Welcome to the **University of Michigan 2021 Invitational Protein Modeling** test! Make sure you have a stable internet connection and are ready to compete!

For this test, you are allowed the following resources:

- A Google Meet/Zoom/Skype/Phone/Video call with your partner
- A cheat sheet/binder, printed or in pdf format on your computer
- Programmable/Non-programmable calculator
- Scratch paper

You **MAY NOT** take advantage of the following resources. Doing so will result in a disqualification plus 30 points added to your team's overall score.

- ANY internet resource (other than linked resources within the test)
- Help from any person other than your partner
- A printed version of the test

This event requires a pre-build. You may view the pre-build instructions here https://docs.google.com/document/d/1NQ8rOeijZXYB54KCKXodD6gwaHsJwxqf2Rb8jLAisk8/edit?usp=sharing
and submit here:https://forms.gle/wFaiaKLLim6FoPnq9 . If you have not already done so, please turn in your pre-build using this google form by the END OF YOUR TIME SLOT.

This test consists of 75 questions and you will have 50 minutes to complete it.

The Tiebreakers for this test will be: **Part 2 Q20, Part 3 Q50, Q55, Q61, Q66**

If you experience technical difficulties during the test:

- Immediately contact the event supervisor through the classroom feature on Scilympiad, stating clearly what issue you are having.
- If your work is not saving/submitting, take screenshots of your answers on Scilympiad and submit them to this google form. Try to stay within your allotted 50 minutes.

This event requires a pre-build. You may view the pre-build instructions here [EMBED LINK TO PDF/GOOGLE DOC]. If you have not already done so, please turn in your pre-build using this google form by the END OF YOUR TIME SLOT.

PART 1: PREBUILD - 30 points

The protein that the competitors will model is residues 31-145 of Chain A of the APOBEC3A Cytidine Deaminase Protein as indicated by the rules and the MSOE (national sponsor for protein modeling).

Students will submit a document containing images of their prebuild model along with a notecard detailing their creative additions through a google form: https://forms.gle/wFaiaKLLim6FoPnq9

Further instructions about what images should be submitted can be found on this google document: https://docs.google.com/document/d/1NQ8rOeijZXYB54KCKXodD6gwaHsJwxqf2Rb8jLAisk8/edit?usp=sharing

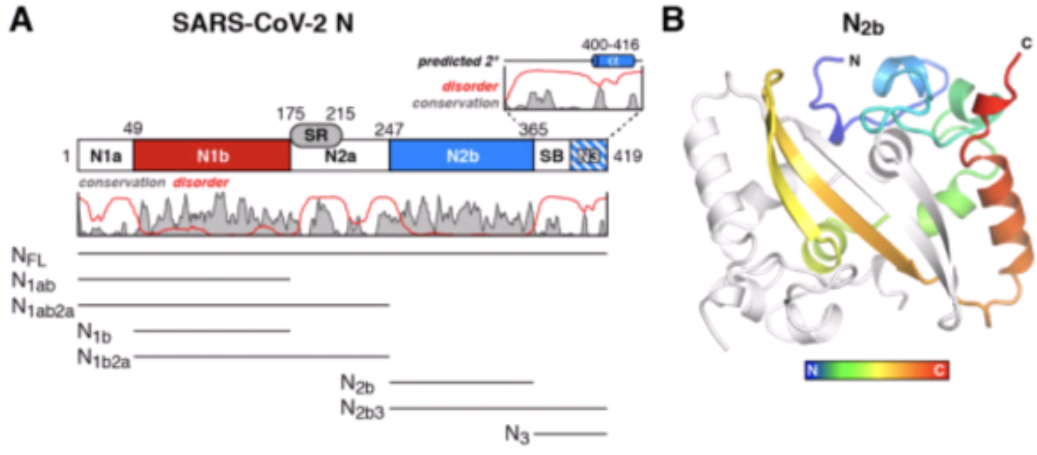
PART 2: JMOL EXPLORATION - 30 points

"In December 2019, a new type of coronavirus (SARS-CoV-2 or 2019-nCoV) causing a novel pneumonia now named COVID-19 broke out in Wuhan, China. The SARS-CoV-2 genome is composed of approximately 30,000 nucleotides, which encodes four structural proteins including spike (S) protein, envelope (E) protein, membrane (M) protein, and nucleocapsid (N) protein. The N protein is a highly immunogenic and abundantly expressed protein during infection. Furthermore, the N protein is frequently used in vaccine development and serological assays. At present, there are few reports focusing on SARS-CoV-2 N protein, and the updated understanding of SARS-CoV-2 N protein is in urgent need."

* Zeng, Weihong et al. "Biochemical characterization of SARS-CoV-2 nucleocapsid protein." Biochemical and biophysical research communications vol. 527,3 (2020): 618-623. doi:10.1016/j.bbrc.2020.04.136,https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7190499/

"As the most prominent surface component of the virus, the spike protein is the major target of antibodies in patients, and is the focus of several current efforts at SARS-CoV-2 vaccine development. Initial trials using antibody-containing plasma of convalescent COVID-19 patients has also shown promise in lessening the severity of the disease.While the above efforts target viral entry, RNA synthesis, and protein processing, there has so far been less emphasis on other steps in the viral life cycle. One critical step in coronavirus replication is the assembly of the viral genomic RNA and nucleocapsid (N) protein into a ribonucleoprotein (RNP) complex, which in betacoronaviruses like SARS-CoV-2 is thought to form a helical filament structure that is packaged into virions through interactions with the membrane-spanning membrane (M) protein. Despite its location within the viral particle rather than on its surface, patients infected with SARS-CoV-2 show higher and earlier antibody responses to the nucleocapsid protein than the surface spike protein. As such, a better understanding of the SARS-CoV-2 N protein's structure, and structural differences between it and N proteins of related coronaviruses including SARS-CoV, may aid the development of sensitive and specific immunological tests."

*Ye, Qiaozhen et al. "Architecture and self-assembly of the SARS-CoV-2 nucleocapsid protein." *bioRxiv : the preprint server for biology* 2020.05.17.100685. 17 May. 2020, doi:10.1101/2020.05.17.100685. Preprint. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7263487/



Now that you have some background information of the nucleocapsid protein of the SARS-CoV-2 genome, we will delve deeper into the structure of SARS-CoV-2 Nucleocapsid dimerization domain.

Protein Bank: https://www.rcsb.org/structure/6WZO
Research/Experiment: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7263487/
Online version of JMOL: https://cbm.msoe.edu/markMyweb/jmolDesignEnvironment/#forward

To answer the following questions, you have access to the above protein bank link, the pdb file on jmol (available for download in the protein bank link), and the research/experiment paper.
To access the structure on jmol, press the tab labeled "Load a new structure" >> type in 6WZO in the second blank labeled "Public .PDB file (by ID)" >> press load

However keep in mind that not all questions refer to the above resources.

1. What resolution (Å) did the researchers use in this experiment to identify the structure of SARS-CoV-2 Nucleocapsid dimerization domain? (1 pt)
   1.42 (must be exact for all points since research paper explicitly states this value)
2. List at least one advantage and one disadvantage of using X-ray Crystallography in an experiment (be specific)? (2 pts)
   Advantages: High resolution up to 0.5 angstroms, No protein mass limit
   Disadvantages: physical crystals needed, Structure is static average -- hence limited information provided about protein dynamics.
   One point for advantage and one point for disadvantage.

Use the resolution values (Å) provided below to answer the following four multiple choice questions. Answer with the minimum resolution value necessary to observe the following situations:

I) 1.0

II) 1.5

III) 2.5

IV) 3.5

V) 4.0

VI) 4.5

3. Hydrogen atoms are visible (0.5 pts)
    a. I
    b. II
    c. III
    d. IV
    e. V
    f. VI
4. Secondary structures well defined and visible (0.5 pts)
    a. I
    b. II
    c. III
    d. IV
    e. V
    f. VI
5. Global Fold (limited view of secondary structures) (0.5 pts)
    a. I
    b. II
    c. III
    d. IV
    e. V
    f. VI
6. All phi-psi angles are well defined and visible (0.5 pts)
    a. I
    b. II
    c. III
    d. IV
    e. V
    f. VI

Now for the following questions refer to either the JMOL version of the protein or the RCSB Protein Bank:

7. How many chains are present in the nucleoprotein? (1 pt)
    4

The following questions refer to chain A of the nucleoprotein:

8. How many beta strands are present? (1 pt)
   2
9. Classify beta strand(s) as either parallel or antiparallel? Briefly state which classification is more stable and explain why at an atomic level? (2 pts)
   Both beta strands are antiparallel (1pt). Antiparallel strands are stronger because hydrogen bonding is optimal in comparison to parallel strands as the double bonded oxygens line up with the hydrogen atoms -- closer hence more stable (1pt).
   Answer must include atomic reasoning to receive full credit

10. How many alpha helices are present? (1 pt)
    5
11. There are other secondary structures present in chain A (True or False) -- answer with a capital T for true or a capital F for false. (0.5 pt)
    T
12. Based on the answer to the previous question, name the secondary structure. If you answered false, enter "N/A". (0.5 pt)
    3-10 Helix
13. Indicate the x, y, z coordinates of the C-terminus (first blank -- x coordinate, second blank -- y coordinate, third blank -- z coordinate) (1.5 pts)
    (9.534 , 8.026 , 14.218)
14. Indicate the x, y, z coordinates of the N-terminus (first blank -- x coordinate, second blank -- y coordinate, third blank -- z coordinate) (1.5 pts)
    (-19.185 ,16.351 ,6.902)
15. Which amino acid serves as a rotamer outlier in chain A? (indicate with full amino acid name) (1 pt)
    Arginine
16. How many hydrogen bonds are present? (1 pt)
    186
17. How many sulfur atoms are interacting with chain A? (1 pt)
    3

The following questions require the use of the research article linked at the beginning of the section:

18. Typically nucleocapsid proteins have a general common structure with the presence of an ordered RNA binding domain (N1b) and a dimerization domain (N2b).

a. The dimerization domain (N2b) of SARS-COV-2 is most similar to the dimerization domain of which structure? (1 pt)

The structure is particularly similar to that of SARS-CoV, with which the N2b domain shares 96% sequence identity.

All or nothing graded question.

b. List the residual differences between both structures. (2 pts)

Only five residues differ between these proteins' N2b domains (SARS-CoV Gln268 → SARS-CoV-2 A267, D291 → E290, H335 → Thr334, Gln346 → Asn345, and Asn350 → Gln349)

0.5 points for each residual difference listed.

c. Why do you think these differences/modifications are significant? (2 pts)

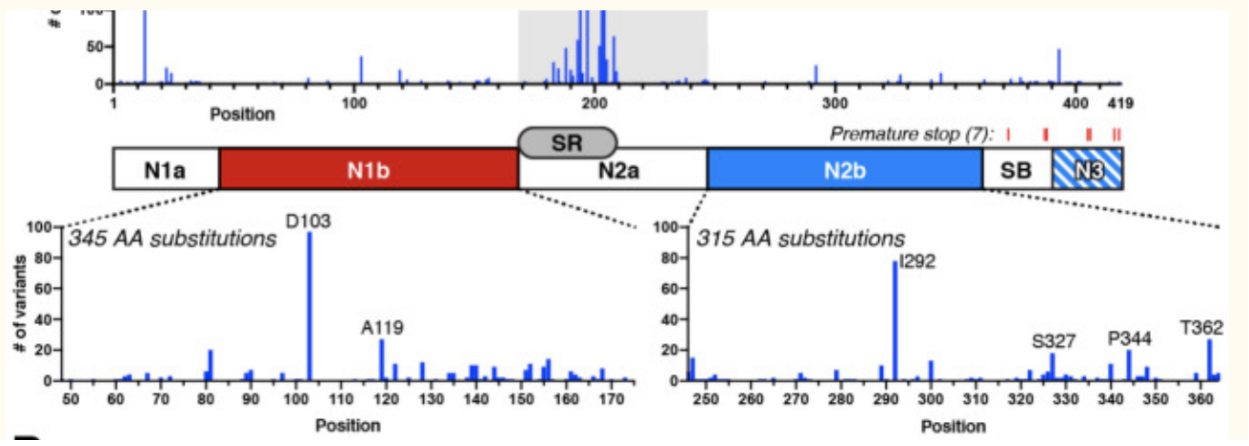
Differences are significant -- answer should mention changing residues, changes targeting practices (relate to big picture of why disease is not exactly the same and mutationally different thus potentially resistant).

19. Which residues are most commonly mutated in the N2b dimerization domain? (2 pts)

S327, I292, T362, P344

0.5 points per residue

20. What are the functionalities of the N2b domain, C-terminal spacer B/N3, N2a linker domain. What is the main drawback of targeting the N2a domain with individual patient antibodies? What conclusions can be made? (6 pts) *Tiebreaker 1*



While the N2b domain mediates dimer formation, we find that addition of the C-terminal spacer B/N3 domain mediates formation of a robust homotetramer. N2a linker domain links N2b and N1b domains and ensures no steric clashes. N2a domain is uniquely tolerant of mutations, in keeping with its likely structural role as a disordered linker between the RNA-binding N1b domain and the N2b dimerization domain. (2pts)

The high variability of the N2a domain means that individual patient antibodies targeting this domain may not be reliably detected with tests using the reference N protein; especially if these antibodies recognize residues 203 and 204, which are mutated in a large fraction of infections. (2 pts)

Need to Proceed with caution -- Given the early and strong antibody responses to the nucleocapsid displayed by SARS-CoV-2 infected patients, the distribution of mutations within this protein should be carefully considered as antibody-based tests are developed. (2pts) -- for their conclusion

PART 3 - GENERAL EXAM - 40 points

We have divided the general exam into three different sections, each covering a topic indicated by the National Science Olympiad Rules:

Section 1: Protein Folding Chemistry

Section 2: CRISPR Mechanisms

Section 3: APOBEC3A Cytidine Deaminase Protein

There are different formats for questions, please follow the directions. Tiebreakers will also be indicated

Section 1: Protein Folding Chemistry

1. A right handed helix is typically how long in angstroms (Å)? And each residue corresponds to a rise of _____ (Å). (0.5 pt)
   5.4, 1.5
2. What is the length of a polypeptide (angstroms) with 150 amino acids in a helical shape (make up an alpha helix)? (0.5 pt)
   225 angstroms
3. Define chirality (0.5 pt):
   Chirality describes the ability of a molecule to present with unique L and D enantiomers -- creating mirror images of the structure. A molecule is chiral if it cannot be superimposed on its mirror image.

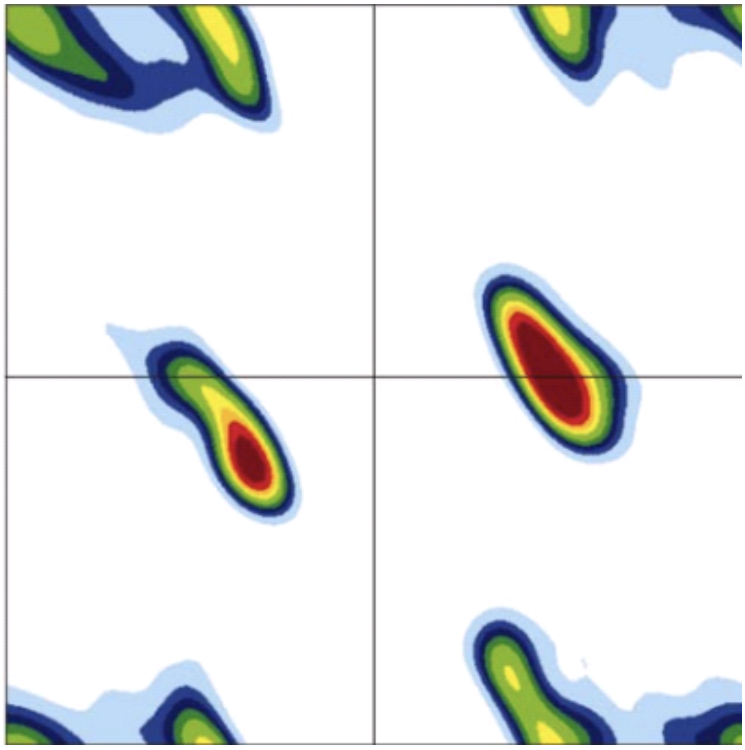
4. Which of the following is a chiral amino acid (circle all that apply) (0.5 pt)
   a. Proline
   b. Glycine
   c. Histidine
   d. Tryptophan
   e. Alanine
   f. All of the above

5. Which of the following is an achiral amino acid (circle all that apply)? (0.5 pt)
   a. Proline
   b. Glycine
   c. Histidine
   d. Tryptophan
   e. Alanine
   f. All of the above
6. Based on the answers to question 4 and question 5, provide structural evidence to explain why the chosen amino acids are chiral and achiral respectively. (1 pt)
   Glycine is achiral because its mirror image is the same, in other words their L and D enantiomers are identical -- thus Glycine is not chiral. Answer can include Proline, Histidine, Tryptophan, or Alanine when describing chiral amino acids. These amino acids are chiral because their L and D enantiomers are different since mirror image would create a flipped structure.

7. What effect does chirality have on protein folding? (1 pt)

Chiral molecules have less protein chain flexibility and lower protein folding entropy.

8. Name this plot (first blank), define the horizontal axis (second blank), define the vertical axis (third blank) -- spell out/do not use symbols (0.5 pts)

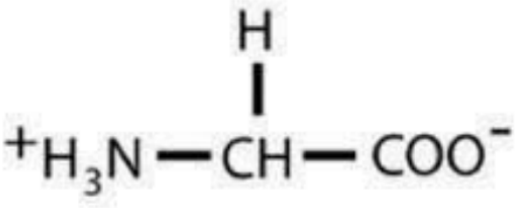


Ramachandran Plot, phi, psi

9. From the above plot image, which amino acid does it most closely resemble? (0.5 pts)

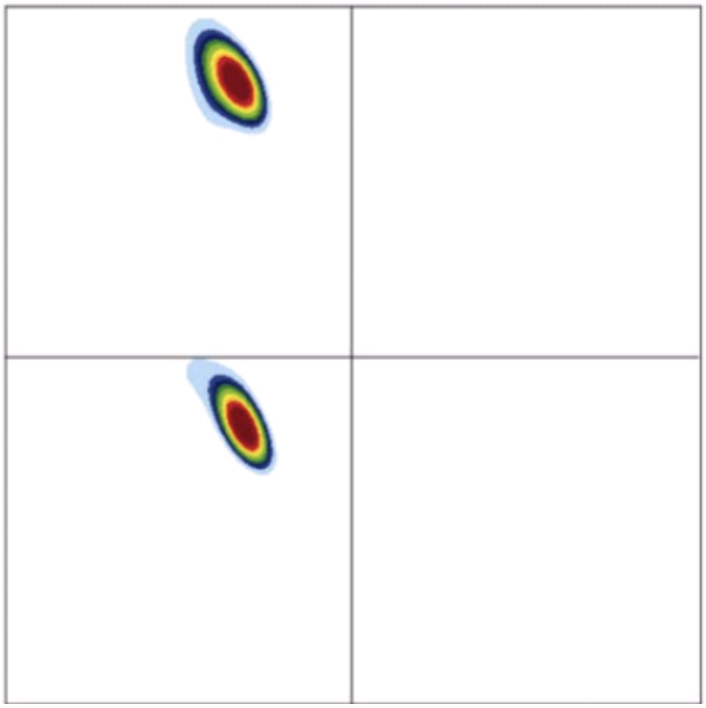
Glycine

10. Briefly describe this amino acid (explain atomic structure in words), detail any important properties relevant to protein folding? (1 pt)

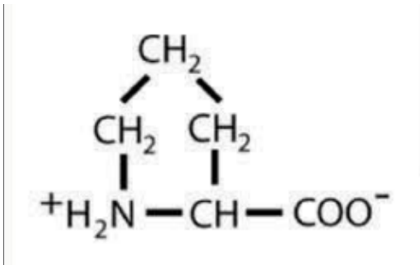


Description should match:

Glycine only has one hydrogen which makes it really flexible to have a set dihedral angle (0.5 pts).
Glycine is hydrophobic (0.5 pts).



11. What amino acid does this plot resemble? (0.5 pts)
    Proline
12. Briefly describe this amino acid (explain atomic structure in words), detail any important properties relevant to protein folding? (1 pt)

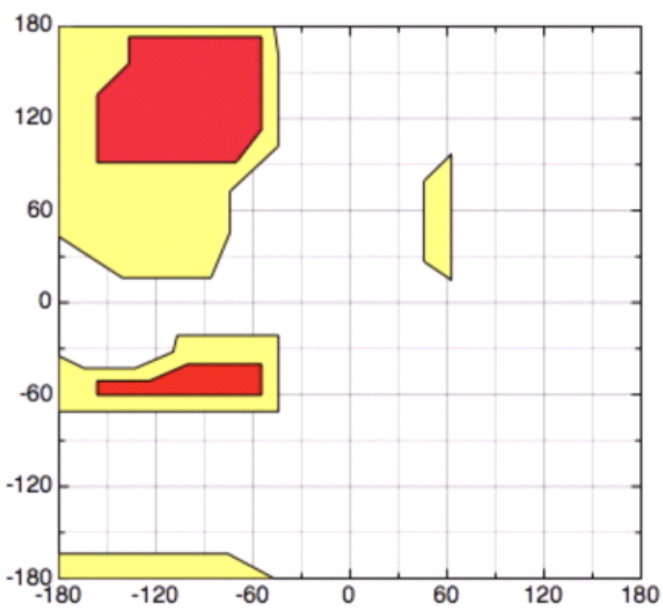


Description should match:
Proline has a 5 membered ring and is the only cyclic amino acid. (0.5 pts)
Proline is also hydrophobic. (0.5 pts)

13. What is the significance of these two amino acids in each of the secondary structures (alpha helices, beta sheets, beta turns)? (1 pt)

Proline and Glycine are frequently found in beta turns, proline because its cyclic structure is ideally suited for the beta turn, and glycine because, with the smallest side chain of all the amino acids, it is the most sterically flexible.
Glycine is an intrinsically destabilizing residue in β sheets. In natural proteins, however, this destabilization can be 'rescued' by specific cross-strand pairing with aromatic residues.
Proline acts as a structural disruptor in the middle of regular secondary structure elements such as alpha helices and beta sheets
Glycine also tends to disrupt helices because its high conformational flexibility makes it entropically expensive to adopt the relatively constrained α-helical structure.

14. Given this image, explain the differences between the red, yellow, and white colored areas. (1 pt)



White: Impossible areas, Yellow: some steric hindrance/resistance, Red: No steric constraints

15. If a plot had some density in the lower left quadrant what secondary structure would you expect to be marked? (0.5 pts)
Right-Handed Alpha Helix
16. Ribonuclease is a small protein responsible for splitting and controlling RNA strands. It also contains cysteine molecules connected via disulfide bonds.

What two components when combined can denature this protein (ignore general salts) (0.5 pts)?

Urea and 2-mercaptoethanol

17. When these two components are added, ribonuclease denatures -- describe this denatured state. (1 pt)

ribonuclease is unfolded and catalytically inactive. The disulfide cross links are reduced with the presence of urea and 2-mercaptoethanol which produces cysteine residues.

18. If the two components are taken out at the **same time**, ribonuclease returns to its original structure and disulfide cross-links are correctly reformed. (0.5 pts)
    a. True
    b. False

19. Which of the following protein precipitation techniques are valid (choose all that apply) (0.5 pts):
    a. Salting out
    b. Hydrophilic polymers
    c. Using acetone as a solvent
    d. Isoelectric precipitation
    e. All of the above

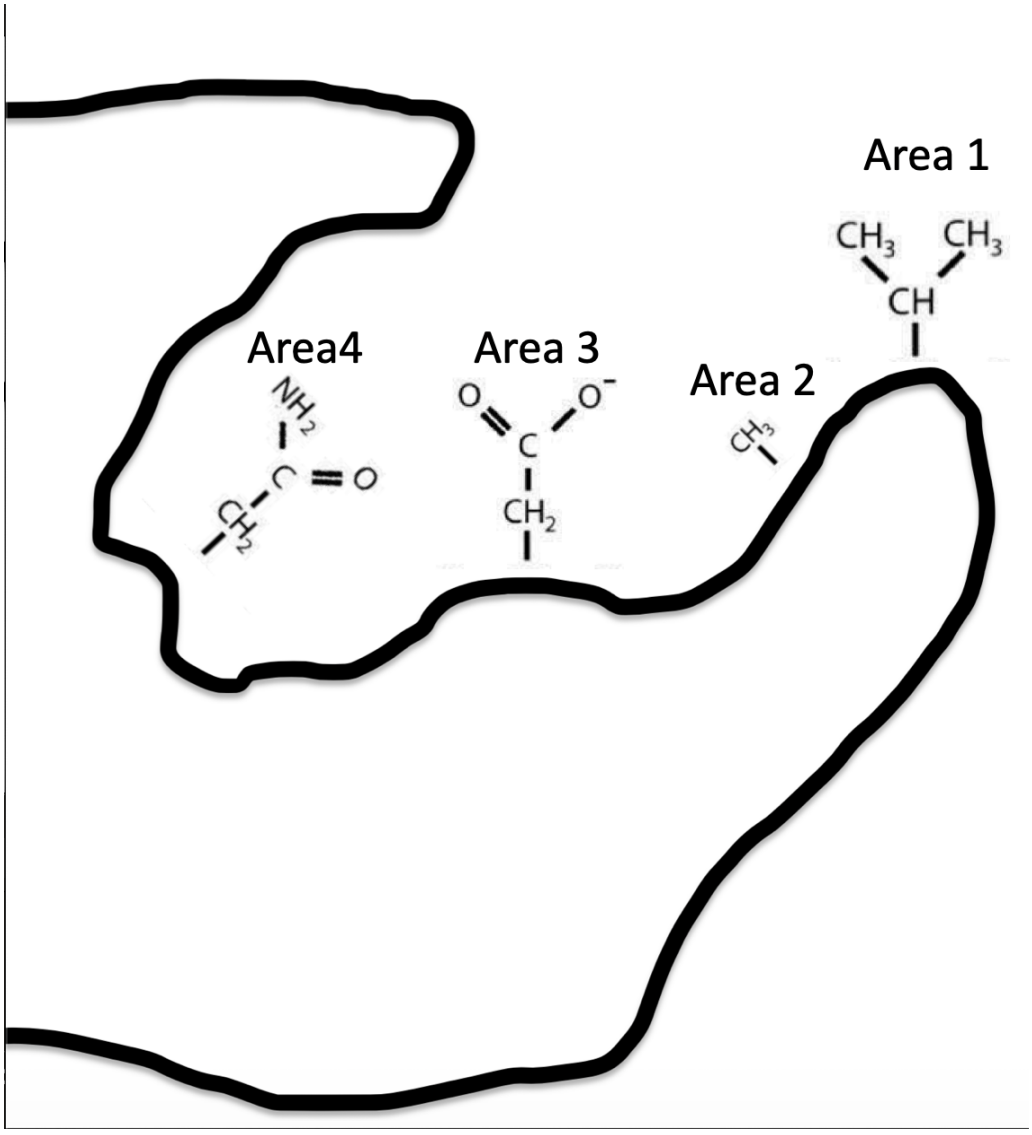
20. Which amino acid can form disulfide bonds? (0.5 pts)
    a. Cysteine
    b. Tyrosine
    c. Glycine
    d. Glutamine

e. All of the above

21. Name two components that can be used in a cleaving reaction of disulfide bridges? (0.5 pts)
Thiolates and B-mercaptoethanol

A pharmaceutical drug company has recently uncovered a protein-substrate complex. The picture below details the rough active site for the protein. Use the picture below to answer the following

set of questions (23 - 31):



22. What amino acid is present in Area 1? (0.5 pts)
    a. Leucine
    b. Isoleucine
    c. Valine
    d. Serine
    e. Alanine
23. What amino acid is present in Area 2? (0.5 pts)
    a. Glycine
    b. Proline
    c. Serine
    d. Lysine
    e. Alanine
24. What amino acid is present in Area 3? (0.5 pts)

a. Glutamic Acid
b. Aspartic Acid
c. Arginine
d. Histidine
e. Asparagine

25. What amino acid is present in Area 4? (0.5 pts)
    a. Glutamic Acid
    b. Lysine
    c. Aspartic Acid
    d. Asparagine
    e. Glutamine

For the purposes of this question, the specific substrate amino acids involved with this active site and complex is unknown.

26. What is the strongest interaction/force that can exist between the Area 2 amino acid (on the drug surface) and the substrate surface? (1 pt)
    a. Hydrogen Bonds
    b. Van der Waal Forces
    c. Ionic Bonds
    d. Covalent Bonds
27. What is the strongest interaction/force that can exist between the Area 3 amino acid (on the drug surface) and the substrate surface? (1 pt)
    a. Hydrogen Bonds
    b. Van der Waal Forces
    c. Ionic Bonds
    d. Covalent Bonds
28. What is the strongest interaction/force that can exist between the Area 4 amino acid (on the drug surface) and the substrate surface? (1 pt)
    a. Hydrogen Bonds
    b. Van der Waal Forces
    c. Ionic Bonds
    d. Covalent Bonds

Amazed by your knowledge of proteins and their mechanisms, the pharmaceutical company hires you to help design a potential protein that would bind more effectively and fit better within the active site of the image above. On your first day of work, you're faced with three questions:

29. Would Serine or Isoleucine (on the substrate surface) allow for a greater interaction with Area 4? (0.5 pts)
    a. Serine

b. Isoleucine

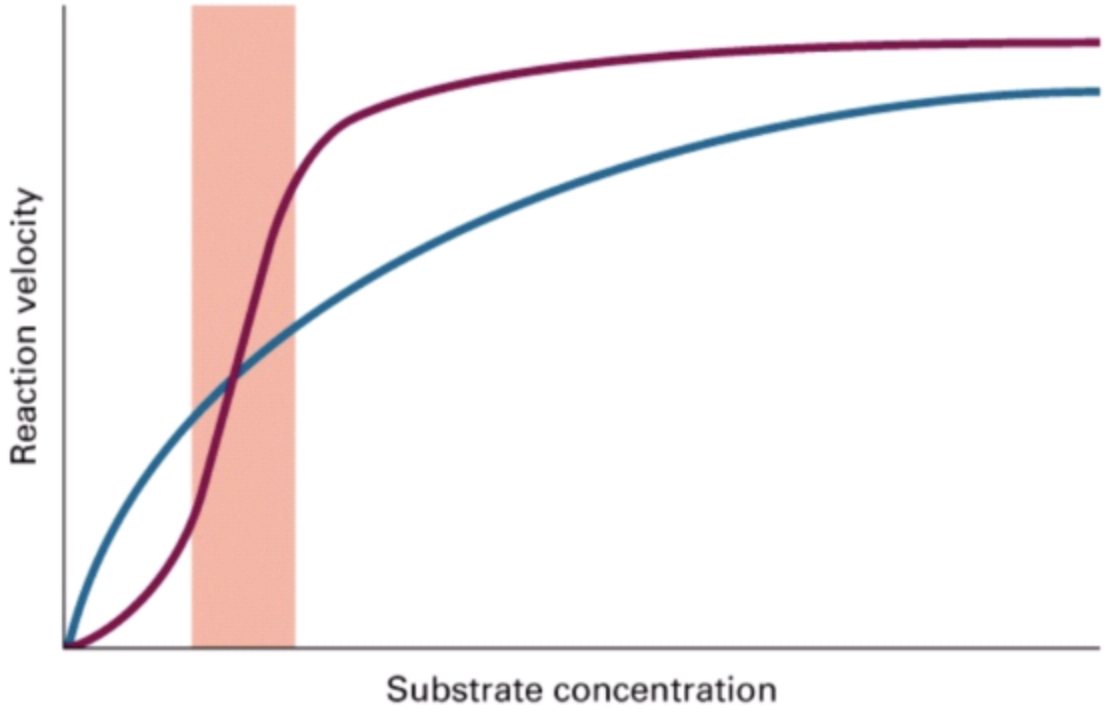
30. Would Glutamic Acid or Lysine (on the substrate surface) allow for a greater interaction with Area 3? (0.5 pts)
    a. Glutamic Acid
    b. Lysine
31. Explain your answer choices in further detail from both part a and part b. (2 pts) *Tiebreaker 2*

Serine can form hydrogen bonds and is a polar amino acid while the strongest interaction that Isoleucine can form are van der Waal forces.
Lysine is positively charged and thus can form ionic bonds with Area 3 -- Though Glutamic Acid is also charged, it is negative and thus will repel with Aspartic Acid.

32. During an enzyme reaction, if the enzyme concentration is decreased by a factor of 5, how will the Km value change correspondingly? (0.5 pts)
    a. Km decreases by a factor of 5
    b. Km decreases by a factor of 2.5
    c. Km increases by a factor of 5
    d. Km increases by a factor of 2.5
    e. Km is unchanged
33. An enzyme has a Vmax of 50 mM product formed per min. The Km for the substrate of the enzyme is 10 mM. What is the initial reaction rate (v0) when [S] is 5.0 mM? (0.5 pts)
    a. 41.6 mM per second
    b. 16.6 mM per second
    c. 47.3 mM per second
    d. 27.6 mM per second
    e. 33.3 mM per second
    f. 14.9 mM per second

34. Which of the following combinations of choices will describe an enzyme with high catalytic efficiency? (0.5 pts)
    a. Low kcat, High Km
    b. High kcat, High Km
    c. High kcat, Low Km
    d. Low kcat, Low Km

35. Select the most accurate statement regarding the purple curve (0.5 pts):



a. Enzyme shown is allosteric because graph of substrate concentration and velocity is sigmoidal
b. Enzyme shown is not allosteric because graph of substrate concentration and velocity is sigmoidal
c. Enzyme shown is allosteric because graph of substrate concentration and velocity follows the standard Michaelis-Menten curve
d. Enzyme shown is not allosteric because graph of substrate concentration and velocity follows the standard Michaelis-Menten curve

36. Describe the formation and structure of an Enzyme-substrate complex (be sure to address specific interactions and proximity) (2 pts) *Tiebreaker 3*

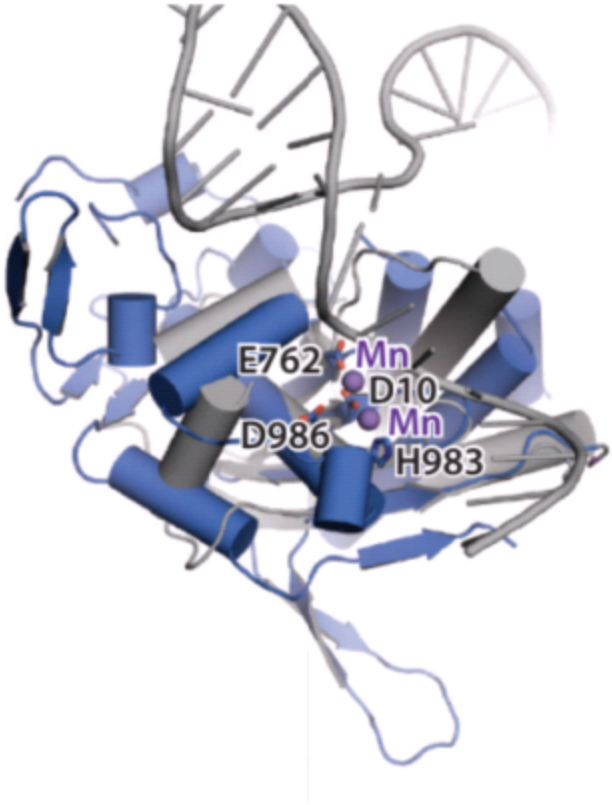
An enzyme-substrate complex is a combination of the enzyme and the substrate in which the two are bound together very closely so that atoms on each are essentially in physical contact with each other. The physical contact regions involve H-bonding, ionic bonds, hydrophobic interactions, and occasionally, covalent bonds.

Section 2: CRISPR Mechanisms

37. What are the three main steps for the CRISPR - Cas9 immune response (name of step followed by description of step)? (1 pt)

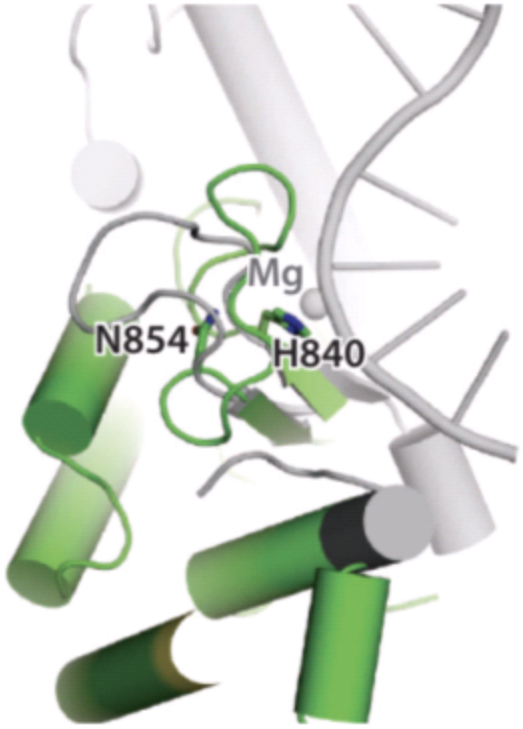
Adaptive stage: spacer acquisition, the processed foreign DNA (known as the protospacer) is integrated into the CRISPR array locus, yielding a new spacer. 2. crRNA expression and processing stage involves transcription of the CRISPR locus into a single pre-CRISPR RNA (pre-crRNA) and further processing into mature crRNAs 3. In the interference stage, a single Cas protein (or complex) uses the crRNA as a guide to cleave phage nucleic acid or plasmid bearing a complementary sequence to the spacer sequence of the crRNA.

38. Name the domain involved in the CRISPR/CAS9 system diagram below: (0.5 pts)



RuvC

39. Name the domain involved in the CRISPR/CAS9 system diagram below: (0.5 pts)



HNH

40. What are the functions of the two domains above and how do they differ? (1 pt)

    HNH cleaves target strand, RuvC cleaves non-target strand, different cleaving mechanisms

41. Which one of the above domains is most likely to use a two-metal-ion catalytic mechanism for cleavage? (0.5 pts)
    RuvC

42. What is the main marker to differentiate one-metal-ion-dependent and two-metal-ion-dependent nucleic acid cleaving enzymes in terms of conserved residues? (1 pt) *Tiebreaker 4*
    The main marker of one-metal-ion-dependent and two-metal-ion-dependent nucleic acid cleaving enzymes is the presence of a conserved general base histidine and an absolutely conserved aspartate residue (108), respectively.

43. Compare the two types of DNA repair involved with cleavage within the Cas9 system. (1 pt)

Homology directed repair, Nonhomologous end joiningNon-homologous end joining (NHEJ) is a pathway that repairs double-strand breaks in DNA. NHEJ is referred to as non-homologous

because the break ends are directly ligated without the need for a homologous template, in contrast to homology directed repair, which requires a homologous sequence to guide repair. The DSB created by Cas9 is then repaired either by error-prone nonhomologous end joining (60), resulting in small random insertions and/or deletions (indels) at the cleavage site, or by high-fidelity homology directed repair (86), resulting in precise genome modification at the site of the DSB using a homologous repair template

44. Which one is more accurate? (0.5 pts)
    HDR (Homology directed repair)
45. What are three other uses for the two types of DNA repair other than cleavage? (1 pt)
    Full points for three of the following:
    Gene targeting (dCas9–effector):
    – transcriptional regulation
    – epigenetic modification
    – live-cell imaging
    – nucleotide editing
    Genetic screens/drug screens, Ligation-mediated gene editing by double Cas9 nickases (D10A)

46. Describe how crRNA, tracrRNA, sgRNA, and gRNA are related to one another as well as their overall within the CRISPR/Cas9 system. (1 pt) *Tiebreaker 5*

crRNA contains the target sequence while tracrRNA helps link crRNA to Cas9. Both the crRNA and tracrRNA are fused in the composition of the sgRNA -- sgRNA accepts the gRNA

47. The native PAM sequence for the commonly used SpyCas9 is (0.5 pts):
    5'-NGG-3'
48. How many different main types (not including subtypes) of CRISPR systems exist? (0.5 pts)
    a. 2
    b. 4
    c. 6
    d. 8
    e. 18

49. What are the differences between type I and type II CRISPR systems? (0.5 pt)

Class 1 systems rely on interference complexes called Cascade, using multi-subunit complexes. Class 2 systems rely on single effector proteins such as Cas9.

50. Which type/system of CRISPR can Cas9 be found in? (0.5 pts)
    a. Type I
    b. Type II
    c. Type III
    d. All of the above
    e. None of the above

Section 3: APOBEC3A Cytidine Deaminase Protein

51. What role does H29 play within the protein? (1 pt)
    the unique role of H29 in positioning the substrate ssDNA with a series of coordinated hydrogen bonds and stacking interactions, essentially latching the ssDNA and the target dC0 within the active site.

52. What ion is present within the active site of the Cytidine Deaminase protein -- explain its significance? (1 pt)
    Zn(2+), facilitates interactions around it (ligand with Cl, etc.)

53. Name the motif (present in the central nucleotides) bound at the active site along with the target deoxycytidine (dC0)? (0.5 pts)
    5'-TCT-3'
54. What is one drawback of using cytidine base editors (explain why)? (1 pt)
    cytidine base editors create off-target edits (sometimes inaccurate)

55. Four different cytidine deaminase enzymes were evaluated for ssDNA deamination. Which of the following demonstrated the highest activity of deaminase? (0.5 pt)
    a. Human APOBEC3G
    b. Lamprey CDA1
    c. Rat APOBEC1
    d. Human AID

Congratulations on completing the University of Michigan 2021 Invitational Protein Modeling test!

If you have not done so already, please use this google form to turn in pictures of your prebuild model. This must be submitted before the timeslot ends -- links are at the top of the test.

If you have any questions or concerns pertaining to this event, please email tec.umichscioly@umich.edu, and we will try to get back to you as soon as we can.