Protein Modeling - Division C Master Key

University of Texas-Austin Regional Competition

February 22, 2020

Written by Sophia Velasco, University of Florida, B.S. Biochemistry 2023

Instructions and Clarifications:

- You have **50** minutes to finish this exam and the computer exploration of protein structure. This packet contains **60%** of the overall score.
    - The **Exam** is comprised of Section **I and II**. The exam accounts for **30%** of the overall score.
    - The **Computer Exploration** is comprised of Section **III**. The computer exploration accounts for **30%** of the overall score.
- Each **participant** may bring **one** 8.5" x 11" sheet of paper that may be in a sheet protector or laminated that contain information without any annotations or labels affixed along with writing utensils for each participant.
- Anything written on the exam packet will **not** be graded. Only the **answer sheet** will be graded.
- Write your team number on every page of the answer sheet.
- Tiebreakers are labeled as **TB#**. There are **five** tiebreakers in this exam.
- If you have any questions or comments about this exam, feel free to email me at *velasco.scienceolympiad@gmail.com.* **Happy testing!**

I. Protein Biochemistry

Multiple Choice: Choose the most appropriate response for each answer. For questions that are multiple select, **all** the correct answers must be chosen to earn credit. **No partial credit** will be given for multiple select questions. Each question is worth **two** points. **(30)**

1. Which of the following is the most basic level of protein structure?
   a. Primary
   b. Secondary
   c. Tertiary
   d. Quaternary

2. In which of the following protein structures would hydrogen bonding between peptide -CO-NH groups and NH-OC occur? (*Multiple select*)
   a. Primary
   b. Secondary
   c. Tertiary
   d. Quaternary

3. Most proteins take what structure/shape once they have folded?
   a. Globular
   b. Fibrous
   c. Alpha helices
   d. Beta sheets

4. Which of the following is the biggest determinant in the ability of a protein to fold?
   a. Hydrophobic interactions
   b. Disulfide bonds
   c. Hydrogen bonds
   d. Thermodynamic stability of the complex

5. What will happen if a protein is not in the lowest energy conformation?
   a. A protein will become denatured and stop functioning.
   b. A protein will continue to move and adjust until it finds its most stable state.
   c. A protein will stay at its current energy conformation.
   d. A protein will attempt to reach the highest energy conformation.

6. Which of the following statements below would be correct for a theoretical protein that has reached its conformation?
   a. The interior of the protein would mostly contain amino acids like histidine, arginine and lysine.
   b. The interior of the protein would mostly contain amino acids like cysteine, glycine, and tyrosine.
   c. The interior of the protein would mostly contain amino acids like phenylalanine, leucine and tryptophan.
   d. The interior of the protein would mostly contain amino acids like aspartic acid and glutamic acid.

7. What should be the value of the Gibbs free energy of protein folding?
   a. Positive (>0)
   b. Negative (<0)
   c. 0
   d. None of the above. Gibbs free energy is not used to describe protein folding.

8. Which of the following are techniques that could be used to study protein folding? (*Multiple select*)
   a. X-ray crystallography
   b. Fluorescence spectroscopy
   c. Circular dichroism
   d. Protein nuclear magnetic resonance spectroscopy

9. In one of Cyrus Levinthal's papers, what was his estimation of the number of possible protein conformations?
   a. 3^{300}
   b. 4^{154}
   c. 1×10^{9}
   d. 2^{100}

10. Which principle states that naturally evolved proteins have optimized their folding energy landscapes? **TB#1**
    a. Law of conservation of energy
    b. Principle of minimal frustration
    c. Levinthal's paradox
    d. Anfinsen's dogma

11. Which of the following is/are component(s) of an amino acid? (*Multiple select*)
    a. $-NH_2$
    b. -COOH
    c. C
    d. R group

12. What is the term for amino acids that have both the amine and carboxylic acid groups attached to the alpha carbon atom?
    a. 1-,alpha
    b. 2-,alpha
    c. 3,-alpha
    d. 4,-alpha

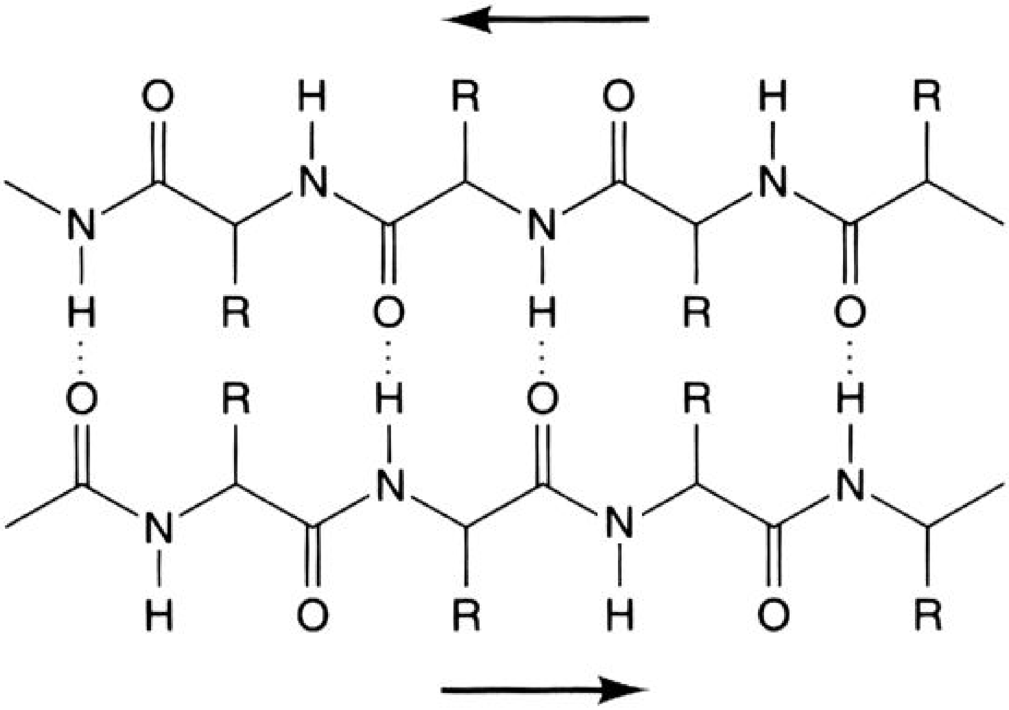
13. What is the term for the carbon atom next to the carboxyl group?
    a. α carbon
    b. β carbon
    c. σ carbon
    d. γ carbon

14. Which of the following amino acids below is considered achiral?
    a. Methionine
    b. Cysteine
    c. Tyrosine
    d. Glycine

15. What is the term for the exact midpoint between two pKa values where the average net charge of all the forms present is zero?
    a. Isomerism
    b. Chiralism
    c. Zwitterions
    d. Isoelectric point

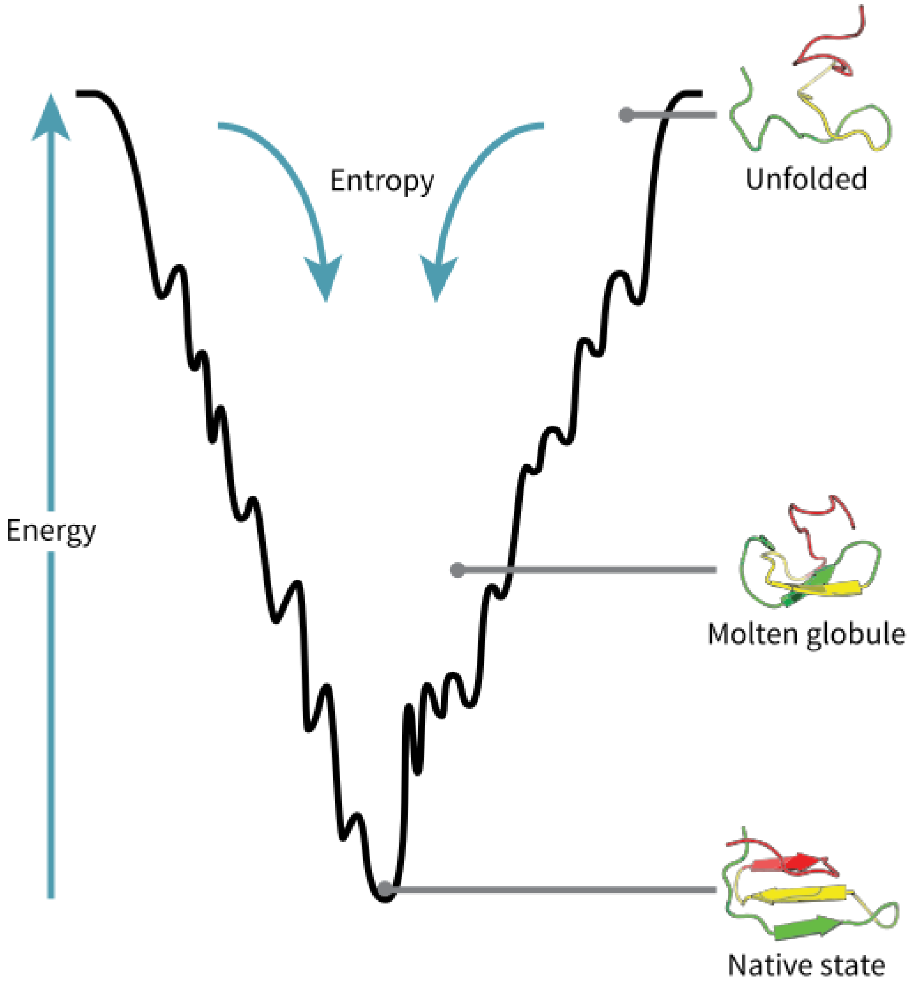
Diagram-based Questions: Answer the questions below referring to the diagram or ideas behind the diagram. Complete sentences are **not** required for this section. **(70)**

16. **Figure 1.1** shows a **beta-sheet**. Answer the following questions below. **(11)**



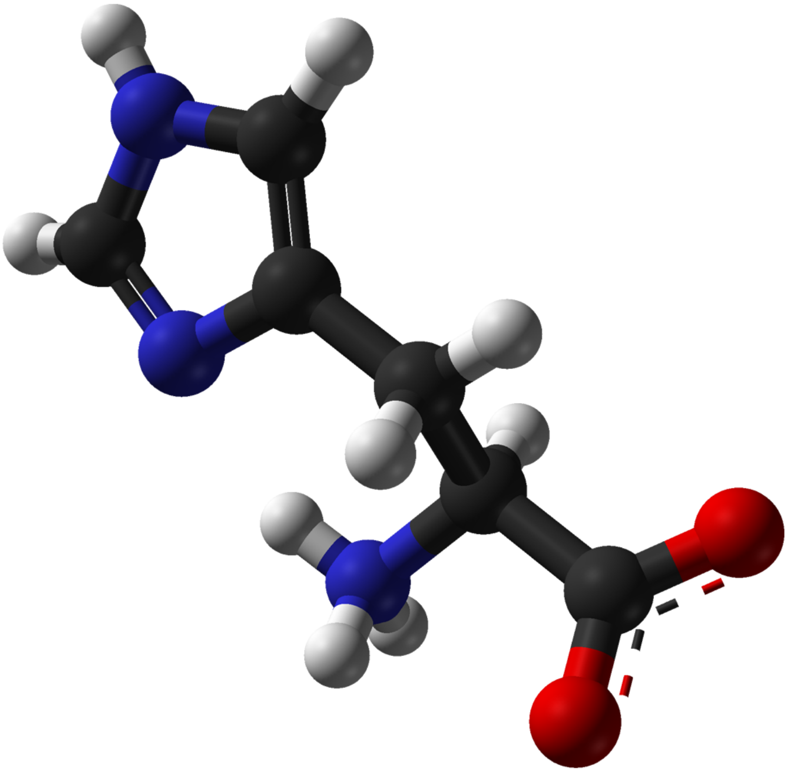
A. What type of beta-sheet is shown above? (1) Antiparallel
B. Is this beta-sheet more stable than the other type? Explain your answer. (3) Yes, it is more stable. (1) The hydrogen bonds has the ideal 180 degree angle compared to the slanted hydrogen bonds in parallel beta sheets. (2)
C. What is the value of ϕ in this type of beta-sheet? (2) -140 degrees
D. What is the value of ψ in this type of beta-sheet? (2) 135 degrees
E. List an amino acid that are favored to be found in beta-strands in the middle of beta-sheets. (3) **TB#2** Tyrosine, phenylalanine, tryptophan, threonine, valine, isoleucine. (Give 3 points for any one of the amino acids listed)

17. **Figure 1.2** shows the **energy funnel** in the process of a protein assuming its native structure. The following questions below refer to this process. **(8)**



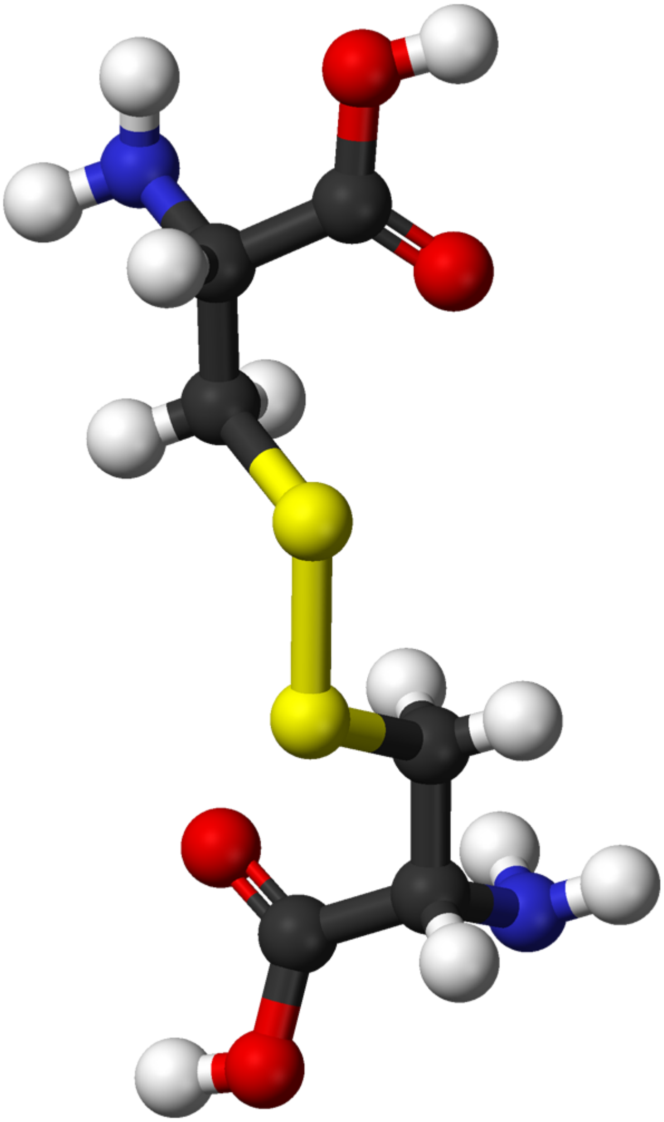
A. Who was the person who coined the term "funneled energy landscape"? (1) José Onuchic
B. In the funnel landscape, is the protein restricted to a single mechanism of folding or is the protein allowed to fold through a large number of pathways and intermediates? (2) Large number of pathways and intermediates
C. At which region of the funnel landscape does a protein have the highest energy state? (2) Top
D. What is the term for a point in the funnel landscape where the transition state for a protein is found? (2) Saddle point
E. **True or False:** No protein could assume its native conformation without passing the transition state. (1)

18. **Figure 1.3** shows a **specific amino acid.** Answer the following questions referring to the amino acid. **(10)**



A. What amino acid is shown above? (1) Histidine
B. Is this protein polar, nonpolar, positively charged, or negatively charged? (1) Positively charged
C. List **one** codon that codes for this amino acid. (2) CAU and CAC (give two points for any of the two listed)
D. Which sidechain of this amino acid is a common coordinating ligand in metalloproteins and is part of the catalytic sides in certain enzymes? (4) Imidazole
E. What conjugate acid of the imidazole side chain of this amino acid has what pK_a? (2) 6.0

19. **Figure 1.4** is a structure present in the **tertiary structure of proteins**. The following questions below refer to this structure. **(11)**



A. What is the name of this structure? (3) Cystine (1 point for cysteine)
B. What is the typical bond dissociation energy of this structure in kcal/mol? (2) 60
C. Is this structure shorter or longer than a C-C bond? (2) Longer
D. Is this structure formed in oxidizing or reducing conditions? (2) Oxidizing
E. Is this structure stable or unstable in the cytosol? (2) Unstable

20. Complete **Figure 1.5,** which shows the different types of alpha helices. Each letter is worth **two** points. Pitch, radius and rise per residue are all in angstroms (Å). **(30)**

S.No.	Types of helix	No. of residue per turn	Pitch	Radius of helix	Rise per residue
1.	α helix	A	E	I	L
2.	B	3	F	J	M
3.	2.2_7 ribbon	C	G		N
4.	D	4.4	H	K	O

A. 3.6
B. 3_{10} helix
C. 2.2
D. Pi (π) helix
E. 5.4
F. 5.8
G. 6
H. 5.2
I. 2.3
J. 1.9
K. 4.8
L. 1.5
M. 2.0
N. 2.72
O. 1.1

II. DNA/RNA Structure and CRISPR-Cas Systems

Multiple Choice: Choose the most appropriate response for each question. Each question is worth **two** points. **(30)**

21. What is a nucleoside?
    a. A ribose sugar connected to a base through the 1’C.
    b. A ribose sugar connected to a base through the 5’C.
    c. A deoxyribose sugar connected to a base through the 1’C.
    d. A deoxyribose sugar connected to a base through the 5’C.
    e. More than one of the above

22. Which chemical interaction is mostly responsible for the stabilization of the RNA secondary structure?
    a. Ionic bonds
    b. Covalent bonds
    c. Disulfide bonds
    d. Hydrogen bonds
    e. More than one of the above

23. Which of the following on RNA causes it to self-hydrolyze?
    a. Nucleoside
    b. Nucleotide
    c. Magnesium ion
    d. 2’ hydroxyl group
    e. More than one of the above

24. Which of the following statements regarding the three-dimensional structure of DNA is correct?
    a. The three-dimensional structure of DNA is the double helix.
    b. In DNA, the two-ring base is always paired up with the single-ring base.
    c. The two chains of DNA are held together by hydrogen bonding.
    d. The complementary base-pairing results in the base pairs being packed in the energetically most favorable arrangement in the interior of the double helix.
    e. More than one of the above

25. Which of the following contributes to the increased stability of DNA in comparison to RNA?
    a. The phosphodiester bond is more susceptible to hydrolytic attack.
    b. The phosphodiester bond of DNA has a negative charge, which repels nucleophilic attacks by hydroxide ions.
    c. There is an extra hydroxide group on the 2'C of DNA.
    d. The phosphodiester bond of DNA has a positive charge, contributing to a stronger attraction between the base pairs.
    e. More than one of the above

26. Which CRISPR-Cas type(s) uses Cas6 endoribonuclease within the Cascade complex for primary processing of the pre-crRNA?
    a. I
    b. II
    c. III
    d. A and B
    e. A, B, and C

27. Which CRISPR-Cas type(s) requires crRNA, tracrRNA, and two domains of Cas9 (RuvC and HNH) for target cleavage?
    a. I
    b. II
    c. III
    d. A and B
    e. A, B, and C

28. Which CRISPR-Cas type(s) requires the annealing of the tracrRNA to the repeat sequences of the pre-crRNA and the subsequent cleavage of the dsRNA by the host RNase III?
    a. I
    b. II
    c. III
    d. A and B
    e. A, B, and C

29. Which of the following statement(s) about Type II CRISPR-Cas systems is correct? **TB#3**
    a. The pairing between tracrRNa and precursor crRNA creates a single-stranded substrate claved by RNase III.
    b. The cleavage of DNA target in this CRISPR-Cas system type is the Cas9 protein.
    c. The formation of a complex with crRNA:tracrRNA drives Cas9 conformation changes that direct target DNA binding and cleavage in a PAM-independent manner as shown in recent studies.
    d. They require more Cas machinery for immunity than Type I and III systems.
    e. More than one of the above

30. Which of the following statement(s) about CRISPR-Cas9's role in genetic engineering is correct?
    a. The crRNa and tracrRNA can be combined into sgRNA, which is capable of guiding Cas9 to complementary sequences.
    b. The NHEJ repair pathway is considered to be more precise than the HDR pathway.
    c. A catalytically inactive Cas9 endonuclease can still be used as a DNA targeting module.
    d. The Cas9 endonuclease and sgRNA can be both engineering to create DSBs.
    e. More than one of the above

31. The spacer acquisition in Type I systems require the overexpression of which *cas* protein(s)?
    a. *cas1*
    b. *cas2*
    c. *cas9*
    d. A and B
    e. All of the above

32. Which of the following would you expect to bind tightly to dsDNA in a sequence-independent manner?
    a. *cas1*
    b. *cas2*
    c. *Cas9*
    d. A and B
    e. All of the above

33. In which step of CRISPR immunity are CRISPR arrays transcribed and processed into crRNAs?
    a. Step 1
    b. Step 2
    c. Step 3
    d. More than one of the above
    e. None of the above

34. In which step of CRISPR immunity are new spacers acquired from exogenous nucleic acid into the CRISPR locus?
    a. Step 1
    b. Step 2
    c. Step 3
    d. More than one of the above
    e. None of the above

35. In which step of CRISPR immunity do crRNAs guide Cas nucleases for specific cleavage of nonhomologous sequences?
    a. Step 1
    b. Step 2
    c. Step 3
    d. More than one of the above
    e. None of the above

Short Answer: Answer the short answer questions below. The point values are indicated by the number in the parentheses. Complete sentences are **not** required for this section. **(62)**

36. There are drastic differences between the structure of DNA and RNA. Answer the following questions about these differences below. **(10)**
    a. Which structure would have a narrow and deep structure in its major groove: DNA or RNA? (1) RNA
    b. Which structure would have a wide and shallow morphology in its major groove: DNA or RNA? (1) DNA
    c. There are at least how many forms of right-handed DNA double-helical structures? (2) 2
    d. Which form of right-handed DNA double-helical structure is said to be the most stable in high humidity? (*Hint: This structure was proposed by Watson and Crick*). (2) B-DNA structure
    e. Why can RNA only form an A-type double helix? (4) steric restrictions (2) imposed by 2' hydroxyl (-OH) residue on ribose (2)

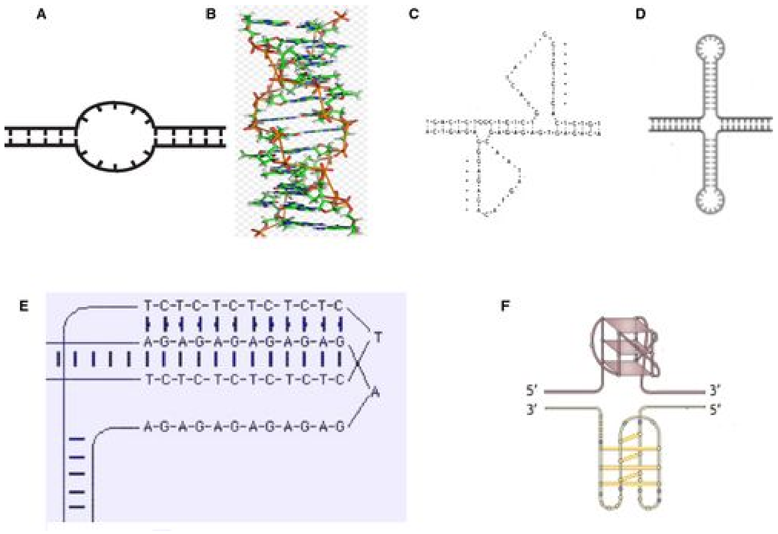
37. Answer the following questions about the biochemistry and structure of RNA molecules. **(12)** **TB#4**
    a. How many hydrogen bonds can form between guanine and cytosine in RNA? (2) 3
    b. How many hydrogen bonds can form between adenine and uracil in RNA? (2) 2
    c. How many hydrogen bonds can form between guanine and uracil in RNA? (2) 2
    d. Which **two** chemical interactions contribute to the folding of RNA into its three-dimensional structure? (4) hydrogen bonding (2) and base stacking (2)
    e. **True or False:** Noncanonical pairing and hydrogen bonding does **not** contribute to RNA folding. (2) False

38. There have been recent developments in CRISPR technology in terms of base-editing technology. These so called "*second-generation genome-editing tools*" can precisely convert a single base into another **without** causing double strand DNA breaks (DSBs). Answer the following questions about the evolution of second-generation CRISPR gene-editing tools. **(20)**
    a. Komor et al. recently demonstrated that a fusion complex composed of nickase Cas9 fused with two other proteins can effectively convert Cytosine (C) to Thymine (T) at the target site without DNA DSBs. What are these **two** proteins? (10) APOBEC1 deaminase enzyme (5) and Uracyl Glycosylase (UGI) protein (5); (provide 2 points partial credit for APOBEC; do **not** accept cytidine deaminase)
    b. What is the name of the protein can be evolved and fused to nickase Cas9 to develop a base editor that achieves direct A-G conversion at the target sites? (5) transfer RNA adenosine deaminase
    c. A complex resulting from the fusion of dCas9 and another protein can act as a gain of function screening tool. What is this protein fused with dCas9? (5) (*Hint: This protein can convert cytosine to uracil through deamination*). Activation-induced adenosine deaminase (AID)

39. The incorporation of CRISPR-Cas9 systems in genetic engineering has increased in recent times. Answer the following questions about CRISPR-Cas9 and its scientific uses. **(20)**
    a. Does CRISPR-Cas9 perform better in site selection in comparison to ZFNs and TALENs? (1) yes
    b. How do most CRISPR-Cas9 systems express the guide RNA from a plasmid (*Hint: What general promoter type is used?*) (3) RNA polymerase III promoter
    c. What two additional structures are needed if we were to deliver Cas9 via viral vectors? (6) promoter (3) and polyadenylation sequence (3)
    d. How can scientists create an action similar to a DSB with 5' overhangs (that is later repaired by NHEJ or HDR) by only introducing a single-strand nick into the target DNA? There is **no** partial credit for this question. (8) use a pair of "nickase" (4) CRISPR-Cas9 complexes with binding sites on opposite strands flanking the target site (4)
    e. **True or False:** Despite the fact that each nucleotide within the 20 nucleotide protospacer contributes to the overall *S. pyogenes* Cas9 binding and specificity, single mismatches are still tolerated. (2) True

Diagram-based Questions: Answer the questions referring to **Figure 2.1** or ideas behind **Figure 2.1**. Complete sentences are **not** required for this section. **(18)**

40. Under physiological conditions, the energetic environment of DNA sequences favors the formation of alternative DNA structures. Determine the name of the alternative DNA structures shown for each letter. Acronyms will be accepted for some of the answers. Each letter is worth **three** points. **(18)**



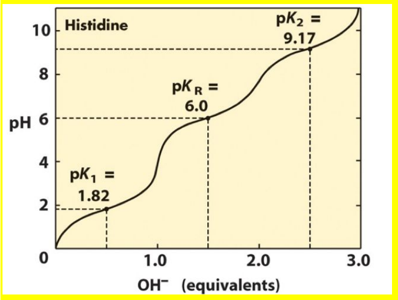
A. DNA bubble
B. Z-DNA
C. Slipped loop
D. Cruciform
E. H-DNA
F. G-quadruplex/*i-motif* in double stranded DNA

III. Computer Exploration - Jmol

Directions: The following questions below is based on the crystal structure of cytidine deaminase from *Burkholderia pseudomallei* (**3DMO.pdb**). These questions are related to the **structure** of this protein. Answer these questions in the most concise and specific way possible. The point values for each question are addressed in the parentheses. **(123)** TB#5

41. The following questions are based on **residue 20**. **(11)**
    a. Identify this amino acid. (1) Proline
    b. What is the XYZ coordinates of this atom on Jmol for atom number 1135? (3) -2.815, 27.041, -27.552
    c. The side chain of this amino acid classifies it as a nonpolar aliphatic amino acid. What is the name of this side chain? (3) pyrrolidine
    d. **True or False:** This is the only proteinogenic amino acid with a secondary amine. (2) True
    e. **True or False:** This amino acid is considered to be conformationally rigid due to its cyclic side chain. (2) True

42. The following questions are based on **residue 4**. **(34)**
    a. What is the **one letter name** of this amino acid? (2) H
    b. Is the α-amino group of this amino acid protonated or deprotonated under biological conditions? (2) protonated
    c. Considering the structure of this amino acid, infer why it can participate in acid-base catalysis. (8) it can **switch** (2) between **protonated** (3) and **unprotonated** (3) states
    d. Draw the titration curve for this amino acid in its **protonated form.** Be sure to label the axes and pK points (plus values). (22)



Point breakdown:

- 2 points for curve having three hills
- 2 points per pK label (pK_1, pK_r, and pK_2) for a total of 6 points
- 2 points for having the axes labeled (1 point for pH, 1 point for OH- or equivalents like NaOH)
- 2 points each for correct values of pK (1.82, 6.0, and 9.17 for a total of 6 points)
- 2 points each for correct general area of pH values (~2, 6, and ~9 for a total of 6 points)

43. The following questions are based on **residue 38**. **(9)**
    a. Identify this amino acid. (1) phenylalanine, Phe, F
    b. State the IUPAC name of this amino acid. (2) (*S*)-2-Amino-3-phenylpropanoic acid
    c. This amino acid is a precursor for which other amino acid? (2) tyrosine
    d. What is the chemical formula of this amino acid? (2) $C_9H_{11}NO_2$
    e. The benzyl side chain of this amino acid classifies the amino acid as what? (2) nonpolar

44. The following questions are based on **residue 78**. **(19)**
    a. Identify this amino acid. (1) Valine, Val, V
    b. The isopropyl side chain of this amino acid classifies it what? (4) nonpolar (2) aliphatic (2) amino acid
    c. This amino acid is associated with sickle-cell disease. In sickle cell disease, a single glutamic acid in beta-globin is replaced with this amino acid (residue 78). As a result, the hemoglobin (Hb) is prone to abnormal aggregation. Based on your knowledge of the structure of this amino acid, infer why this mutation could lead to abnormal aggregation of Hb. (10) glutamic acid is hydrophilic (5) whereas valine/residue 78 is hydrophobic (5)
    d. Is the α-amino group of this amino acid protonated or deprotonated under biological conditions? (2) protonated
    e. Is the α-carboxylic acid group of this amino acid protonated or deprotonated under biological conditions? (2) deprotonated

45. The following questions are based on **residue 75**. **(16)**
    a. Identify the three letter name of this amino acid. (2) Ile
    b. Describe the side chain of this amino acid. (8) hydrocarbon (2) side chain with a branch (2) (central carbon atom [2] bound to three other carbon atoms [2])
    c. What is the chemical formula of this amino acid? (2) $C_6H_{13}NO_2$
    d. Is this amino acid polar or nonpolar? (2) nonpolar
    e. Is this amino acid charged or uncharged at physiological pH? (2) uncharged

46. Calculate the hydrogen bonds for the following **(10)**:
    a. Entire protein structure (2) 306
    b. Residues 1-19 (2) 48
    c. Residues 28-39 (2) 16
    d. Residues 46-73 (2) 58
    e. Residues 82-88 (2) 0

47. Calculate the struts for the following **(10)**:
    a. Entire protein structure (2) 89
    b. Residues 4-6 (2) 0
    c. Residues 78-90 (2) 5
    d. Residues 54-66 (2) 2
    e. Residues 10-30 (2) 4

48. How many atoms are there for the helices present in this protein? **(2)** 1457

49. How many atoms are there for the sheets present in this protein? **(2)** 838

50. Answer the following questions about the **overall** protein structure **(10)**:
    a. How many hydrophilic atoms are there in the overall protein structure? (2) 0
    b. How many hydrophobic atoms are there in the overall protein structure? (2) 2119
    c. How many charged atoms are there in the overall protein structure? (2) 1047
    d. How many hetero atoms are there in the overall protein structure? (2) 403
    e. How many water **atoms** are there in the overall protein structure? (2) 362