

## C - Protein Modeling - C - Nov 7 Country-wide SO Practice - 11-07-2020

Hello all competitors! This test contains Part 2 (Computer Exploration) and Part 3 (Written Exam) of the Protein Modeling event (Total of 180 points).

Try to format your answers as instructed to facilitate auto-grading. We will, however, review all answers to make sure points are properly given for all correct answers.

IMPORTANT: The Scilympiad platform will not penalize you for using the offline version of Jmol, so don't be afraid to open Jmol application and leave the test window.

### Post-competition notes:

Question 27 is not scored due to missing picture.

### PART 2: COMPUTER EXPLORATION

For this part, load the 6vxx PDB file using Jmol by clicking File > Get PDB and entering "6vxx" at the prompt.

If this is not available, you can download from PDB website.

Make sure you open the website in a new window (Right click, Open link in new window)

<https://www.rcsb.org/structure/6vxx> (<https://www.rcsb.org/structure/6vxx>) (<https://www.rcsb.org/structure/6vxx>)

After using "Download files" button, please close the PDB web site and return to scilympiad window, or to Jmol.

1. (2.00 pts) How many total residues are there in this PDB file?

2. (2.00 pts) For chain A, how many atoms are there from residue 319 to 514?

3. (2.00 pts) For chain A, how many Alpha helix are there between residue 319 and residue 514?

4. (3.00 pts) What is the residue 511:A - N atom location in (x, y, z)?  
Enter the (x,y,z) coordinate, with three decimal points, in the box below in order.  
Make sure there is no extra space after your answer.

5. (2.00 pts) What is the type of protein for residue 511?

- A) Alanine  
 B) Valine  
 C) Glycine  
 D) Lysine

6. (2.00 pts) What is the secondary structure type that residue 511 belong to?

- A) Alpha helix
- B) Beta turn
- C) Beta sheet
- D) Acidic side chain

7. (2.00 pts) How many Hydrogen bonds are there **within** the residue 319-511 on chain A?

0

8. (1.00 pts) The sequence between residue 350 to 500 is a more stable configuration of such secondary structure.

- True
- False

9. (2.00 pts) What is the bonding type between residue 560 and residue 567?

- A) Disulfuric bond
- B) Hydrogen bond
- C) Ionic bond
- D) Hydrophobic force
- E) Peptide bond

10. (2.00 pts) This spike protein of SARS-CoV2 has two units:  
First is a \_\_\_\_\_ S1 subunit, which forms of the globular head of the S protein,  
and the \_\_\_\_\_ S2 region that forms the stalk of the protein.  
Select the correct phase to fill in the blank above regarding the overall protein structure.

- A) N terminal, C terminal
- B) P terminal, C terminal
- C) C terminal, N terminal
- D) C terminal, P terminal

11. (2.00 pts) Identify the residue type at position (188.904 189.161 249.845), in its three-letter name.

ARG

12. (6.00 pts) Identify the AA connected to the sidechain of the residue in previous question. List them in order of their residue number.

TYR

ASP

PHE

13. (3.00 pts) List the three-letter names of residues 603, 604, 605 in that order.

ASN

THR

SER

14. (2.00 pts) Based on the attached molecules at residue 603, what is functional name for the sequence above?

glycosylation sequon

15. (3.00 pts) Based on the connection identified in the question above, list two or more key roles of residue 603 in this protein.

attracts carbohydrates

determines folding or confo

specific recognition

16. (3.00 pts) If such sites are abundant in subunit of SARS-CoV-2, how does it influence its transmissibility?

**Expected Answer:** Increase binding affinity of SARS\_Cov-2 virus to human ACE2 receptor (host cell receptor) (2pts) Glycan shielding may also limit recognition by the immune response of infected hosts (1pt)

### Part 3.A. CRISPR, Cas9, APOBEC3A (Written Section)

17. (2.00 pts) In what type of organisms can CRISPR/Cas system be found naturally?

- A) Eukaryotes
- B) Prokaryotes
- C) Viruses
- D) Humans

18. (2.00 pts) What is Cas9 and what does it do?

- A) an RNA molecule that binds to target DNA via complementary base pairing
- B) a DNA sequence that binds the Cas9 protein
- C) a viral protein that disrupts bacterial membranes
- D) a protein enzyme that cuts both strands of DNA at sites specified by an RNA guide

19. (4.00 pts) What are the possible outcomes following DNA cleavage by Cas9?

(Mark **ALL** correct answers)

- A) Disruption of a gene
- B) Repair of a gene
- C) Deletion of a chromosome
- D) Disruption of more than one gene

20. (2.00 pts) In base editing, deamination of Cytosine (C) is catalysed by cytidine deaminases to create

- A) Thymine (T)
- B) Adenine (A)
- C) Uracil (U)
- D) Guanine (G)
- E) None of the above

21. (2.00 pts) What type of enzyme is APOBEC3A?

- A) Transfrase
- B) Hydrolase
- C) Isomerase
- D) Ligase

22. (2.00 pts) The protein APOBEC3A is a protein encoded by the APOBEC3 gene in which organism?

- A) Streptococcus pyogenes
- B) Mycobacterium tuberculosis
- C) Haloferax volcanii
- D) Homo sapiens
- E) Saccharomyces cerevisiae

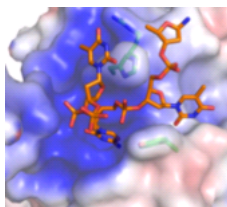
23. (3.00 pts) In the third-generation base editor (BE3), Uracil DNA glycosylase inhibitor (UGI) was fused to the C terminus of BE1-version of Cytidine deaminases, What is the benefit of this design?

**Expected Answer:** Improve editing efficiencies in human cells (2 pts), favor U-A conversion (1pt).

24. (2.00 pts) In all the Cytidine base editor designs by Professor Liu's group at Harvard, a C-G pair is always converted to what base pair eventually?

T-A

25. (3.00 pts) When A3A (short for APOBEC3A) binds to DNA, a snapshot of the DNA binding site is shown below. Specify in first box what the color contour is for. In second and third boxes, specify what the red and blue represent, respectively.



surface electrical potential

negative

positive

26. (2.00 pts) Which mutation in wildtype A3A-BE3 causes the complex to exhibit the most specificity for the target cognate site over the bystander site?

- A) N57A
- B) N57G
- C) N57Q
- D) Y130F
- E) N57Q and Y130F

27. (2.00 pts) What does the magenta-colored sphere represent in the picture below (of A3A), and what is its role in the deamination process?

Zn ion

Catalyst

28. (2.00 pts) N57 site of A3A is central to the recognition of ssDNA (target nucleotide base) because

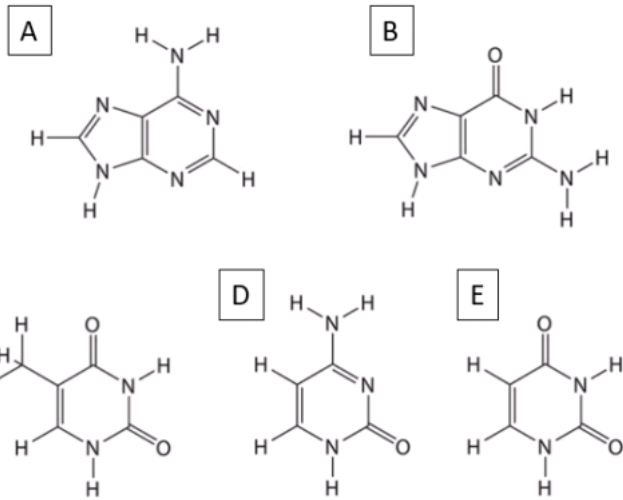
(Mark ALL correct answers)

- A) Side chain of N57 determines the 5'-3' direction of binding by forming hydrogen bond with target
- B) N57 side chain forms a hydrogen bond with the backbone to position it to hydrogen bond to target
- C) N57 has extensive van der Waals contacts with thymine bases near by
- D) N57 has a salt bridge to nearby helix, thus stabilizes the overall hydrogen bonding configuration
- E) N57 side chain packs against the deoxyribose ring of target, stabilizing the orientation of its sugar plane
- F) N57 is an Asn, therefore it acts as a catalytic active site

29. (5.00 pts) Describe the functions of the C-terminal domain, N-terminal domain, and the groove of APOBEC proteins.

**Expected Answer:** C-terminal domain: catalytic site, catalytically active, deaminates cytidine (2 pts) N-terminal domain: pseudo-catalytic site, catalytically inactive, binds to nucleic acids (2 pts) Groove: binds to ssDNA (1 pt)

30. (2.00 pts) Which of the following is the correct nucleobase substrate for Cytidine deaminase?

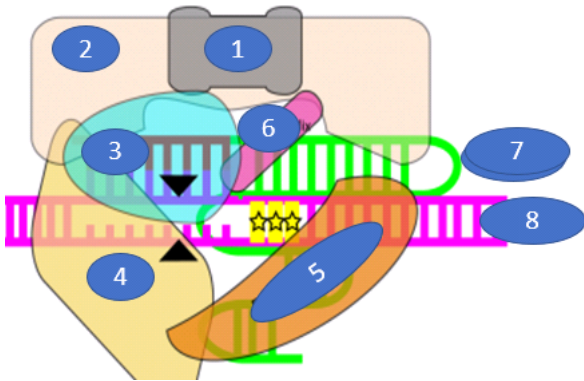


D

31. (2.00 pts) The CRISPR Cas9 system is

- A) A bacterial immune defense system
- B) Composed of protein structures alone.
- C) Composed of protein and RNA components
- D) Both A and B
- E) Both A and C

The CRISPR Cas9 is illustrated in the diagram below with labels for key conformational and recognition domains. Answer the eight following questions with the abbreviated name or description of each domain.



32. (1.00 pts) Block 1

REC II

33. (1.00 pts) Block 2

34. (1.00 pts) Block 3

35. (1.00 pts) Block 4

36. (1.00 pts) Block 5

37. (1.00 pts) Block 6

38. (1.00 pts) Block 7

39. (1.00 pts) Block 8

40. (3.00 pts) The stars in the diagram above represents PAM.  
Specify the three letter Codon sequence for PAM in *S. Pyogenes*, starting with 5' direction. One letter in each box.

41. (3.00 pts) Among the eight domains marked above, which two are critical in performing the function of cleaving RNA?  
List first the domain responsible for cutting nearest to PAM.

In last box, specify how many bp upstream of PAM the nearest cleavage occurred.

### Part 3.B Protein Structures

42. (2.00 pts) Which of the following interactions make protein folding possible (the tertiary structure of the protein)?

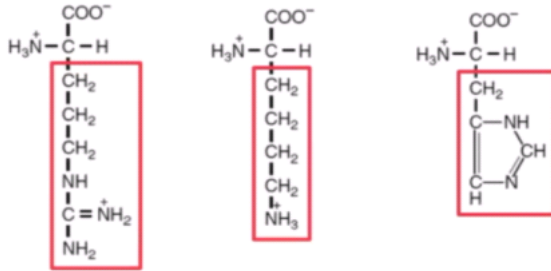
(Mark ALL correct answers)

- A) Peptide bonds
- B) Hydrogen bonds
- C) Ionic bonds
- D) Disulfide bridges
- E) Hydrophobic interactions

43. (2.00 pts) The inactivation and subsequent degradation of unwanted proteins is most commonly achieved in a process known as:

- A) Phosphorylation/Dephosphorylation
- B) Heat Shock Response
- C) Ubiquitination
- D) Autoinhibition
- E) Pp ppl

44. (6.00 pts) Image below shows three Amino Acids that are of polar type. Write the full name and 1-letter abbreviation in parentheses in order. Example of format: Glycine (G).



Arginine (R)

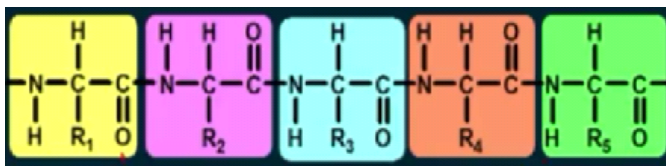
Lysine (K)

Histidine (H)

45. (2.00 pts) There are five different types of tight turns found in the secondary structure of proteins. Which of the following turns involves the presence of four amino acid residues?

- A)  $\alpha$ -turn
- B)  $\beta$ -turn
- C)  $\gamma$ -turn
- D)  $\delta$ -turn
- E)  $\pi$ -turn

46. (2.00 pts) Following is a bonding pattern of the alpha helix: carbonyl oxygen of the n-th residue bound with amide proton of the (n+4)-th residue. What type of bonding does this show?



Hydrogen bonding



47. (4.00 pts) Helix propensity of an AA is a measure of the likelihood for the amino acid to be in a helix. Which of the following have high propensities?

(Mark ALL correct answers)

- A) Glu
- B) Gly
- C) Ala
- D) Leu
- E) Met
- F) Proline

48. (1.00 pts) How many residues per turn is there in a normal alpha helix?

3.6

49. (2.00 pts) Which two amino acids are most likely to appear in beta turns? Use single letter representation. Fill in the blank in alphabetical order.

G

P

50. (2.00 pts) Which of the following peptides would be most likely to form a stable alpha helix?

- A) TKAGQESTDVVEW
- B) NRATLDMAKVQTA
- C) TDHQPNGAYLWA

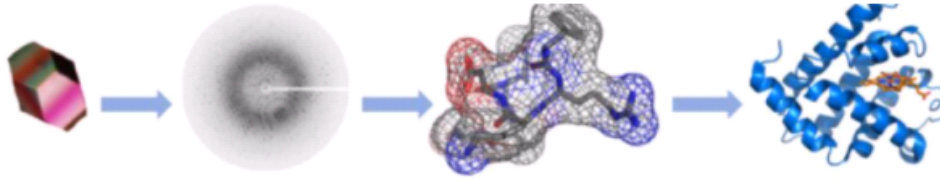
51. (3.00 pts) Explain why gelatin, which is mostly collagen, is nutritionally inferior to other types of dietary protein.

**Expected Answer:** The amino acid composition of collagen mostly consists of Gly, Pro, and different Pro derivatives. Good sources of protein usually contain a much greater variety of amino acids.

52. (6.00 pts) A mutation that changes an **Ala** residue in the interior of a globular protein to a **Val** is found to lead to a loss of activity. However, the activity of the protein is restored when a second mutation at a different position changes an **ile** residue to **gly**. How might this second mutation lead to a restoration of activity?

**Expected Answer:** Ala to Val causes a greater hydrophobicity and more steric strains because of the increased number of atoms in Val compared to Ala. [2 pts] Ile to Gly creates a lesser hydrophobicity and there is less steric due to fewer number of atoms in Gly compared to Ile. [2 pts] The switch to Gly cancels out the changes caused by switched to Val so the protein restores to its normal state. [2 pts]

53. (2.00 pts) Image below shows a typical techniques for deciphering the protein structure. What is the technique called?



X-ray diffraction

54. (2.00 pts) Cysteine is an important amino acid because it can form \_\_\_\_\_ that connect two Cysteines together.

- A) hydrogen bonds
- B) van de Waal bonds
- C) salt bridges
- D) disulfide bridges

55. (2.00 pts) Which of the following amino acid is most likely to occur in the transmembrane domain of a protein?

- A) Pro
- B) Glu
- C) Lys
- D) Leu
- E) Arg

56. (2.00 pts) Which of the follow residue is most unlikely to occur in a beta turn?

- A) Gly
- B) Ser
- C) Pro
- D) Asn
- E) Trp

57. (2.00 pts) Which of the following mutation has the biggest impact for the tertiary structure of a protein?

- A) ile -> leu
- B) ser -> thr
- C) val -> pro
- D) asp -> glu
- E) ile-> lys

58. (2.00 pts) In the Ramachandran plot, what type of secondary structure would have a positive psi and positive phi angle?

- A) Right turn Alpha helix

- B) Antiparallel beta sheet
- C) Beta turn
- D) Parallel beta sheet
- E) Left turn Alpha helix

59. (4.00 pts) Select the correct statement(s) about beta sheet from below.

(Mark ALL correct answers)

- A) Made up of more than two peptides
- B) Antiparallel beta sheets occur more frequently in nature than parallel ones
- C) Beta sheet can form between two different protein molecules
- D) Peptides in beta sheet is more extended compared to that in Alpha helix
- E) Sidechains of amino acids on the same beta sheet align on one side

### Part 3.C Protein, Genes and Biochemistry

60. (2.00 pts) A study in 2015 showed that a full-length crRNA targeting the human VEGFA gene also cleaves 21 off-target sites (unintended error). The human genome contains 3.2 billion base pairs. Approximately at what frequency of the off-target effect did they observe?

- A) 1 in 15,000
- B) 1 in 1,600,000
- C) 1 in 150,000,000
- D) 1 in 1,600,000,000

61. (2.00 pts) The pKa value for Alanine are pKa1=2.34, pKa2=9.66. What is the **pI** value for Alanine?

62. (2.00 pts) Glutamic acid (E) has three ionization states with pKa1=2.19, pKa2=9.67 and pKa3=4.25. What is the **pI** value for Glu?

63. (2.00 pts) How is histidine affected by physiological pH condition?

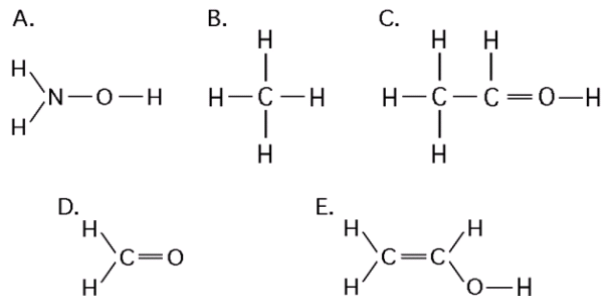
- A) Physiological pH is approximately equal to pKa of histidine just to stabilize substrates at enzyme active sites.
- B) Physiological pH is approximately equal to pKa of histidine, so it can exist in both protonated and deprotonated forms.
- C) Physiological pH is much greater than pKa of histidine, allowing histidine to exist in its deprotonated form.
- D) Physiological pH is much less than pKa of histidine, allowing histidine to exist in its deprotonated form

64. (3.00 pts) Identify the amino acid attached to the tRNA, if the **anticodon** found on such tRNAs are: 5'-AGU-3', 5'-AUG-3', 5'-CUG-3'. Fill the boxes of the full name in the same order.

65. (2.00 pts) Which two functional groups are **always** found in amino acids?

- A) ketone and aldehyde
- B) carbonyl and carboxyl
- C) carboxyl and amino
- D) phosphate and sulfhydryl
- E) hydroxyl and aldehyde

66. (2.00 pts) Which of the following structures illustrate an impossible covalently bonded molecule?



- A)
- B)
- C)
- D)
- E)

67. (2.00 pts) Which of the following statements regarding enzymes is true?

- A) Enzymes increase the energy required for a chemical reaction.
- B) Enzymes increase the rate of a reaction.
- C) Enzymes change the direction of chemical reactions.
- D) Enzymes change the direction of chemical reactions.
- E) Enzymes prevent changes in substrate concentrations.

68. (2.00 pts) The active site of an enzyme is the region that

- A) is involved in the catalytic reaction of the enzyme
- B) binds the products of the catalytic reaction by absorbing water
- C) adds water to the substrate during dehydration synthesis
- D) maintains the induced fit in extreme pH environments
- E) both b and c

69. (2.00 pts) Which of the following is synthesized 3' to 5'?

- A) The leading strand

- B) The lagging strand
- C) mRNA
- D) A & C
- E) None of the above

70. (2.00 pts) An activation domain and a DNA binding domain are found in which of the following?

- A) Enhancer
- B) TFIIIB
- C) Insulators
- D) Transcriptional activators
- E) RNA polymerase

71. (3.00 pts) How many different polypeptides of 61 residues can be made from the 20 naturally occurring amino acids?  
Express your answer in scientific notation, with first box the number (base) in single digit, and second box with the power of 10.

72. (4.00 pts) The amino group of Lys has a pKa of 10.5.  
What fraction of these groups will be protonated in a dilute solution of Lys at pH =9.5?  
Write your answer in percentage with two significant figures.

73. (4.00 pts) Calculate the net charge of **LCYRAIDG** at a pH of 1.5. Be sure to indicate positive or negative.

74. (6.00 pts) Calculate the approximate **pI** of the peptide **DSYRLKCF**, to two significant figures.

75. (2.00 pts) A protein has the amino acid sequence: **DSRLSKTMSIEAPAKLEQNMAL**  
How many peptide fragments would result from cleaving the sequence with trypsin?

Congratulations! You finished the Protein Modeling test!