the values of S_{LL} and S_{WW} for the unfiltered matrix? Circle the best option and briefly explain your choice.

(a) They are the same
(b) Their values would be closer together
(c) Their values would be more different from each other.

Answer: (b) - the values would be closer together. Inclusion of very similar proteins, e.g., all mammalian variants of a protein, in the sequence blocks would skew the statistics. If many proteins that are nearly identical are included, it would look as if many positions in the sequence are highly conserved. Thus, the self-substitution score for leucine would become larger, and tryptophan would not stand out so much as a residue that is rarely substituted.

Q3C, continued
(ii) (5 points) Consider the following sequence alignment:

```
Protein A:     N L E A D F W N L
Protein B:     N L E A D F D N L
Protein C:     D L E A D F W N L
```

Calculate the ratio of the likelihoods that proteins A and B are evolutionarily related versus proteins A and C are evolutionarily related. That is, calculate the following ratio:

$$\frac{\left[\dfrac{\text{likelihood that sequences A and B evolved from a common ancestor}}{\text{sequences A and B are randomly related}}\right]}{\left[\dfrac{\text{likelihood that sequences A and C evolved from a common ancestor}}{\text{sequences A and C are randomly related}}\right]}$$

Answer: The ratio of likelihoods is related to the difference between the Blosum scores:

$$L_{AB} = 2^{S_{AB}/2} \quad L_{AC} = 2^{S_{AC}/2} \text{ and so: } \frac{L_{AB}}{L_{AC}} = 2^{(S_{AB}/2)-(S_{AC}/2)} = 2^{\Delta S/2}$$

Because we are considering the difference of BLOSUM scores, we need to only look at positions that are different in one or the other sequence:

$$\Delta S = S_{AB} - S_{AC} = [S_{NN} + S_{WD}] - [S_{ND} + S_{WW}]$$
$$= [6-4] - [1+11] = \text{-10}$$

$$\frac{L_{AB}}{L_{AC}} = 2^{\Delta S/2} = 2^{\text{-10/2}} = 2^{\text{-5}} = 0.03125$$

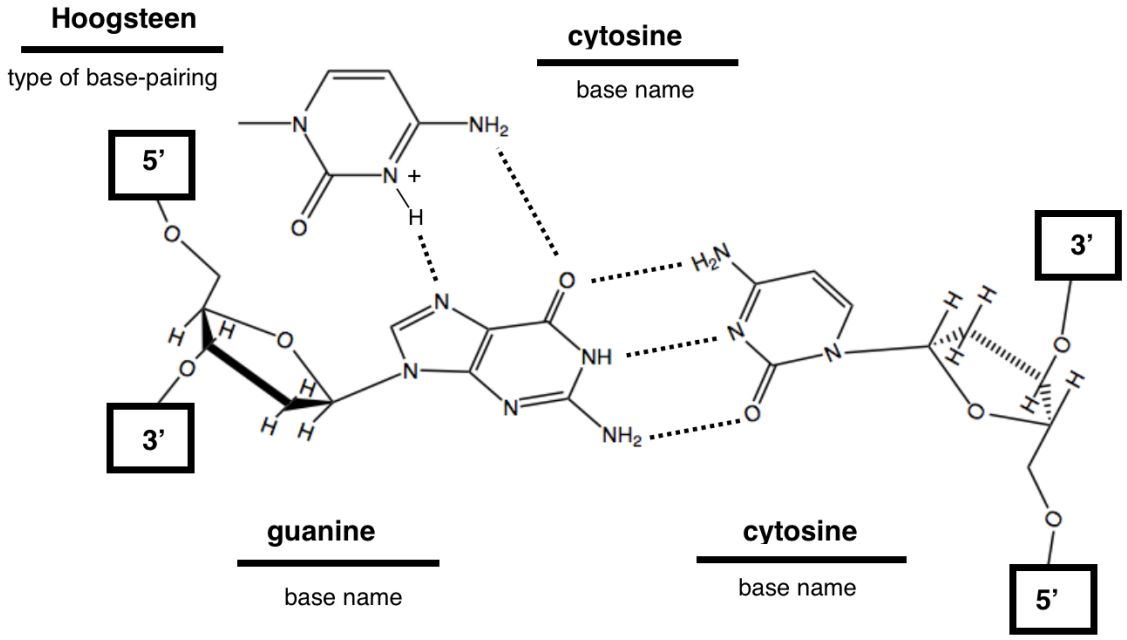
Q4. (20 points)
Q4A (11 points) Shown below is a single base-pair of DNA sequence upstream of the gene Myc associated with human cancer. Interestingly, if one had a powerful enough microscope to observe this base pair at atomic resolution in living cells, one would occasionally find a third base making a significant interaction as also shown below. Please complete the structure and answer the following questions.

(i) (2 points) Label the 3' and 5' ends of the antiparallel strands, in the boxes provided.

(ii) (5 points) For the Watson-Crick pair: fill in the missing chemical functionalities, label the names of both bases, and complete the hydrogen bonding pattern.

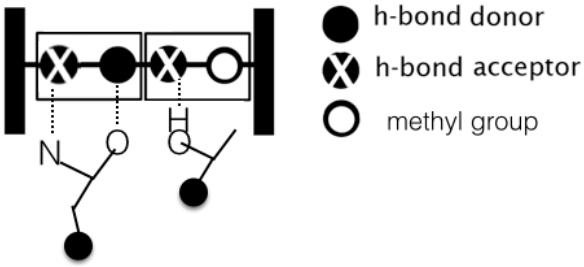
(iii) (3 points) What is the name of the interaction the third base is making? Fill in the missing chemical functionalities and give the name of the third base.

(iv) (1 point) Given your knowledge of the functionalities of these macromolecules, what process might this interaction be regulating (3 points)?
Answer:



Hoogsteen base pairing is often used to regulate transcription.

Q4B. (5 points)
(i) (2 points) The following schematic diagram indicates the edge of an A-T base pair that faces the major groove. Indicate which is A and which is T.
Answer: The base on the left is A. The base on the right is T.



(ii) (2 points) A protein inserts into the major groove of DNA and interacts with this base pair using an asparagine residue and a threonine residue. Draw in the chemical structures of the two sidechains in the diagram above, using the two black circles as the Cα atoms. Indicate the hydrogen bonds by dotted lines.

(iii) (1 point) Sequence-specific DNA-binding proteins usually recognize DNA in the major groove rather than the minor groove. One reason is that α helices cannot readily enter the narrower minor groove. What is the other important reason that the recognition occurs in the major groove?
Answer: Interaction with the major groove edges of the base pairs is more specific because the pattern of interaction sites is unique to each of the 4 base pairs (AT, CG, TA, CG).

Q4C (4 points) A circular piece of DNA is cut, and one end is rotated with respect to the other. The ends are then resealed. The resulting DNA molecule is shown in the diagram below.

(i) (2 points) How many turns was one end of the DNA rotated before the ends were resealed?
4

(ii) (2 points) Was the rotation done in a right-handed or a left-handed sense?

The DNA is negatively supercoiled, so the rotation was done in a left-handed sense (the DNA is underwound).

Q5. (20 points) Multiple choice and True/False questions. Circle the best option (or circle either TRUE or FALSE). **+2 points for each correct answer, –1 point 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. If the total score is zero or negative, the total score will be set to zero.

(i) The BLOSUM substitution score for Cys-Cys substitutions is higher than for Leu-Leu substitutions because:
(a) Cysteine is a rarer amino acid than leucine and so it is more highly conserved.
(b) Leucine can be replaced with more residues than cysteine, without loss of function.
(c) Leucine is a more common residue in proteins than cysteine, so it is less important.

(ii) Across evolution, the globin fold is relatively unchanged even though the sequences of different globins are very divergent because:
(a) Heme groups only bind to the globin fold, and heme binding is necessary for oxygen binding.
(b) The globin fold is required for the formation of quaternary structure.
(c) The globin fold contains α helices, and the stable packing of α helices against each other is restricted to specific inter-helical angles.
(d) The globin fold is required to form the hydrophobic core.

(iii) Choose the amino acid substitution that results in the greatest change of hydrophobicity:
(a) A $\longrightarrow$ T
(b) W $\longrightarrow$ M
(c) P $\longrightarrow$ S
(d) **F** $\longrightarrow$ **R**

(iv) Membrane proteins occasionally contain polar sidechains, and these residues are usually more conserved in evolution than other residues because:
(a) Their presence is required for hydrogen bonding interactions with the phosphate headgroups of the lipids.
(b) They are likely to play a critical role in the biological function of the protein.
(b) Polar sidechains have a unique structure that cannot readily be substituted.
(c) The interior of the lipid bilayer occasionally has polar groups that interact with the protein.

(v) You align the sequences of two proteins, and find that they share 20% sequence identity. You can be reasonably sure that the two proteins have the same fold.
TRUE / **FALSE**

(vi) Active sites in proteins are often located at the interfaces between domains in the folded structures of proteins because:
  (a) The structures and sequences of individual domains have to satisfy the constraints of folding, while interdomain regions can evolve rapidly.
  (b) Interdomain orientations can change easily, allowing evolution to accommodate different ligands at the active site.
  (c) The interdomain regions can provide crevices for the binding of small molecules.
  (d) All of the above are true.

(vii) The stability of a folded protein is determined to a large extent by the formation of hydrogen bonds between backbone amide and carbonyl groups.
  TRUE / **FALSE**

(viii) The B-form double helix is a characteristic feature of DNA but not RNA because:
  (a) The extra oxygen atom in RNA collides with a base.
  (b) The lack of a methyl group in uracil versus thymidine reduces the hydrophobic stabilization of RNA.
  (c) Protein molecules do not have to interact with RNA in the major groove, so the B-form is not needed.
  (d) The C2' endo conformation of the sugar is disfavored in RNA.
  (e) The C3' exo conformation of the sugar is disfavored in RNA.

(ix) The three-dimensional structures of folded RNA molecules contain non-Watson-Crick base pairs such as G-U, as well as many alternative interactions between bases because:
  (a) the constraint of uniform base pair geometry does not apply in RNA as it does in DNA.
  (b) the presence of the 2' OH group in RNA allows more variation in structure
  (c) metal ions interact differently with RNA than DNA
  (d) RNA is less structurally stable than DNA

(x) The following interactions are arranged in *increasing* order of strength (TRUE/ **FALSE**):
  Van der Waals attraction / hydrogen bonds / ion pairs / peptide bonds / base stacking interactions

(xi) Single-stranded DNA molecules containing only adenine bases (poly-A) form helical structures in water. The formation of the single-stranded helix is due to:
  (a) minimization of the exposed surface of the bases
  (b) favorable electrostatic interactions between the partial charges on the bases
  (c) shape complementarity of the bases
  (d) minimization of phosphate-phosphate repulsion

(xii) The Ramachandran diagram is calculated for an alanine-alanine dipeptide. Nevertheless, the forbidden regions in the diagram apply to all pairs of amino acids except proline and glycine because:

(a) alanine is representative of all the other amino acids
(b) the forbidden regions are characteristic of the protein backbone atoms only
(c) additional interactions in the amino acids with larger sidechains cannot overcome the repulsions present in the alanine-alanine dipeptide
(d) all residues have to be compatible with α helices and β sheets

(xiii) The size of a typical protein domain is:

(a) 50 residues
(b) **150 residues**
(c) 300 residues
(d) 500 residues