UNIVERSITY OF CALIFORNIA, BERKELEY
CHEM C130/MCB C100A MIDTERM EXAMINATION #1. SEPTEMBER 26, 2006
INSTRUCTOR: John Kuriyan

THE TIME LIMIT FOR THIS EXAMINATION IS 1 HOUR AND 20 MINUTES

SIGNATURE:

_____

Please **SIGN** your name (**in indelible ink**) on the line above.

YOUR NAME:

_____

Please **PRINT** your name (**in indelible ink**) on the line above (& on the top right hand corner of every page). Also, please write all of your answers **as legibly as possible**.

PLEASE **CIRCLE** THE NAME OF THE GSI FROM WHOM YOU WILL PICK UP YOUR GRADED MID-TERM EXAM:

Natasha Keith Kyle Simonetta

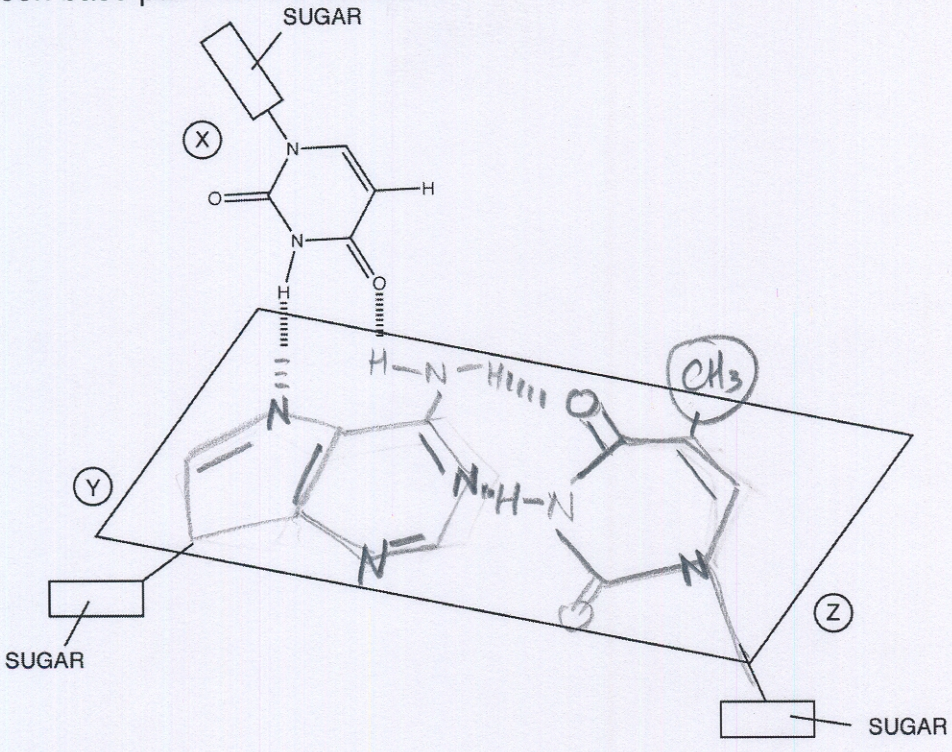
Jonas Lee James Fraser

Allen Liu

This exam consists of 5 questions, each worth 20 points total, as indicated below, for a total of 100 points. This exam counts for 200 points out of the final score of 1100 for the course, and so your score on this exam will be multiplied by 2.0 when the final score is calculated.

Question	Part A	Part B	Part C	Part D	Part E	Your Total	Maximum Score
1.	(10)	(10)	----------	----------	----------		20
2.	(7)	(8)	(5)	----------	----------		20
3.	(10)	(5)	(5)	----------	----------		20
4.	(10)	(10)	----------	----------	----------		20
5.	(20)	----------	----------	----------	----------		20
TOTAL	----------	----------	----------	----------	----------		100

Q1 A (10 points). In the schematic diagram shown below, the nucleotide labeled X forms two hydrogen bonds with the major groove edge of two DNA nucleotides, labeled Y and Z. The bases of Y and Z occupy the space indicated by the box, and form a Watson-Crick basepair: The base of nucleotide X forms a Hoogsteen base pair with the base of Y.



(i) Give the single letter code for the nucleotide labeled X. (2 points)

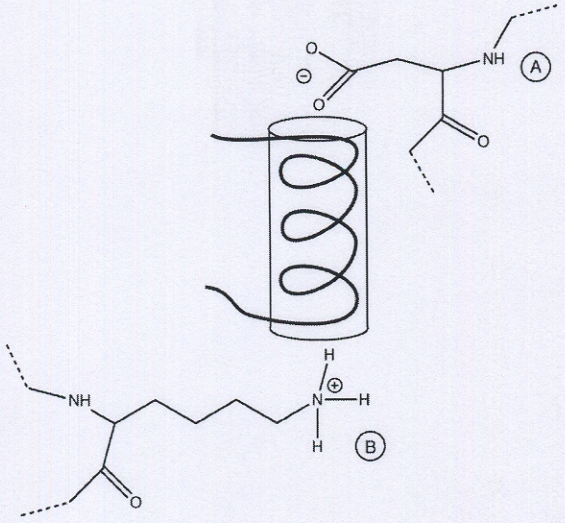
U

(i) Give the single letter codes for the nucleotides labeled Y and Z. (2 points)

A, T

(iii) Draw the chemical structure of the Watson-Crick basepair formed by Y and Z clearly, indicating all hydrogen bonds. Hydrogen bonds made between X and Y are indicated by dashed lines. Draw the basepair inside the diagram above. (6 points)

Q1B. (10 points) In the diagram shown below, two amino acid sidechains, labeled A and B, are shown interacting with the two ends of an α helix, as indicated.



(i) What are the two amino acids? Give the three letter codes. (4 points)

A = ASP, B = LYS

(i) Which is the most likely C-terminal end of the α helix? The end near the A sidechain or near the B sidechain? Clearly explain why you have made your choice. (3 points)

The end near the B residue is the likely C-terminal end. The helix dipole effect causes the C-terminal end of a helix to have net + charge, which interacts favorably with the lysine.

(i) Due to a mutation, sidechain B is replaced with a histidine sidechain. The pKa of the histidine sidechain at position B is expected to be:

(a) ~6·5 pH units

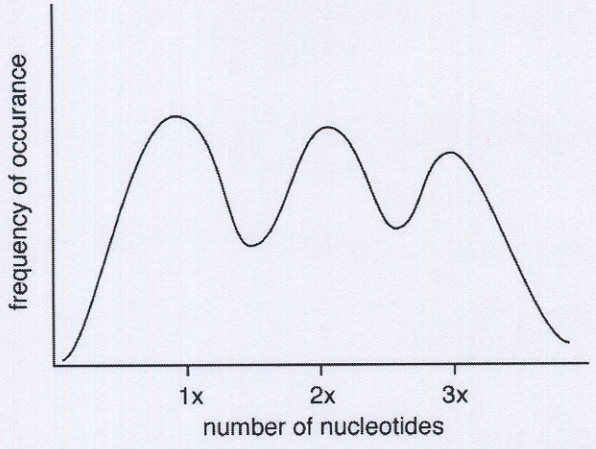
(a) increased to more than ~6·5 pH units

(a) decreased to below ~6·5 pH units

Circle the correct choice, and clearly explain your answer. (3 points)

The C-terminal end will stabilize the protonated form of His (+ charge). Hence deprotonation will require higher pH than normal ⇒ pKa will be increased.

Q2A. (7 points) The RNA molecules produced by the enzyme RNA polymerase are referred to as RNA transcripts. The diagram shown below shows the length distribution (number of nucleotides) of RNA transcripts in a bacterial cell.



(i) The distribution is seen to be periodic, with peak values occurring at lengths of 1x, 2x, 3x nucleotides. What is a reasonable value for x? Clearly explain the reasoning behind your answer. (4 points)

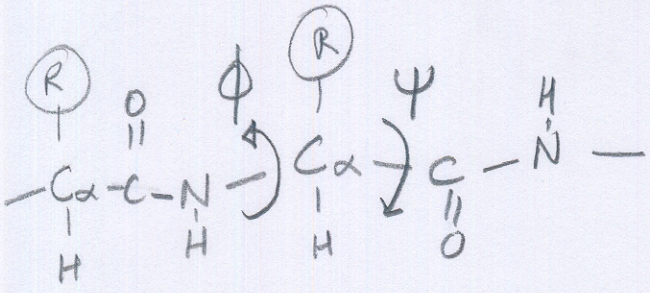
Assuming that domains in proteins are ~100 - 200 residues long, each domain is coded for by ~ 300 - 600 bases in the transcript. The periodicity is due to the domain structure of proteins, and X = 300-600

(i) The diagram shown above is for RNA transcripts in a bacterial cell. Would the length distribution of RNA transcripts (before processing) be expected to show the same kind of distribution in a eukaryotic cell (e.g., mammalian)? Clearly explain your answer. (3 points)

Eukaryotic RNA transcripts undergo splicing before becoming mRNA. Hence the periodicity in the unprocessed transcripts will be much less clear.

Q 2B. (8 points)

(i) Draw out the structure of a peptide backbone and indicate the torsion angles denoted Φ and ψ. (4 points)



(i) Which of the following amino acids is most frequently found in forbidden regions of the Ramachandran diagram?

(a) Proline
(a) Glycine
(a) Tryptophan

Clearly explain your choice. (2 points)

Because glycine has no sidechain (i.e., the R group is hydrogen) it suffers fewer vdw collisions as Φ and Ψ are changed.

(i) Which of the following amino acids is most restricted in its Φ and ψ values?

(a) Proline
(a) Glycine
(a) Isoleucine

Clearly explain your choice. (2 points)

Because the sidechain of proline is fused to the backbone its Φ and Ψ values are restricted.

Q 2C. (5 points) A bacterial DNA polymerase moves along DNA at a speed of 1000 base pairs per second while replicating the chromosome.

(a) The polymerase molecule is 100 Å long. By what multiple of its length does the polymerase move forward along the axis of the DNA helix in 1 second? Assume that the DNA stays fixed and ignore the rotational component of the motion of the polymerase. (5 points)

10 bps of DNA = 34 Å

1000 bps of DNA = 3400 Å (movement in 1 second)

= 34 × length of Polymerase

Answer: The polymerase moves ~34 × its length in 1 second.

Q. 3A (10 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 fourth 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

What purification method would be the best for each of the separations given below? In each case, clearly explain the reason for your choice of method.

(i) Separate proteins A and D. (3 points)

Gel filtration chromatography, which will resolve the large A protein from the small D protein

(i) Separate proteins B and D. (3 points)

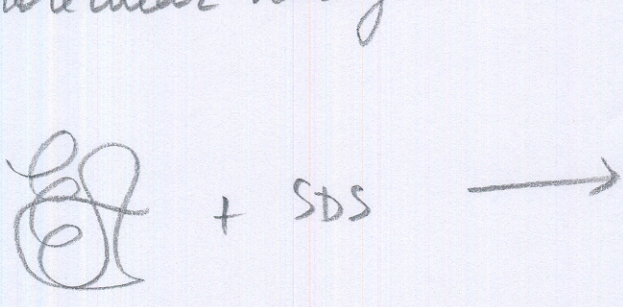
Ion exchage chromatography. For example, if a positively charged column is used at pH7, protein B will be neutral and protein D will be negatively charged. Protein B will emerge from the column first.

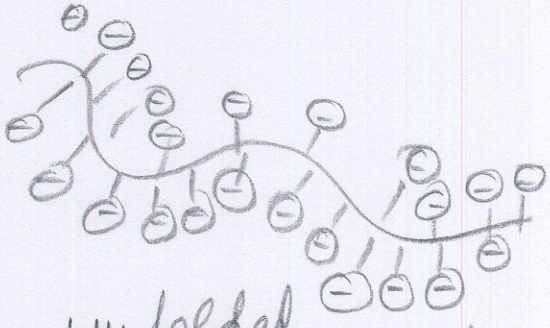
(i) Separate proteins C and D. (4 points)

Affinity chromatography. A column containing ATP linked to beads will retard the flow of D but not C.

Q 3B. (5 points) Urea and SDS are both chemicals that denature proteins. A urea-PAGE gel is one in which the proteins are denatured with urea, and then separated across an electric field in the gel. When a mixture of proteins is analyzed by urea-PAGE would we expect to see a ladder of proteins sorted on molecular weight, as we do in SDS-PAGE? Provide a clear and complete explanation for your answer, including a schematic diagram that indicates the relevant effect of SDS on proteins.

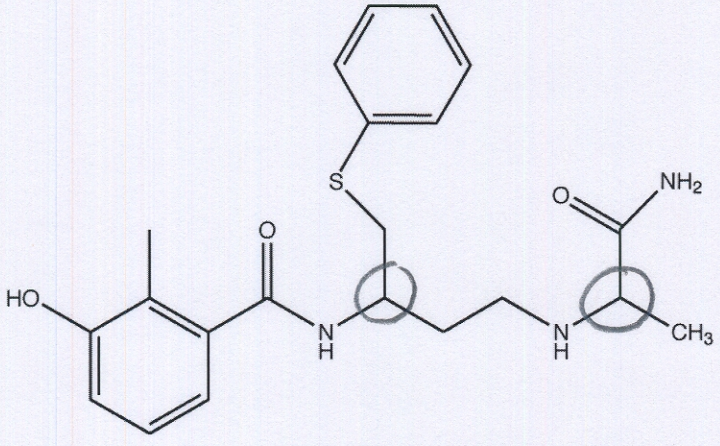
When urea is used to denature proteins the unfolded protein molecules will have a charge that depends on their pI, not on their size. In contrast, unfolded proteins are coated by so many SDS molecules that the net charge (negative) on the protein is proportional to the size of the protein. Since the gel resolves proteins based on net charge, SDS and not urea-denatured proteins will run in a ladder based on molecular weight





unfolded protein coated by SDS
Great excess of negative charge masks protein charge

Q 3C. (5 points)Shown below is the chemical structure of a drug that works by binding to and blocking the active site of a protein. A hospital purchases this drug from two different chemical suppliers. Both sources of the drug are shown to be 100% chemically pure (i.e., only compounds with this chemical structure are present in the samples). Nevertheless, one of the two samples is only 25% as potent as the other.



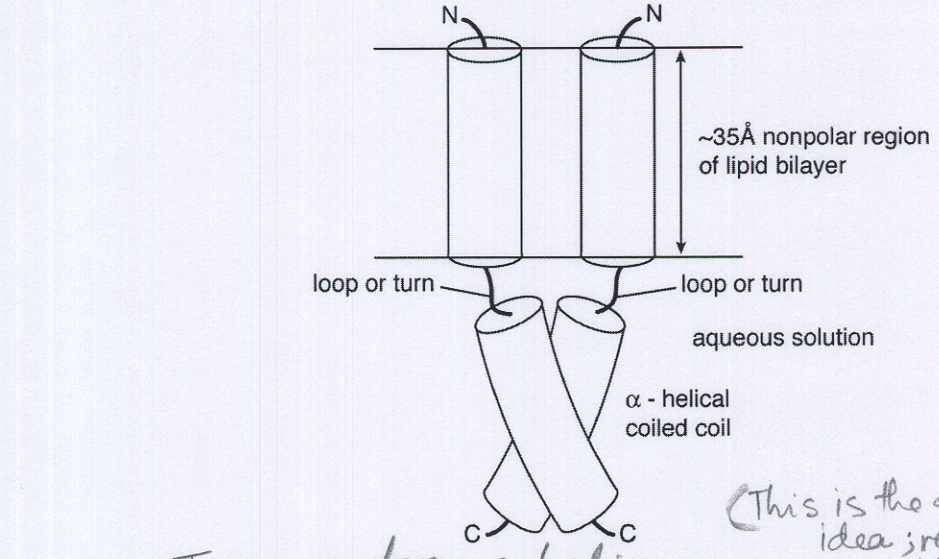
(i) Provide a clear and complete explanation for why the activity of one sample is 25% that of the other. (3 points)

(Hint: How can two molecules differ in their interaction with a protein even when they have the same chemical structure?)

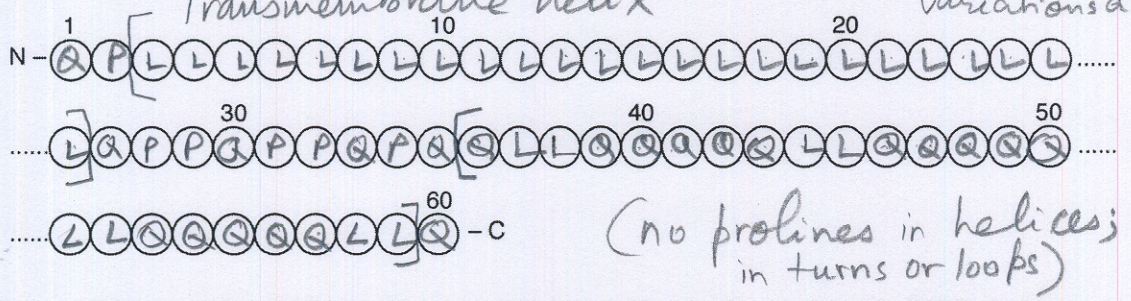
The molecule contains 2 chiral centers (circled). Hence there are 4 possible Stereo-isomers. Since protein molecules are themselves chiral, a protein will be expected to bind to a specific stereoisomer. One sample contains the pure form of the correct stereoisomer, whereas the other sample contains an equal mixture of all 4, of which only 1 (25%) is active.

(ii) Circle or highlight on the drawing any relevant features and explain what they are. (2 points)

Q 4A. (10 points) The diagram below shows the schematic structure of a 60 residue trans-membrane protein. The protein forms a single transmembrane helix, and dimerizes with another molecule by forming a symmetric coiled coil in the aqueous phase. The transmembrane helix and the helix outside the membrane are the same length.



(This is the general idea; reasonable variations are OK)



(no prolines in helices; prolines in turns or loops)

(i) How many residues do you expect each α helix to contain? Mark the boundaries of each helix in the linear sequence diagram at the bottom of the picture. Explain your answer. (4 points)

Rise per residue in α helix = 1.5 Å. Therefore 35 Å are spanned by 23-24 residues. Hence each helix is ~ 24 residues long.

Question 4A., continued.

(ii) Using the single letter code, and using only the amino acids Gln, Leu, Pro, write down a reasonable sequence for the transmembrane segment in the diagram above, within the circles in the linear sequence at the bottom of the diagram. (2 points)

(iii) Do the same for the coiled coil, and fill in the remaining spaces with reasonable residues. (4 points)

Q 4B. (10 points) A folded protein structure contains 5 ion pairs between lysine and glutamate, and there are no other lysine and glutamate sidechains in the protein. A chemical known as a cross-linker forms a covalent bridge between lysine and glutamate sidechains if they are close together in an ion pairing interaction. By analogy to the Anfinsen experiment, the following two experiments are done:

(a) folded protein $\xrightarrow{\text{add urea}}$ unfolded protein $\xrightarrow{\text{remove urea}}$ wait $\xrightarrow{\text{add cross-linker}}$ $\xrightarrow{\text{wait, then remove cross-linker}}$ measure activity.

(b) folded protein $\xrightarrow{\text{add urea}}$ unfolded protein $\xrightarrow{\text{add cross-linker}}$ wait $\xrightarrow{\text{remove urea, cross-linker}}$ wait $\longrightarrow$ measure activity.

The covalent cross-linking of glutamate and lysine sidechains does not by itself affect the activity of the protein, and the activity measured at the end of experiment (a) is taken as 100%.

(i) 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. (6 points)

Lys | Glu

1 (5) — 2 (5) — 5 ways of picking first lysine, 5 ways of picking first glu = 25 combinations

3 (4) — 4 (4) — 4 ways of picking 2nd lysine, 4 ways of picking second glu = 16 combinations

5 (3) — 6 (3) — 9 combinations

7 (2) — 8 (2) — 4 combinations

9 (1) — 10 (1) — 1 combination.

Correction term for vertical order = 5 × 4 × 3 × 2 × 1 = 120

Total = 25 × 16 × 9 × 4 × 1 = 14,400

Q4B, continued.

Correcting for equivalence of vertical orders we get:

$$\text{Combinations} = \frac{14,400}{120} = 120$$

$$\therefore \text{Activity expected} = \frac{1}{120} = 0.008 = 0.8\%$$

(Assuming only 1 ion pairing pattern is correct)

Note that in this case the left/right entries are not equivalent because two different kinds of residues are involved.

(ii) In a variation of the second experiment (b), varying amounts of salt (sodium chloride) are added along with the cross-linker in the second step. The salt is then removed with the urea and the cross-linker. Explain what you would expect to see in terms of the recovered activity as the concentration of the salt is increased. (4 points)

As more salt is added fewer ion pairs will form. Salt destabilizes ion pairs, and as salt concentrations are increased, an increasing fraction of the unfolded protein will not form random crosslinks. Hence recovered protein levels will increase.

Q 5. (20 points) Multiple choice and True/False questions. Circle the correct option (or 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.
Maximum points: 20. Minimum points: 0.

(i) All of the following are weak interactions in soluble proteins, except:
(a) Van der Waals attraction
(b) hydrogen bonds
(c) ion pairs
(d) peptide bonds
(e) hydrophobic interactions

(ii) Most interactions between site specific DNA binding proteins and DNA occur in the major groove because:
(a) hydrogen-bounding and other interactions are unique for each base pair in the major groove
(b) although the minor groove also provides unique interactions it is too narrow to allow a protein chain to enter.
(c) the phosphate groups that proteins interact with are properly aligned in the major groove.
(d) the hydrophobic effect is strongest in the major groove.

(iii) The B-form double helix is not favored in 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.

(iv) The formation of hydrogen bonds between backbone amide and carbonyl groups is the dominant force driving protein folding.
TRUE / FALSE

(v) A protein molecule forms a random coil structure in water. When the protein is transferred from water to a solvent X, it forms an α helix.
(a) Solvent X is more polar than water.
(b) Solvent X is less polar than water.
(c) Solvent X has a molecular shape that is complementary to that of a α helix.
(d) Solvent X mimics the structure of a peptide backbone.

(vi) Active sites in proteins are often located at interdomain boundaries 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 sequences of two proteins share 15% sequence identity. The two proteins are likely to have the same three dimensional structure.
TRUE / FALSE

(viii) Membrane proteins very rarely insert polar sidechains within the lipid bilayer. Nevertheless, polar sidechains within the bilayer-spanning region are highly conserved because:
(a) Their presence is required for hydrogen bonding interactions with the phosphate headgroups of the lipids. 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.

(ix) Porins are transmembrane proteins that often form water-filled channels in the membrane. This does not cause a problem for leakage of the contents of the cell because the bacteria that contain them have an inner cell membrane with no porins.
TRUE / FALSE

(d) these polar sidechains perform a critical function

(e) polar sidechains within the membrane are involved in proton transfer.

(x) Choose the amino acid substitution that results in the greatest change of hydrophobicity:

(a) A ⟶ P

(b) W ⟶ A

(c) P ⟶ Y

(d) F ⟶ R

(xi) RNA structures contain non-Watson-Crick base pairs such as G-U because the constraint of uniform base pair geometry does not apply in RNA as it does in DNA.

TRUE / FALSE

(xii) DNA is the repository of genetic information because the absence of a hydroxyl group at the 2' position makes it less susceptible to cleavage.

TRUE / FALSE

(xiii) The structure of the globin fold is relatively unchanged even though the sequences of different globins can be virtually unrelated because:

(a) Heme groups only bind to the globin fold

(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 interhelical angles.

(d) The globin fold is required to form the hydrophobic core.

(xiv) Urea is very polar molecule and SDS is very nonpolar. Yet they both denature proteins by weakening the hydrophobic effect.

TRUE / FALSE

(xv) A chemist synthesizes a peptide with the same sequence as a natural protein, but uses D-amino acids for every residue instead of L-amino acids. The natural peptide forms dimeric coiled coils. The synthetic (all D) peptide will:

(a) not form α helices

(b) form α helices, but not coiled coils

(c) form a coiled coil with a right handed super coil

(d) form a coiled coil with a left handed super coil

-END OF EXAM-