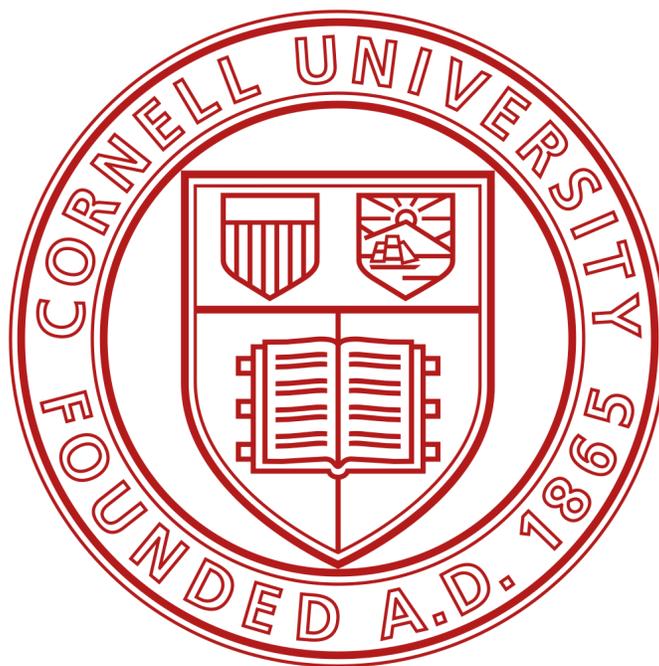


Protein Modeling Division C - Master Key

Cornell University Science Olympiad Invitational

November 23, 2019



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

Instructions and Clarifications:

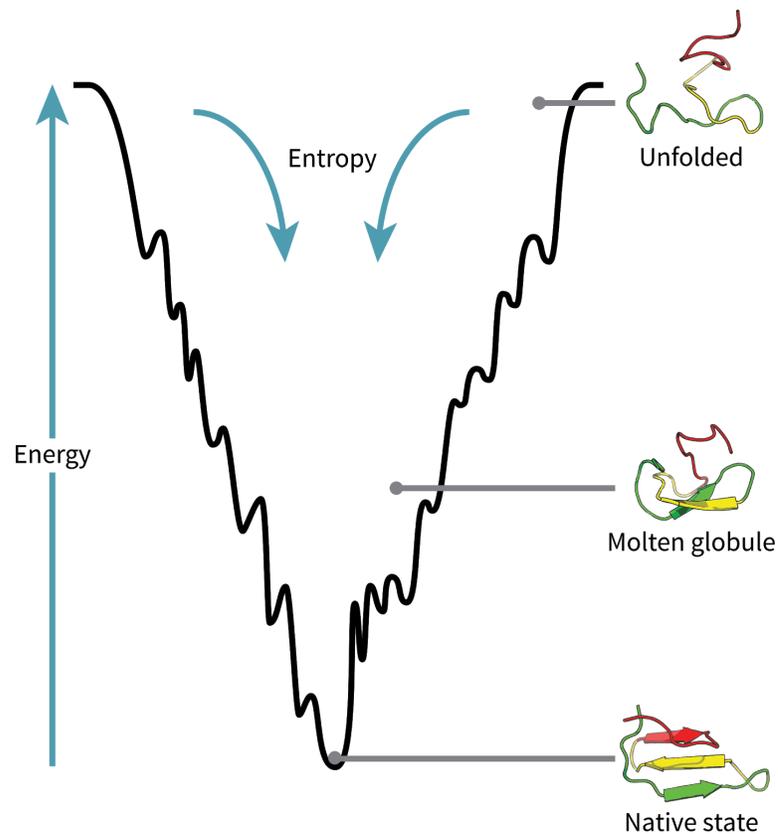
- You have **50** minutes to finish this exam and the computer exploration of protein structure. The **Exam** portion is composed of Section **I** and **II** and is worth 30% of the overall score. The **Computer Exploration** portion is composed of Section **III** and is worth 30% of the overall score.
- Each **participant** may bring **one** 8.5" x 11" sheet of paper that contains information about this event without any annotations or labels affixed along with writing utensils for each participant.
- You **may** split the exam but you are responsible for placing them in the correct order afterwards.
- Write your team number in **every** page of the answer sheet.
- There are **5** tiebreakers in this exam. They are marked **TB#**.
- Anything written on the exam will **not** be graded. Only the **answer sheet** will be graded.
- If you have any questions or comments about this exam, feel free to email me at velasco.scienceolympiad@gmail.com. **Happy testing!**

Section I: General Biochemistry

Directions: Answer the following questions about biochemistry in the most concise way possible. The point values for each question are addressed in the parentheses. **(85)**

1. Hydrophobic interactions is an essential aspect of protein folding. Explain the hydrophobic effect. (2) Hydrophobic chains of a protein collapse into the core of the protein. (1) As a result, the hydrophilic chains of a protein is located in the protein's exterior. (1)
2. When a protein is placed in an aqueous environment, the water molecules tend to aggregate around hydrophobic regions and form ordered shells of water molecules. The following questions below refer to this phenomenon. (3)
 - a. How does this phenomenon affect the entropy of the system? (1) Negative change/less entropy in the system/decreases entropy in the system/less disorder in the system/more order in the system.
 - b. How do the hydrophobic groups react against this phenomenon? (1) They fold inwards and form the interior of the protein.
 - c. How do the hydrophobic groups reintroduce entropy back into the system? (1) The hydrophobic collapse (hydrophobic groups folding inward) causes the breaking of the ordered watered cages. This frees the water molecules.
3. According to Cyrus Levinthal, there is an estimation of 10^{143} possible conformations in an unfolded polypeptide chain due to all the following conformations of a protein. The following questions below refer to Levinthal's paradox. (3)
 - a. True or False: If the conformations were folded in a picosecond scale, the time it takes to reach the desired conformation is possible within five minutes. (1)
 - b. Levinthal proposed that random conformation search does not occur in the process of protein folding. Based on his information, infer how Levinthal believed proteins reach their desired conformation. (2) Proteins fold through a series of metastable intermediate states.

4. Joseph Bryngelson and Peter Wolynes proposed that proteins follow the principle of minimal frustration. Explain this principle. (6) Natural evolved proteins optimize their folding energy landscapes. (2) Nature has chosen amino acid sequences so that the folded state of the protein is stable. (2) The process of acquiring the folded state is relatively fast. (2)
5. In the protein folding energy funnel landscape, there is a saddle point where the transition state for a protein is found. What is the transition state? (2) The conformation that must be assumed by every molecule of the protein if the protein wants to assume its native structure.
6. Draw the energy funnel by which an unfolded polypeptide chain achieves its native structure. To acquire all points, address energy, entropy, and the location of the unfolded protein, molten globule, and native state of a protein relative to the funnel. (6) **TB#1**



Answer:

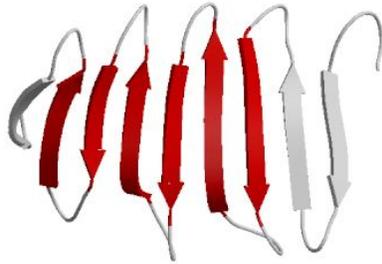
- (1) - drawing the energy arrow pointing up
- (1) - drawing entropy pointing downwards
- (1) - addressing unfolded protein in the top of the drawing
- (1) - addressing molten globule in the middle of the drawing
- (1) - addressing the native state in the bottom of the drawing
- (1) - drawing the squiggly downwards cone representing energy

7. Hydrogen bonds in alpha helices occur between the N-H group of an amino acid and a C=O of a different amino acid. However, the distance between the hydrogen bonds differ for different types of alpha helices. Determine the distance of the carboxyl group of an amino acid that would form a hydrogen bond with an amine group of a different amino acid relative in the following alpha helices in terms of residues. (6) **TB#2**
- a. Alpha helix $n+4$ (four residues earlier) (2)
 - b. 3_{10} helix $n+3$ (three residues earlier) (2)
 - c. Pi-helix $n+5$ (five residues earlier) (2)
8. Some amino acids have a propensity to be in different regions of a beta-strand due to their characteristics. Determine if the amino acids below would be located in the edge or middle of beta-sheets. (7)
- a. Tyrosine (1) middle
 - b. Proline (1) edge
 - c. Tryptophan (1) middle
 - d. Threonine (1) middle
 - e. Valine (1) middle
 - f. Isoleucine (1) middle
 - g. Phenylalanine (1) middle

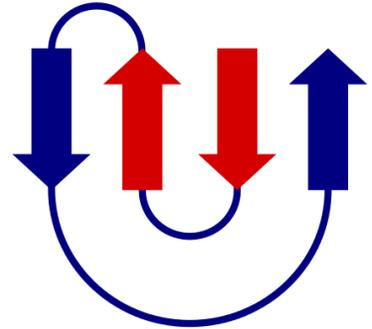
9. Structural motifs are short segments found in a protein that are considered to be spatially close but not necessarily adjacent in the sequence. There are different structural motifs that can be found within a beta-sheet. Determine what structural motif is shown by the pictures below. (8)

Answers: a. Beta meander (2) b. Beta-hairpin (2) c. Greek key (2) d. Psi-loop (2)

a. (2)



c. (2)



b. (2)



d. (2)



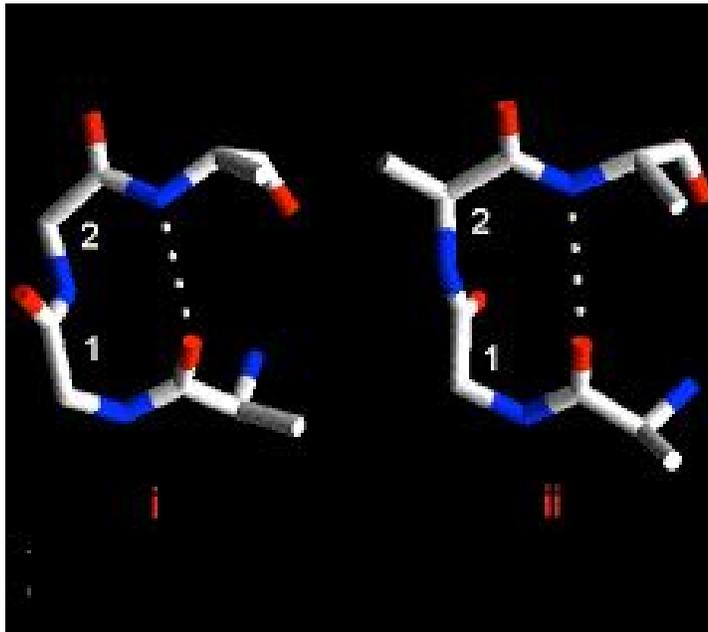
10. The secondary structure of a protein can be described by its topology. A topology of a beta-sheet describes the order of the hydrogen-bonded beta-strands along the backbone. Let's say a hypothetical protein called Cornellion has a five-stranded, parallel beta-sheet with the topology of 41235. Based on this topology, determine the following: (10)

- What are the edge strands of this protein? **4 and 5** (2)
- Determine the hydrogen bond relationship between these five beta-sheets based on the topology. **Beta-strand 4 is hydrogen bonded to beta-strand 1.** (2) **Beta-strand 1 is hydrogen bonded to beta-strand 2.** (2) **Beta-strand 2 is hydrogen bonded to beta-strand 3.** (2) **Beta-strand 3 is hydrogen bonded to beta-strand 5.** (2)

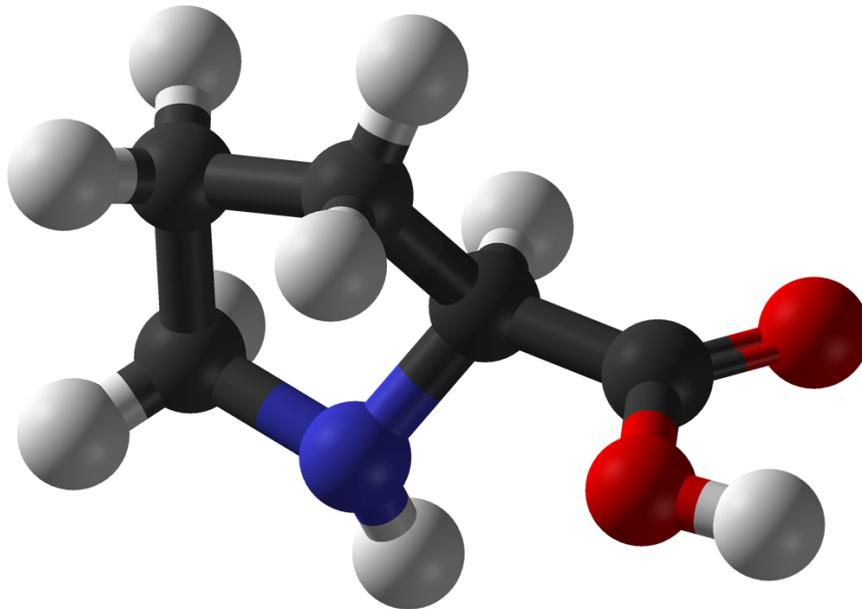
11. A beta-turn is a structure that can be found within a protein that has four amino acids turned back on themselves. The following questions below refer to this structure. (10)

TB#3

- How many hydrogen bonds are found at the corners of beta turns? (1) **One**
- How many amino acids are located at each corner of beta turns? (1) **Two**
- What three symbols are used to define the first through last residue found in a turn if read from the N-terminus to the C-terminus? (6) **i** (2), **i+2** (2), **i+3** (2)
- Using the image below, what type of beta turn is (i)? (1) **Type I**
- Using the image below, what type of beta turn is (ii)? (1) **Type II**

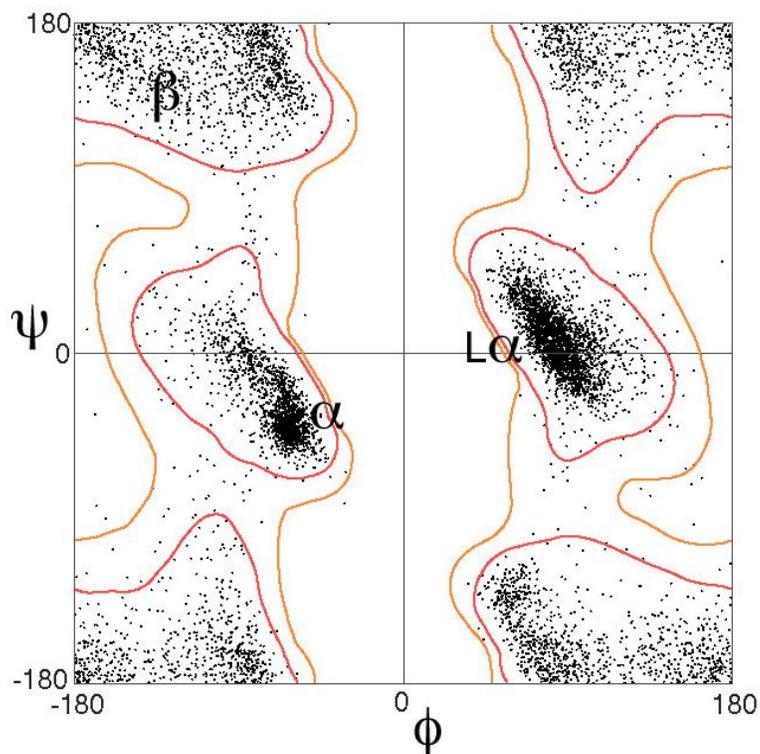


12. The picture below refers to an amino acid that offers unique properties when it comes to protein folding. The following questions below refer to this amino acid. (6)



- A. What is the amino acid shown above? (1) **Proline**
- B. True or **False**: When this amino acid is bound as an amide in a peptide bond, its nitrogen is not bound to any hydrogen. It can act as a hydrogen bond donor but not as an acceptor. (1)
- C. Where can this amino acid be usually found in alpha helices? (1) **First residue of an alpha helix**
- D. Where can this amino acid be usually found in beta sheets? (1) **Edge strands**
- E. **True** or False: This amino acid is usually solvent-exposed even though it has a completely aliphatic side chain. (1)
- F. Is this amino acid conformationally flexible or **inflexible**? (1)

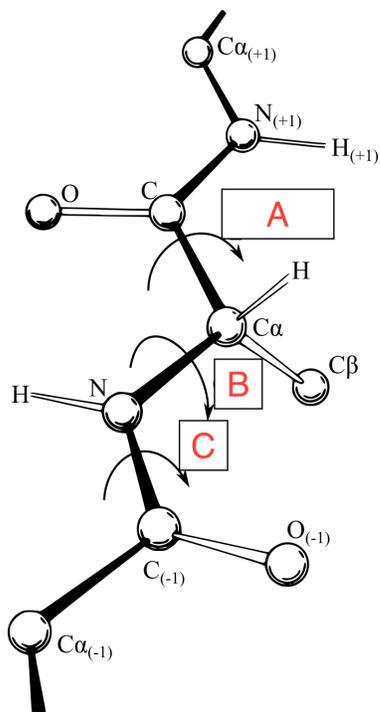
13. The following questions refer to the Ramachandran plot below. (5)



(5)

- What amino acid has this characteristic Ramachandran plot? (1) **Glycine**
- Support your answer by referring to the characteristics of the amino acid and analyzing the Ramachandran plot. (4) **Glycine has a hydrogen atom for its side chain. (1) It also has a smaller van der Waals radius than the CH_3 , CH_2 , or CH group that is usually characteristic of all other amino acids. (1) As a result, glycine is the least restricted of all amino acids. (1) Because of this characteristic, the Ramachandran plot has a lot of allowable area, as shown in this plot. (1)**

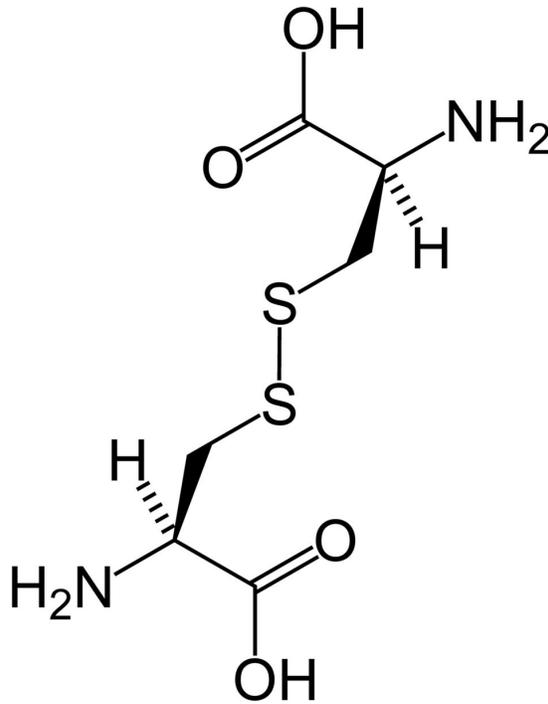
14. Draw the corresponding symbols for the dihedral angles in the diagram below. Each letter is worth **one** point. (3)



- A. (1)
- B. (1)
- C. (1)

15. Disulfide bonds is a bond that can be found in tertiary structures in proteins. The following questions refer to disulfide bonds. (8)

- Are disulfide bonds usually formed in an **oxidation** process or reduction process? (1)
- Write down the chemical formula for the formation of disulfide bonds in proteins. There will be no partial credit for this question. (3) **$RS-SR + R'SH \rightarrow R'S-SR + RSH$**
- True or **False**: Disulfide bonds are stable in the cytosol of the cell. (1)
- Draw two cysteines linked by a disulfide bond. (3)
(1) - give one point for drawing the sulfide bond (S-S)
(2) - give two points for correctly drawing two cysteines



Section II: DNA/RNA Structure, CRISPR-Cas 9 and Cytidine Deaminase

Directions: Answer the following questions about DNA/RNA, CRISPR-Cas 9 and Cytidine Deaminase. The point values for each question are addressed in the parentheses. (124)

1. RNA is one of the molecules that is an essential part of most living organisms. Answer the following questions about this nucleic acid. (9)
 - a. Describe how and **which** nitrogen of the pyrimidine base and nitrogen of purines base are bonded **which** carbon of pentose sugar. Be as specific as possible. (5) **1' nitrogen (1) of pyrimidine base (1) and 9' nitrogen (1) of purines base are bonded to 1' carbon (1) of the pentose sugar (1).** (Award two points if locations were not specified).
 - b. What bond forms in question (1a)? (1) **glycosidic bond**
 - c. How are the carbons of ribose numbered? (2) **1' to 5' (one point for each number)**
 - d. In which carbon is the hydroxyl group of the ribose sugar located? (1) **2'**

2. Answer the following questions about RNA structure. (8)
 - a. True or False: The nucleotides of RNA is linked by a 3'5' phosphodiester bridge. (1) **True**
 - b. What is a nucleoside? (2) **ribose sugar connected to a base through the 1'C**
 - c. The intrastrand pairing of bases can often result in what structure? (1) **hairpin loop**
 - d. Describe an "irregular pairing" in RNA. (3) **hydrogen bond donors/acceptors (1) that are not in Watson and Crick base pairs (AU and GC) (1) interact and hydrogen bond (1)**
 - e. Referring to the structure of RNA, why is it not as stable as DNA? (2) **the extra hydroxyl group attached to 2' C (anomeric C)**

3. DNA is another essential nucleic molecule in living organisms. Answer the following questions about its structure. (9)
 - a. True or False: A DNA nucleotide can have more than one phosphate group. (1) **True**
 - b. True or False: 2'-deoxyribose and phosphate groups are always present in DNA. (1) **True**
 - c. What is covalently bonded to the 5' of 2'-deoxyribose? (2) **phosphate group**
 - d. Why are purines and pyrimidines heterocyclic aromatic compounds? (2) **they both have N (nitrogen atoms) (1) in the C (carbon) based ring (1)**
 - e. How are the nitrogenous bases covalently bonded? Be as specific as possible. (3) **nitrogen atom (1) to 1' carbon (1) of deoxyribose sugar in a nucleotide (1)**

4. The following questions refer to the specific bonding characteristics in a DNA molecule. Answer as concise and specific as possible. (6)
- What type of bond joins two chains of polymerized nucleotides running side-by-side? (1) **hydrogen bond**
 - How many of the bonds in (4a) occur between A and T? (1) **2**
 - How many of the bonds in (4a) occur between C and G? (1) **3**
 - What rule explains the 1:1 ratio of pyrimidines and purines in any cell? (2) **Chargaff's rule**
 - True or False: The sugars of each DNA strand is located inside the DNA molecule. (1) **False**
5. DNA structure is divided in different levels of structure. Answer the following questions regarding these levels. (7)
- Describe the secondary structure of DNA. (2) **base pairing of molecules between two DNA strands of a single DNA molecule.**
 - Describe the tertiary structure of DNA. (2) **three-dimensional shape of a DNA molecule**
 - Is the DNA double helix left or right-handed? (1) **right**
 - Using the words parallel or perpendicular, how are the bonded bases on each strand stacked in relation to the sugar-phosphate backbone? (1) **the bonded bases are parallel in relation**
 - Using the words parallel or perpendicular, how do the bonded bases on each strand run in relation to the sugar-phosphate backbone? (1) **the bases runs perpendicular to the sugar phosphate backbone**
6. Chemical interactions are not only present in proteins but also in genetic material, such as DNA. The following questions refer to the chemistry behind DNA structure. (9)
- Are the bonds holding two DNA nucleotides together weak or strong? (1) **weak**
 - True or False: Each sugar phosphate backbone is said to have a polarity of 3'-to-5'. (1) **False**
 - Explain why each hydrogen atom in the NH_2 group has a slight positive charge. (3). **The nitrogen atom (1) has a tendency to attract the electrons in the N-H bond (1), leaving the hydrogen short of electrons (1).**
 - An oxygen atom has six unbonded electrons in its outermost shell. How does this affect the charge of the oxygen atom? (2) **Makes the charge slightly negative**
 - Why is the DNA molecule relatively stable even though it is held by weak hydrogen bonds? (2) **There are a lot of hydrogen bonds.**

7. Similar to DNA, the structure of RNA is also stabilized by chemical interactions. The following questions refer to the chemistry behind RNA structure. (4)
- What stabilizes the secondary structure of RNA? (1) **hydrogen bonds**
 - True or False: RNA can form double helices easier than DNA since it is more stable. (1) **False**
 - Why is RNA not the molecule that stores genetic information in humans? (2) **RNA is more unstable than DNA.**
8. CRISPR-Cas9 is an adaptive immune system used by bacteria and archaea for defense against phages and foreign plasmids. The following questions below is about this system. (6)
- What are the two components of CRISPR-Cas9? (2) **Cas9 nuclease and sgRNA (single artificial chimeric guide RNA)**
 - Complete the following statement regarding Cas9: Cas9 serves as the signature Type ____ gene. (1) **II**
 - Type IIC CRISPR systems contain how many *cas* genes? (1) **three**
 - There are about 1,000 Cas9 nucleases identified from UniProt in terms of homology. With this information, is protein length homogenous or heterogenous? (1) **Heterogenous**
 - What component of the CRISPR-Cas9 system is an alpha-helix-rich region that contains an Arg-rich bridge helix? (2) **REC domain**
9. The protospacer adjacent motif (PAM) is associated with Cas9 target range and search mechanism. The following questions below refers to PAM. (5)
- The PAM flanks at which end of the DNA target site? (1) **3'**
 - According to single-molecule imaging, the PAM sequences associate with what complex first before Cas9 initiates DNA strand separation? (2) **Cas9-crRNA-tracrRNA**
 - If PAM is used for engineering, would it generate short Cas9 orthologs with flexible 3'-NGG or 5'-NGG? (2) **5'NGG**

10. The following questions below are about the mechanism of how CRISPR-Cas9 works in bacteria. (7) **TB#4**
- What transcript is cleaved within the repeats by CRISPR-associated ribonucleases in CRISPR type I and III? (2) **pre-crRNA**
 - In what CRISPR type are crRNA intermediates further processed at the 3' end by RNases to produce a mature transcript? (1) **III**
 - In what CRISPR type do crRNA-tracrRNA hybrids complex with Cas9 to mediate interference? (1) **II**
 - What CRISPR type has a cascade complex loaded with a crRNA molecule that constitutes catalytically inert surveillance complex that recognizes target DNA?(1) **I**
 - In type I CRISPR, what is recruited to the cascade-bound R-loop and leads to the mediation of target degradation? (2) **Cas3 nuclease (accept Cas3)**
11. There have been instances to improve the fidelity of Cas9 target recognition. The following questions refer to this topic. (12)
- What component can be converted into a DNA “nickase” that can create a SSB? (1) **SpCas9**
 - How does the component in part (a) create a SSB? Describe the process. (4) **catalytically inactivating (1) the RuvC (1) or HNH nuclease domains (1) via point mutations (1)**
 - DNA SSBs are repaired via the BER pathway that is high-fidelity. What does BER stand for? (2) **base excision repair**
 - How would a double-nicking approach analogous to dimeric ZFNs or TALENs improve on-target DSB specificity? (4) **It can be used to increase the overall bases (2) that are specifically recognized in the target DNA (2).**
 - Would Cas9 nickases with single sgRNAs be used to mediate HR or NHEJ? (1) **HR**

12. Recently, Cas9 has been used as an engineering platform. The following cases below refer to Cas9's engineering uses. (10)
- a. Two Cas9 nickase complexes that have spaced target sites can mimic targeted DSBs through cooperative nicks. Based on this information, answer the following questions (3):
 - i. What happens to the length of target recognition? (2) **It doubles (2) (give 1 point for "gets longer")**
 - ii. Would cleavage efficiency be affected by this process? (1) **No, it would not.**
 - b. Some expression plasmids can be directly transfected into cell lines of interest. What promoter would be associated with this expression plasmid that also encodes the Cas9 gene and a short sgRNA cassette? (4) **U6 (1) RNA (1) polymerase (1) III (1) promoter**
 - c. Infer how Cas9 can be used in DNA imaging. (4) **Cas9 can be coupled to fluorescent receptors (2) that can be used for live imaging of DNA loci (2).**
13. CRISPR-Cas 9 is utilized by bacteria as a form of immunity. The following questions refer to this system and the immune system of bacteria. (9)
- a. In what phase of immunity are new spacers derived from the genome of the virus invading the bacteria? (2) **adaptation**
 - b. In what phase is the nucleolytic cleavage of an invading nucleic acid specified? (2) **targeting**
 - c. How many subtypes of CRISPR-Cas are there? (1) **12**
 - d. What is a protospacer? (2) **sequence in exogenous nucleic acid element (1) that corresponds to a CRISPR spacer (1)**
 - e. What CRISPR-Cas9 system primarily occur on the 3' end of the protospacer? (2) **Type II systems**
14. There are different types of CRISPR-Cas systems that have unique characteristics. The following questions refer to these types. (9) **TB#5**
- a. What type of CRISPR-Cas system cleave DNA molecules *in vitro*? (2) **Type III-B (1 point for Type III, 1 point for -B)**
 - b. Which CRISPR-Cas system requires the least Cas machinery for immunity? (2) **Type II**
 - c. What is generated when there is pairing between tracrRNA and precursor crRNA? (2) **double-stranded substrate**
 - d. What cleaves the structure generated in part (c)? (2) **RNase III**
 - e. What cleaves the DNA target in CRISPR-Cas Type II systems? (1) **Cas9**

15. Cytidine deaminase is the protein of the pre-build section of this season. The following questions refer to this protein's function and structure. (13)

- a. What element is this enzyme dependent on? (1) **Zinc**
- b. The deaminase of cytidine/deoxycytidine forms what compounds? (2) **Uridine (1) or deoxyuridine (1)**
- c. What amino acid is located in position 67? (2) **Glutamic acid**
- d. The K27 variant has an increased stability due to an ionic interaction between what two residues? Include the positions of the residues in your answer. (8)
Lysine residue (2) at position 27 (2) and glutamate residue (2) at position 24 (2)

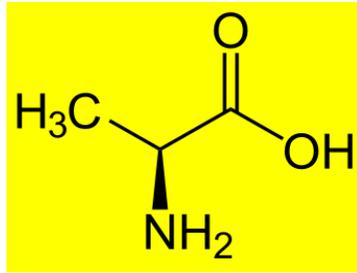
Section III: Computer Exploration

Directions: Use Jmol to explore the crystal structure of *Bacillus subtilis* cytidine deaminase with an Arg56 - Gln substitution (**1ux0.pdb**). The point values are addressed in the parentheses. Be as concise and specific as possible in your answers. Complete sentences are **not** required. (44)

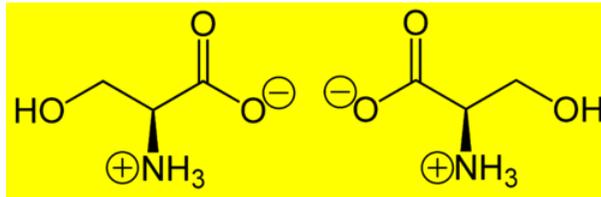
1. The following questions refer to **residue 39**. (6)
 - a. What amino acid is residue 39? (1) **Arginine**
 - b. At a physiological pH, is the carboxylic acid protonated or deprotonated? (1) **deprotonated**
 - c. At a physiological pH, is the amino group protonated or deprotonated? (1) **protonated**
 - d. Is the 3-carbon straight chain of the side group of this amino acid aromatic or aliphatic? (1) **aliphatic**
 - e. What happens to the chain this amino acid is located in if the amino acid is replaced with tyrosine? (2) **Since tyrosine is nonpolar, the protein chain would move from the exterior and move more towards the interior (hydrophobic core).**

2. The following questions refer to **residue 24**. (8)
 - a. What amino acid is residue 24? (1) **Phenylalanine**
 - b. Why is this amino acid nonpolar? (2) **inert and hydrophobic nature (1) of the benzyl side chain (1)**
 - c. Which isomer is used to biochemically form proteins? (1) **L-isomer**
 - d. This amino acid is a precursor for what amino acid? (2) **tyrosine**
 - e. What is the formula of this amino acid? (2) **C₉H₁₁NO₂**

3. The following questions refer to **residue 88**. (7)
 - a. What amino acid is residue 88? (1) **Alanine**
 - b. True or False: In biological conditions, this amino acid appears in its zwitterionic form. (1) **True**
 - c. True or False: This is the simplest amino acid. (1) **False**
 - d. True or False: The side-chain of this amino acid is non-reactive. (1) **True**
 - e. Draw the L-isomer of this amino acid. No partial credit will be awarded for this question. (3)



4. The following questions refer to **residue 49**. (8)
- What amino acid is residue 49. (1) **Serine**
 - Because the side chain of this amino acid has hydroxymethyl, is this amino acid polar or nonpolar? (1) **Polar**
 - Draw the zwitterionic form of this amino acid. Include the S and R version. (6) **S-version is the one on the left. R version is the one on the right. 3 points per version. No partial credit granted.**



5. The following questions refer to the **entire** protein structure. (5)
- How many hydrophobic atoms are in this structure? (1) **970**
 - How many hydrophilic atoms are in this structure? (1) **0**
 - How many charged atoms are in this structure? (1) **508**
 - How many atoms are in sheets? (1) **440**
 - When the structure is restricted to water molecules, how many **atoms** are listed in the console of Jmol? (1) **145**
6. How many hydrogen bonds are in the nucleic acid portion of this complex? (2) **0**
7. How many hydrogen bonds are in the entire protein structure? (2) **59**
8. How many hydrogen bonds are in **69-101** amino acids? (2) **22**
9. How many struts can you make in the entire protein structure? (2) **39**
10. How many struts can you make in the entire nucleic structure? (2) **0**