Protein Modeling - Division C Exam

University of Texas-Austin Invitational October 23-24, 2020

Written by Sameer Rajesh (UC Berkeley), Nithin Parsan (MIT)

Team Number: _____

Team Name: _____

Total Points: _____/319 **Rank:** _____

Instructions and Clarifications:

- You have **50** minutes to finish this exam and there is no computer exploration of protein structure. This exam accounts for **100%** of the event score.
- Each **participant** may use **one** 8.5" x 11" sheet of paper (for online exams, this can be in digital format)
- Please read the detailed instructions below and make sure to follow them

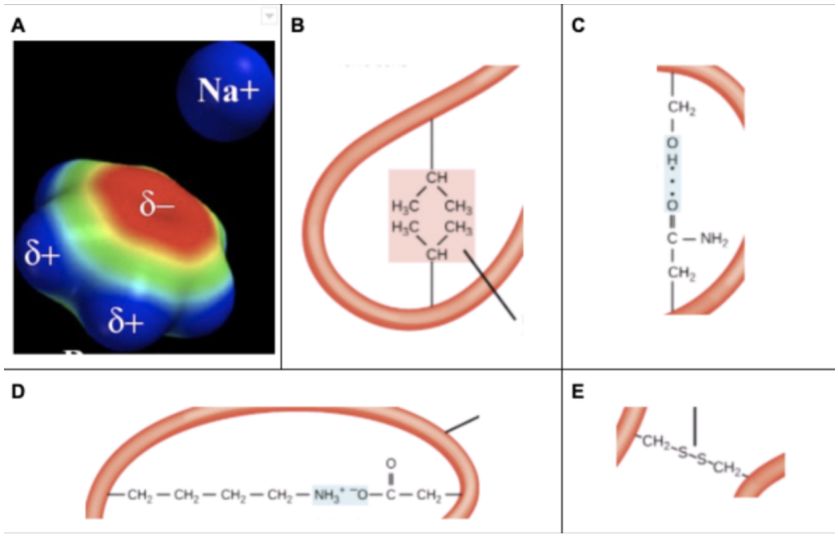
Instructions: Please read

- The last 20 questions are far easier than most questions on the test–we encourage you to take a look at those early on so you can get as many points there as possible. There is no penalty for skipping any questions
- Many FRQ questions may have a particular designated answer format–please follow this format! If you do not, you will not be awarded points. Examples follow:
    - When asked for full name, 3 letter code, or one letter code of a Nucleotide or amino acid, provide your answer in the requested format. Answering "Glycine" when only 1 letter code is requested will result in 0 points.
    - Do not explain answers unless asked to explain (less reading for graders)
    - When asked to explain an answer, limit responses to between 1 and 3 sentences. Overly long "essays" which fail to capture the essence of the problem and answer will be penalized–there is no benefit to you writing every detail you can think of, only mention the most salient aspects of the situation. Answers do not need to be grammatically correct or in full sentences, they just need to be comprehensible.
    - Dont use the internet.

Feel free to reach out regarding any questions, comments or concerns about the test or answer key, as well as any feedback you may have, to sameer.rajesh6626@berkeley.edu or nparsan@mit.edu.

Section 1: Protein Biochemistry

1. Based on the diagram below, identify the likely interactions from the word bank. Fill in the blank with the appropriate lowercase letter from the below list.



1) HYDROPHOBIC INTERACTIONS
2) DISULFIDE BOND
3) ELECTROSTATIC INTERACTIONS
4) CONJUGATION
5) PI-STACKING
6) HYDROGEN BONDING
7) CATION-PI
8) ENTROPIC SHELTERING

(a) (2 points) Interaction A:

(a) 7

(b) (2 points) Interaction B:

(b) 1

(c) (2 points) Interaction C:

(c) 6

(d) (2 points) Interaction D:

(d) 3

(e) (2 points) Interaction E:

(e) 2

2. (3 points) You identify a peptide in a new species of archaebacteria that displays a unique helix due to the presence of different amino acids. Your professor tells you that these AAs are chemically identical to standard AAs, but are stereochemically different. Which of the following helix structures might you observe?

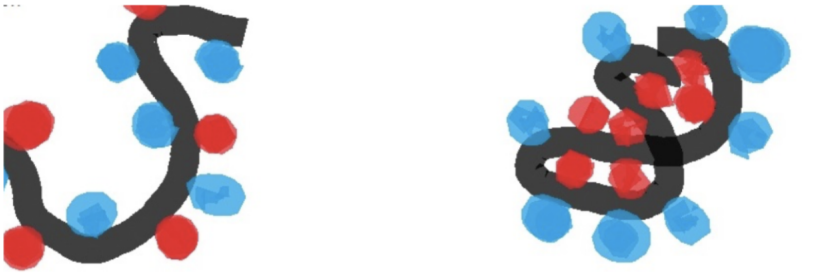
A. Left handed π helix

B. Right Handed α helix

C. Z form

D. Right handed β sheet

3. (3 points) Which of the following phenomena does the diagram below represent?



A. Molten Globule

B. Levinthal's Paradox

C. Hydrophobic Collapse

D. Electrostatic Precipitation

4. (3 points) A scientist isolates a sample of dsRNA from a cell and attempts to find proteins which bind it. He uses transcription factors which generally bind to dsDNA double helices, but has no luck. Why? (TB1)

A. Transcription Factors dont bind nucleic acids

B. RNA does not form double helices

C. RNA adopts an A-form helix when double stranded, and its major groove is incapable of admitting helices in for sequence recognition

D. The minor groove of dsRNA is clustered by zinc ions and thus prevents protein binding

5. (3 points) Leucine zippers are a typical example of what type of structure?

A. Beta Barrel

B. Greek Key

C. Coiled Coil

D. Hairpin Loop

6. (3 points) A scientist has found a high copy number of kinases in a cell which has many membrane receptor domains. Which of the following amino acids might be most related to kinase activity?

A. Alanine

B. Glutamine

C. Lysine

D. Tyrosine

7. (5 points) Which of the following residues would most favorably interact with the C terminal end of an alpha helix at the water-protein interface?

   A. Lysine

   B. Glutamate

   C. Tryptophan

   D. Ornithine

8. (5 points) DNA is found to have a persistence length equal to about 150 base pairs, which means isolated DNA of length 150 base pairs or less will not bend, and instead act like a rigid rod. Despite this, DNA segments of about this length are easily wrapped more than once around histones. How is this possible?

   A. DNA base pairs separate and single stranded DNA molecules easily wrap around histones to form the nucleosome

   B. Positively charged residues such as lysine are able to facilitate wrapping by forming favorable electrostatic interactions with the DNA as it wraps around histones

   C. Polar residues such as threonine intercalate between base pairs to generate favorable linking interactions between histones and DNA

   D. DNA winding around histones is a topological effect which can be explained by modifications of the Lennard-Jones potential

9. (5 points) The major groove of B form DNA allows for sequence-specific recognition of DNA by protein alpha helices, which are often found in transcription factors. Suppose a scientist has identified a transcription alpha helix to interact with the major groove edge of the base pair A-T. She identifies two amino acids that interact with the bases (one with adenine, one with thymine). Which of the following residues will provide the least specific interaction with either of these two base pairs?

   A. Serine

   B. Glutamine

   C. Asparagine

   D. Threonine

10. (5 points) A scientist generates a solution NMR to study the structure of a protein. Using a computer program, he generates a list of all the dihedral angles for each of the residues. Upon plotting these on a Ramachandran diagram, he finds a large concentration of residues to lie outside of the allowed regions on the diagram. He attributes this to the proteins ability to withstand extreme pH values since the protein was isolated from an acidophile. Which of the following statements is true?

    A. The scientist is wrong because proteins cant withstand extreme pH so this structure is a denatured form of the protein

    B. The scientist is wrong because the allowed regions of Ramachandran plots are not affected by non-physiological pH values and temperatures

    C. The scientist is right because forbidden Ramachandran regions can be explored by peptides at extremely low pH values due to denaturation.

    D. The scientist is right because the allowed ranges of van der Waals contact are expanded due to more protonated residues at the low pH value

11. (5 points) The pKa of the sidechain of aspartic acid is around 4. At a pH of 5, what is the expected value of the net charge of aspartic acid side chains in solution? (Ignore charges on the backbone carbonyl and amine groups).

You may find the following definition of expected value useful. For a quantity Q with n different allowed values given by q_n , each of which occurs with probability p_n the expected value of Q is given by

$$\langle Q \rangle = \sum_n q_n p_n$$

where the sum ranges over all possible different values of n and the quantity being summed is the product of a particular value of Q and the probability that value occurs.

A. -1

B. $-\frac{10}{11}$

C. $-\frac{1}{2}$

D. $-\frac{1}{11}$

12. (5 points) A scientist designs a short polypeptide (8 residues) in such a way that its optimal folding dihedral angles correspond to regions within the alpha-helix portion of a Ramachandran diagram. When she dissolves the peptide in water, which of the following will occur?

A. The protein will rapidly sample many different potential conformations before settling on a helical structure

B. The protein will rapidly fold into the alpha helical structure without sampling intermediates

C. The protein will not fold efficiently

D. The protein will hydrolyze due to instability in the peptide bonds

13. (5 points) Another scientist working with the same peptide from the previous question now dissolves the peptide in Carbon Tetrachloride, a nonpolar solvent. Which of the following will happen?

A. The peptide will fold into a globular shape around a solvent molecule to create a hydrophobic core

B. The peptide will aggregate

C. The peptide will react with the carbon tetrachlroide molecules due to the reaction of an amide group with a Carbon-Chlorine bond

D. The protein will fold into an alpha helix

Beta Sheets

The last few multiple choice questions of this section will be on beta sheets. Consider the following model of a beta strand.

Here, we are looking at the beta-strand edge-on. Blue atoms are nitrogen, black are carbon, red oxygen, white hydrogen, and green represent R groups. Note the up-down orientation of the R groups. This beta strand is oriented in an N-C direction. Multiple beta strands can form a beta-pleated sheet in such a way that their backbones hydrogen bond with each other and that their R groups interact with each other. Beta sheets can be either antiparallel or parallel.

14. (5 points) Suppose there are two beta strands in a beta sheet (directionality is not important for this question). Their R groups are oriented in the same up-down pattern, so that there is cross strand R group interaction. If the lengths of R groups that interacted with each other across the strands were not equal to each other, which of the following would be true?

    A. The beta-sheet would be more stable due to the induced twist strengthening as a result of compensating for the R group length mismatch

    B. The beta-sheet would be less stable because the length mismatch will weaken cross strand interaction

    C. The beta-sheet will be equally stable because the primary factor in determining stability is inter-backbone Hydrogen bonding, which is unaffected

    D. The beta sheet would be less stable because the length mismatch will deform the sheet and cause rapid unfolding

15. (5 points) Suppose a scientist designs a peptide which contains three beta strand regions and two long flexible linker between each pair. The sequence of this peptide in one letter codes is N INIQELN — Linker — INIQKLN — Linker — NLDQINL C. Suppose this protein, when put into water, folds into a beta-sheet along with two linkers. Label the strands in order from N to C terminus as Strand 1, Strand 2, and Strand 3. Which of the following is likely true about the beta-sheet structure? Assume L and I are roughly the same size/shape.

    A. Strand 1 and 2 are parallel, and antiparallel to Strand 3

    B. Strand 1 and 2 are antiparallel, Strand 1 and 3 are parallel

    C. All three strands are parallel

    D. Strand 1 and 2 are antiparallel, Strand 2 and 3 are parallel

16. (5 points) Suppose a mutation caused the lysine in strand 2 and the aspartate in strand 3 to be mutated to a glutamate residues. Which of the following must change for the solution if the peptide is still to form beta sheets?

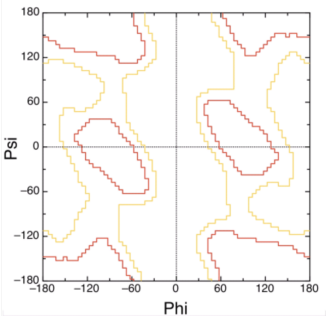
    A. pH should increase

    B. pH should decrease

    C. Temperature should increase

    D. Temperature should decrease

17. This problem will focus on the Ramachandran Diagram. This is a staple of all biochemistry exams. Below is a plot for a particular amino acid.



    (a) (2 points) This amino acid can be incorporated well into β-sheets.

    A. True

    B. False

(b) (2 points) This particular amino acid is not natural because it is in a D stereoconformation, while most amino acids are in the L conformation.

A. True

B. False

(c) (2 points) What is peculiar about the allowed phi angles of this amino acid? (TB2)

Solution: The φ angles of this amino acid extend to the right side of the graph, while most amino acids have a very strong preference for negative φ angles.

(d) (4 points) What amino acid does this represent (full name please)? How do you know?

Solution: The lack of restriction on φ angles seems to suggest that there is more conformational freedom for this amino acid in a protein. This is true only for glycine, whose R group is very small–this allows glycine to adopt a variety of conformations in a protein, many of which are sterically forbidden for other amino acids.

Another possible explanation would be to observe that the plot itself is rotationally symmetric by 180° which indicates a lack of chirality. Since Glycine is the only achiral amino acid, this plot must be of Glycine.

While grading, we did not look for correct explanations and students should have recieved credit for just saying Glycine.

18. This question will focus on exotic proteins. Nith00n, a professor of ecology, has been studying weird organisms (Is that what ecologists do? I never did that event). He has been studying the proteome of a unicellular "prokaryote" from a lake on Titan, a moon of Saturn (astronomy, an event I actually did do). The lakes of Titan are primarily composed of methane and ethane, and according to Nith00n it is a reasonable assumption that the cell's interior has a similar composition.

Nit has found a strange protein that exists as a small, compact polypeptide. Strangely, these foreign organisms have the same genetic code as us and use the same set of amino acids. He has identified the protein sequence to be the following:

AIWYYSTNQRHDEQTLFYW

(a) (5 points) Identify which amino acids you would expect to be trapped inside the core, by 3 letter code. Enter as a comma separated list in alphabetical order. If more than one amino acid of the same kind exists, enter the same name twice–for example, if you think the core has 2 alanines and 1 valine, write Ala, Ala, Val.

Solution: Arg, Asn, Asp, Gln, Gln, Glu, His, Ser, Thr, Thr. Most teams did not answer in alphabetical order, thus we did not grade order. Note that these are all the polar amino acids, which you would normally expect on the exterior and not in the core–note also that in the preamble to this question, it is indicated that the cellular environment of these proteins is largely composed of methane and ethane, both hydrophobic molecules. It stands to reason these proteins would fold in such a way as to contain the polar residues inside the core away from the hydrophobic solvent. .5 points were awarded to each correct amino acid, and .2 points were deducted for each incorrect amino acid, with a minimum score of 0 points.

(b) (3 points) Nit has found that some of the amino acids in the outer portion of the protein are interacting with each other in a way that seems to be stronger than a standard van der Waals interaction. Identify what you think this interaction is. Identify by 3 letter code the amino acids possibly involved in this interaction (in alphabetical order). You only need to identify the type of amino acid–for example, if you think Q is involved, you need only list it once.

Solution: The outerportion is largely composed of W, Y and F. These do interact with van der Waals forces and London Dispersion interactions, but the question indicates a stronger interaction

type–these are Pi stacking or Pi-Pi interactions (aryl-aryl was accepted as well). The amino acids involved are Phe, Trp and Tyr.

(c) (2 points) Identify two amino acids in the core region that can possibly interact in a related (not identical) way with the amino acids you listed above by 3 letter code. The interaction type is not identical

Solution: Slightly ambiguously, this question wanted you to think about cation/anion pi interactions (the "related, but not identical" interaction). The three charged residues in the core are Arg, Glu and Asp–any 2 of these were accepted.

(d) (2 points) Which of the following mutations would have the most deleterious consequences on the protein structure?

A. All D become E

B. All A become L

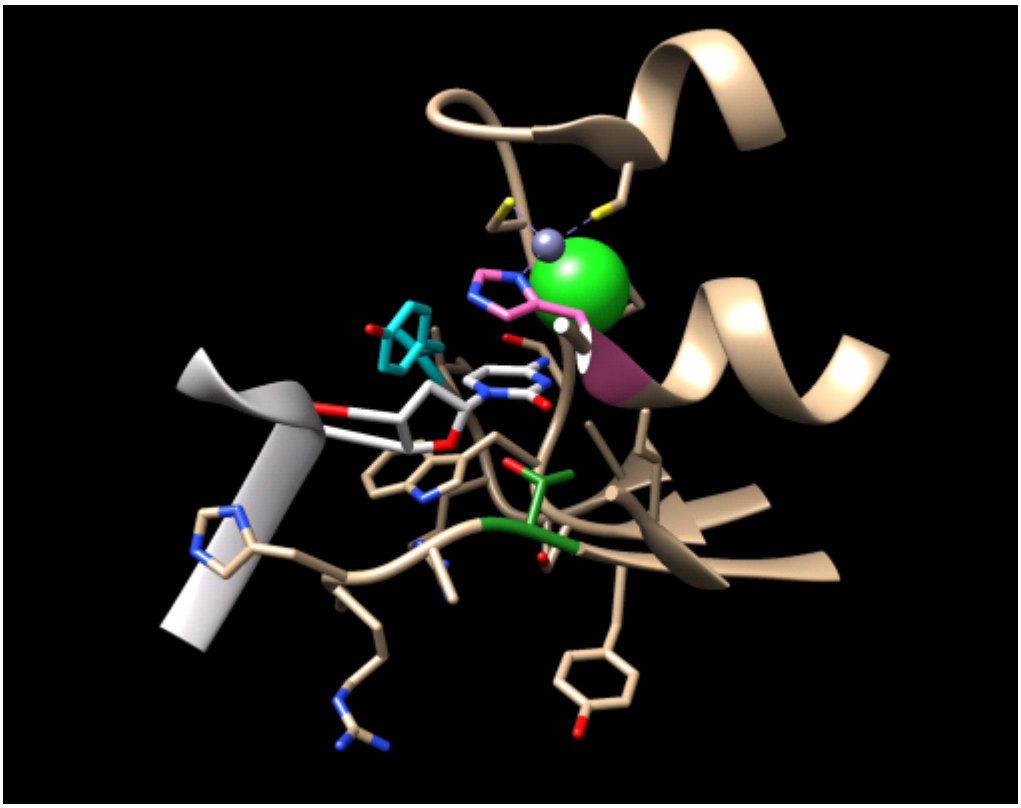
C. All Q become G

(e) (3 points) Alanine and Leucine are a distance of .7 nm. Would you expect this distance to be larger or smaller if this protein was instead dissolved in water? (TB3)

A. Larger

B. Smaller

19. This problem will focus on the structural analysis of a protein based on a diagram of its crystal structure. The following image contains a small region of a protein (in beige) interacting with a single strand of nucleic acid (in gray). There is also a divalent Zinc cation (silver) and a large Chloride ion (green).



Answer the following questions regarding this diagram. All red atoms are oxygen, and blue atoms are nitrogen. In the following, several questions are open-ended–be as concise as possible. Questions asking for interaction type do not need full sentence answers–simply "Hydrogen Bonding" will do, for example.

(a) (2 points) What type of nucleic acid is this?

A. DNA

B. RNA

(b) (3 points) Name the three amino acids colored cyan/light blue, pink, and dark green in that order. Write your answer as a comma-separated list of their 1 letter codes.

Solution: Y, H, T. Cyan can be distinguished because of the presence of the ring and the hydroxyl group (the red oxygen with a missing hydrogen). Pink is clear with its ring structure and two nitrogens in the ring. We did not [forgot to] grade for correct order here, so students who listed any of Y, H, or T in any particular order got credit (1 point for each correct answer).

(c) (2 points) What nucleotide is displayed? One letter code, please.

Solution: C; there is only one hexagonal nitrogenous ring, so it must be a pyrimidine. Since it is DNA it is C or T, but Thymine has a distinct methyl group attacked to the ring which is not seen here, thus it is Cytosine.

(d) (2 points) How does the base interact with the cyan and pink residues? (There is one interaction that the base forms with both of these residues).

Solution: Pi stacking interactions (and synonyms)–other answers, while possibly accurate, were not given points.

(e) (2 points) How does the pink residue interact with the zinc ion?

Solution: Electrostatic interactions (cation pi was given 1 point but it is in fact a correct answer here).

(f) (1 point) What other residue(s) interact with the Zinc ion?

Solution: The two other clearly displayed residues interacting with zinc are the ones containing an atom with a different color from those mentioned before–this atom must be sulfur, since other atoms colors are accounted for. These residues are cysteine.

(g) (3 points) Based on what nucleotide this protein region interacts with, as well as the fact that its activity can be coupled to a Cas9 protein, what protein might this be? [There are likely many answers, but given the theme of this event, one should stand out in particular]

Solution: Cytidine deaminase and APOBEC were both accepted

20. This problem will explore enzyme kinetics from a mathematical standpoint. You will be asked various different questions, some of which will rely on some quantitative reasoning. Kinetics is normally done using calculus, and some reference will be made to topics from calculus in this problem–rest assured, you will only need to know introductory algebra for this problem. Please, simplify your expressions as much as possible and write answers in accordance with the rules of order of operations. Onto the problem!

When studying reaction rates, scientists write differential equations to describe the change of a quantity over a period of time. A differential equation might look like this.

$$\frac{dA}{dt} = -kt$$

This differential equation indicates that a quantity A is changing at a rate of $-kt$ units per second at time t. For example, if $k = 1$ and $t = 5$ then at 5 seconds, A would be decreasing (from the negative sign) at a rate of 5 units per second. We can study the following enzyme catalyzed reaction.

$$\mathrm{S} \overset{\mathrm{E}}{\rightleftharpoons} \mathrm{P}$$

Here S indicates the substrate, P is the product and E is the enzyme catalyzing the reaction. There is also an important chemical species, the enzyme-substrate complex, which I will here denote as ES. There are 4 chemical species whose concentrations are changing in this system, so we can write 4 differential equations which show the relative changes at different times for thes species. These differential equations represent reaction rates.

$$\frac{d[\mathrm{E}]}{dt} = -k_f[\mathrm{E}][\mathrm{S}] + k_r[\mathrm{ES}] + k_{\mathrm{cat}}[\mathrm{ES}]$$

$$\frac{d[\mathrm{S}]}{dt} = -k_f[\mathrm{E}][\mathrm{S}] + k_r[\mathrm{ES}]$$

$$\frac{d[\mathrm{ES}]}{dt} = k_f[\mathrm{E}][\mathrm{S}] - k_r[\mathrm{ES}] - k_{\mathrm{cat}}[\mathrm{ES}]$$

$$\frac{d[\mathrm{P}]}{dt} = k_{\mathrm{cat}}[\mathrm{ES}]$$

Here, k_f is the forward rate constant the enzyme binding the substrate, k_r is the rate constant of the enzyme substrate complex dissociating into enzyme and substrate, and k_{cat} is the catalytic rate of the enzyme turning its substrate to its product. Also, square brackets around chemical species indicate the concentration of that species in molar units.

(a) (4 points) Write down a comma-separated set of two chemical reactions, one showing the enzyme and substrate combining to form the ES complex, and the other showing the ES complex breaking into the enzyme and product. Based on the above equation scheme, which reaction is reversible and which is not?

Solution: 2 points were given to each correct reaction.

$$\text{Enzyme Substrate Complex formation: } \mathrm{E} + \mathrm{S} \rightleftharpoons \mathrm{ES}$$

$$\text{Product formation: } \mathrm{ES} \longrightarrow \mathrm{E} + \mathrm{P}$$

Only the first is in equilibrium because based on the rate equations provided, there does not seem to be a way in which the product concentration decreases to return back to the ES complex. There is a way for the ES complex to degrade back into enzyme and substrate (namely, the second term in the third rate equation).

(b) (2 points) The steady state approximation (SSA) can be applied here. SSA allows us to make simplifications to kinetic models by setting the rate of change of an intermediate species to 0. Identify the only intermediate in your reaction scheme, set its rate of change to 0, and simplify the resulting expression to relate [E], [S], and [ES].

Solution: The intermediate is the enzyme substrate complex (it is not a reactant or product in the overall conversion of substrate to product. If we set the rate of change of the intermediate to 0 we have

$$0 = \frac{d[\mathrm{ES}]}{dt} = k_f[\mathrm{E}][\mathrm{S}] - k_r[\mathrm{ES}] - k_{\text{cat}}[\mathrm{ES}] \implies k_f[\mathrm{E}][\mathrm{S}] = k_r[\mathrm{ES}] + k_{\text{cat}}[\mathrm{ES}]$$

$$k_f[\mathrm{E}][\mathrm{S}] = (k_r + k_{\text{cat}})[\mathrm{ES}]$$

(c) (2 points) Label the total enzyme amount as $[\mathrm{E}_0]$. Write down an equation for the conservation of total enzyme amount. Hint: Conservation of enzyme means that the total concentration of all enzyme related species in the system must equal the total enzyme amount $[\mathrm{E}_0]$.

Solution: Since the total amount of enzyme adds to $[\mathrm{E}_0]$ we must have

$$[\mathrm{E}_0] = [\mathrm{E}] + [\mathrm{ES}]$$

(d) (3 points) Based on the above equations, write down an expression for [ES] in terms of $[\mathrm{E}_0]$ and [S], as well as any of the k values that have been defined in the image above. Then define

$$K_{\mathrm{M}} = \frac{k_f + k_{\text{cat}}}{k_r}$$

and simplify further, if possible. K_{M} is called the Michaelis constant.

Solution: First from the solution to part *b* we can write

$$[\mathrm{E}][\mathrm{S}] = K_{\mathrm{M}}[\mathrm{ES}]$$

We also have from (c) that

$$([\mathrm{E}_0] - [\mathrm{ES}])[\mathrm{S}] = K_{\mathrm{M}}[\mathrm{ES}]$$

Then

$$[\mathrm{E_0}][\mathrm{S}] = (K_\mathrm{M} + [\mathrm{S}])[\mathrm{ES}]$$

Thus

$$[\mathrm{ES}] = \frac{[\mathrm{E_0}][\mathrm{S}]}{K_\mathrm{M} + [\mathrm{S}]}$$

(e) (3 points) Based on your previous equation for [ES], solve for the rate of product formation. That is, solve for $\frac{d[\mathrm{P}]}{dt}$. This gives us the rate of product formation as a function of parameters we know. Refer back to the set of 4 equations you were given in the introduction to this problem to determine this.

Solution: From the rate equations we have that

$$\frac{d[\mathrm{P}]}{dt} = k_\mathrm{cat}[\mathrm{ES}]$$

Plugging in our solution to (d) gives

$$\frac{d[\mathrm{P}]}{dt} = \frac{k_\mathrm{cat}[\mathrm{E_0}][\mathrm{S}]}{K_\mathrm{M} + [\mathrm{S}]}$$

(f) (2 points) Based on your above reaction, what is the maximum enzyme reaction rate? This can be found by allowing [S] to become very large (i.e. much, much larger than K_M), so that the enzyme is saturated.

Solution: If we make [S] very large, we can ignore the K_M in the denominator. Then we have a quantity that is purely proportional to [S] in both the numerator and denominator of the rate–these factors must cancel. We are left with

$$V_\mathrm{max} = k_\mathrm{cat}[\mathrm{E_0}]$$

(g) (4 points) Let $V = \frac{d[\mathrm{P}]}{dt}$ be the reaction rate you found in question 5 and let V_max be the maximum rate you found in question 6. Write down an equation for the quantity $\frac{!}{V}$ as a function of $\frac{1}{[\mathrm{S}]}$. Identify what type of function it is, and if you can do so, provide a slope and y-intercept for the function in terms of constants from this problem. If you plot $1/V$ on a y axis and $\frac{1}{[\mathrm{S}]}$ on an x axis, you will generate a Lineweaver-Burk plot, a famous plot used to analyze enzyme kinetics.

Solution: First we let

$$V = \frac{V_\mathrm{max}[\mathrm{S}]}{K_\mathrm{M} + [\mathrm{S}]}$$

based on all the work we have done so far. Then we can flip both sides of the equation to get

$$\frac{1}{V} = \frac{K_\mathrm{M} + [\mathrm{S}]}{V_\mathrm{max}[\mathrm{S}]}$$

This can be split apart into

$$\frac{1}{V} = \frac{K_\mathrm{M}}{k_\mathrm{cat}}\frac{!}{[\mathrm{S}]} + \frac{1}{V_\mathrm{max}}$$

21. As you can tell, Nithin and I are pretty big fans of enzyme kinetics–here is another question on the topic. Learning about how an enzyme catalyzes its reaction is very important for any biochemist

learning about an enzymatic reaction. The following is an experimental set up that Nithin and I are working on. Nithin and I are amateur biochemists trying to determine some kinetic parameters for our enzyme, ShtU. We had an experimental plan laid out but unfortunately Nithin licked the paper and some of the words were wiped away. We have typed up what we salvaged before, but we are missing important information.

(a) (4 points) Based on the plots and relations you found in the previous problem, fill in the blanks in the following experimental plan that will allow us to characterize our enzyme. There are 4 blanks to fill; enter 4 words in a comma separated list as your answer.

"ShtU is an enzyme which catalyzes the decomposition of purple compound nituene into colorless compounds zootic acid and yootamine. The reaction rate as a function of substrate concentration can be measured with a _____ (device name), based on the principle of _____ (device name(persons name) law. Set up 5 different test tubes containing solutions that are .02M _____ (device name(chemical) and varying concentrations of _____ (device name(chemical) from .2M to 1M. Heat each test tubes in warm water to for 2 minutes to activate the enzyme, then immediately insert into the device to measure the reaction rate. Record initial reaction rate"

Solution: Spectrophotometer (colorimeter); Beer's (Beer-Lambert's); ShtU; Nituene

(b) (4 points) What key piece of an experimental set up is missing from the above? How would you add this element to your protocol? [There are two possible answers, both should be provided for full credit]

Solution: The set up is missing controls. You should set up one tube with only enzyme and no nituene, and another tube with only nituene and no enzyme. The first of these tubes will control for any absorbance or activity of the enzyme in the absence of nituene. The second of these will control for the natural absorbance of nituene.

(c) (3 points) Here is the data that we collected. Make a plot of reaction rate vs substrate concentration [not for credit, just to help you]. Estimate (rounded to nearest .5) the Kinetic parameters K_M and V_{max}. You may use the fact that K_M has units of concentration and is equal to the substrate concentration at which the reaction rate is exactly half of the maximum reaction rate achievable for the enzyme. This can be proved from your answers to the problem on Michaelis-Menten Kinetics.

Nituene Concentration (M)	Reaction Rate (M/s)
0.2	1.26
0.4	2.53
0.6	3.86
0.8	4.97
1.0	5.00

Solution: This question ended up confusing more students, but is actually just estimation based on graph points. It seems that the maximum reaction rate is plateauing around 5M/s (as can be seen by the very small increase from 4.97 to 5 after adding more nituene). Thus the maximum reaction rate V_{max} was 5M/s. Then half the reaction rate is 2.5M/s so K_M is around .4M, which is where the reaction rate is 2.53, or about half the maximum. We did not want students plotting the data or doing any complicated regressions, we just wanted a very simple estimation.

It turns out that the nituene stock used for a follow up experiment was contaminated by phencyclidine, a competitive inhibitor of ShtU. Remember, competitive inhibitors compete for the enzymes active site but do not actually affect the enzymes chemical structure or catalytic activity. Answer the following true/false questions.

(d) (2 points) K_M changes

A. True

B. False

(e) (2 points) $V_{\max}$ changes

A. True

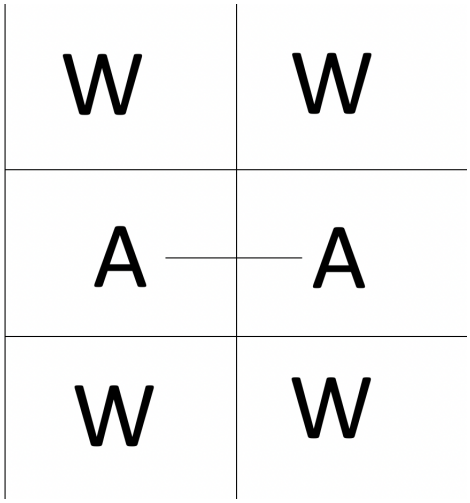
B. False

22. In this question, we will try to quantitatively relate the age of onset for Huntington's Disease (HD) to DNA sequence length for the gene causing HD. Huntington's Disease is a neurodegenerative disease caused by the accumulation/aggregation of fibrillary tangles composed of a particular polypeptide. This peptide is composed of a single amino acid repeated many times, and it forms misfolded tangles and clumps which then form inclusion bodies in neurons. This inhibits neuron function and slowly causes nerve death, leading to progressive motor disease, as well as dementia and cognitive decline. HD typically sets in between the ages of 30-50. Here we will examine this in a bit more detail. (This problem is adapted from Dill And Bromberg's Molecular Driving Forces.)

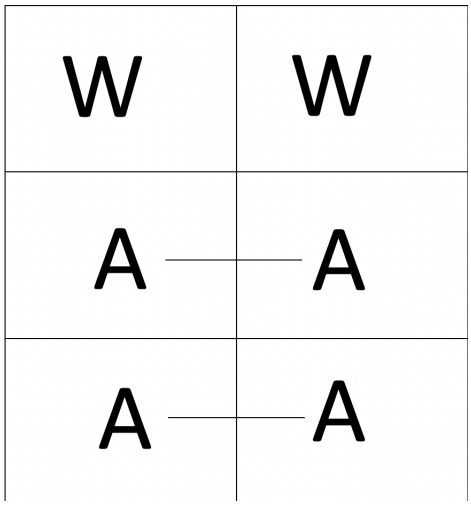
(a) (2 points) The peptide described above is composed of a particular amino acid. What amino acid is this, and what codon codes for it?

Solution: The amino acid is Glutamine and in Huntingtons is caused by repeats of the codon CAG. CAA also codes for Glutamine but this was a question specific to Huntingtons disease, thus only CAG earned credit. If CAG and CAA were both mentioned, full credit was still awarded.

In this question we will introduce and study an elementary lattice model of polymer interaction. It will be useful in this problem for us to introduce the lattice model of interacting peptide chains. For the purposes of this problem, a lattice is given as an $n \times m$ grid of squares, each of which contains either one molecule of solvent or one monomer in a polymer chain. The following image shows a 3×2 lattice representing a dipeptide (two amino acid long peptide) interacting with water.



In this problem, all amino acids will be represented as A, and all solvent molecules are W. In any lattice, any adjacent amino acids which are connected by a covalent bond (same peptide) will have a black line between them, indicating the bond. If two adjacent squares do not have a bond between them, then they interact with intermolecular forces only. The following shows two interacting dipeptides.



Note here that the adjacent dipeptides are not covalently bonded to each other, but they do interact with intermolecular forces. There are three relevant noncovalent interactions we must consider: water-water (ww), water-peptide (wp) and peptide-peptide (pp).

For some of the following problems, we will have you construct your own lattices. When writing your answers, you should construct a small grid with gridboxes and fill in each box with A or W to represent what would be contained in that gridbox. Your lattices should look similar in format to the ones provided (though they will have different sizes and contents)

(b) (4 points) Construct two 5x5 lattices. In the first lattice, represent the following situation.

2 5AA long peptides separated by a later of water one lattice space thick. Draw all peptides horizontally (hint: this means you should have two rows written out as AAAAA)

In the second situation, represent the following situation.

2 5AA long peptides interacting directly with each other along their length,and also interacting with water on both their free sides. (hint: this means the top and bottom rows of your lattice model should be made of water)

Clearly indicate which lattice is the answer for each question. The second situation can be represented in two equally correct ways–either will be accepted.

Solution: For condition one we have the following lattice. Note there is exactly 1 layer of water between the two rows of amino acids.

W W W W W
A A A A A
W W W W W
A A A A A
W W W W W

The other lattice will have adjacent rows of amino acids. one example of this is

W W W W W
A A A A A
A A A A A
W W W W W
W W W W W

Another accurate answer would have been a lattice where the third and fourth rows are all amino acids.

Now consider peptide chains of length L. There are three relevant interactions and thus three relevant energies we need to account for in the process that converts the first condition to the second in the above subpart. That is, the process that causes two separate peptide chains to come together and interact with each other along their length.

(c) (3 points) Let E_{ww} be the energy of ww interactions, E_{wp} be the energy of wp interactions, and E_{pp} be the energy of pp interactions. Fill in the change in number of each interaction type in the process for two length L peptides interacting. Negative numbers indicate interactions broke. Answer in terms of L (you might need to draw a few more conditions with varying lengths to find a pattern if that helps).

pp	wp	ww

For example, if between condition 1 and condition 2, two pw interactions broke and one pp interaction formed, you would fill in the table as

pp	wp	ww
1	-2	0

Solution: When the two strands interact we gain L *pp* interactions. Also, since we now have lost all water layers between the two strands, we are effectively losing $2L$ *pw* interactions, where the 2 comes from having the water layer in between the two strands interacting with both strands. Since in the lattice the total number of interactions must be conserved since every box in a column interacts with its neighbors, we must gain back L interactions from somehwere–these L interactions come from L new *ww* interactions. You can work this out in the two lattices you drew for 5×5 and maybe draw another one for 3×3 to get that the answer is

pp	wp	ww
L	$-2L$	L

(d) (3 points) Consider the following reaction:

$$2\,\mathrm{AAA}\ldots\mathrm{AA} \rightleftharpoons \mathrm{AAA}\ldots\mathrm{AAA}_2$$

This reaction represents two free peptides coming together and interacting with each other. The subscript 2 does not indicate that this is now a new polypeptide of length $2L$. This is merely two L AA-long peptides interacting with each other along their length.

We define the equilibrium constant K for this reaction to be related to the free energy change of the reaction. You can calculate the free energy change ΔE based on the answers you provided for the previous problem from the energies you were given and the numbers you calculated for interactions formed/broken. Assume that

$$E_{pw} < \frac{E_{pp} + E_{ww}}{2}$$

for the purposes of this problem. Use the following equation to determine an expression for K in terms of L. All we expect is a general form of an equation–for example, if you find that K is a linear function of L with a negative slope, then all we want is $K = -mL + b$, a general linear equation with a clear negative sign to indicate negative slope. (hint: it is not a linear function, this was just an example). The following will be useful:

$$\Delta E = -RT \ln K$$

Solution: First we can calculate the change in free energy. We know based on the solution to the previous problem, we are getting an energy contribution of $-LE_{pp}$ and $-LE_{ww}$ from the new interactions formed. We also lose a contribution of $-2LE_{pw}$ from the broken bonds. The signs here are a little bit difficult to rationalize, but the energy contributions of new interactions end up being negative since the interaction is forming and therefore the energy lowers. Then

$$\Delta E = -L(E_{pp} + E_{ww} - 2E_{pw})$$

Note that based on the inequality provided, this free energy is of the form $-mL$ for a positive number m. Thus we have

$$\Delta E = -mL$$

Then from the condition of free energy relating to the equilibrium constant, we have

$$K = e^{mL/RT}$$

All we required was this form of answer–an exponential with $+L$ dependence.

(e) (1 point) If the equilibrium constant is very large, we expect that the concentration of the dimerized, interaction peptide chains will be very high in comparison to the concentration of free peptides. Based on this information, would you expect a higher concentration of dimerized peptides when the peptide length is very long or short?

A. Short

B. Lomg

(f) (2 points) If there are larger amounts of clustered, dimerized, and tangled up peptide chains, Huntington's Diseases begins earlier and progresses faster. Therefore, if we have the following four individuals with a gene coding for our polypeptide of various lengths, who would you predict to get Huntington's earliest?

A. Person A with a 459 bp long gene sequence

B. Person B with a 849 bp long gene sequence

C. Person C with a 298 AA long polypeptide post-translation

D. Person C with a 168 AA long polypeptide post-translation

Section 2: CRISPR-Cas systems

23. Here we will try to do some single base editing with a modified Cas9 Protein. We will be using the Cas9 variant protien formed by fusing Cas9 nickase to APOBEC 1. Consider the following Space+PAM sequence in human DNA which is targeted by an sgRNA tied to the fusion protein described.

5' ATATCAATCCCTGCTATTAACGG 3'

(a) (3 points) Identify the PAM site. Give your answer as a nucleotide sequence (1 letter codes) from 5' to 3'.

Solution: CGG; PAM sites are generally NGG

(b) (3 points) A relevant mutation for this DNA sequence would be a base substitution that directly affects the protein coded by the strand. Assuming the fusion protein acts with 100% efficiency on bases between positions 4-8 in the sequence (bases are numbered starting at 1 on the 5' end), how many relevant base substitutions will occur? Assume that base 2 is the first base of a codon which is in frame for the strand.

Solution: The key to note here is that we are looking only for relevant mutations, i.e. things that will end up manifesting in the protein. The C at position 5 is replaced by a T which ends up turning the CAA codon to a TAA, which ends up being a stop codon. Thus the rest of the sequence doesnt matter. There is only 1 relevant mutation.

(c) (4 points) What sort of mutation occurs in this protein after editing with the fusion protein?

A. Missense

B. Nonsense

C. Frameshift

D. Silent

(d) (5 points) dCas9 proteins do not localize to DNA as they have lost their exonuclease capability.

A. True

B. False

(e) (5 points) A common inhibitor strategy to prevent off-target editing is the introduction of Cas9 nickases to cleave regions of the Cas9 protein.

A. True

B. False

(f) (5 points) A scientist trying to suppress DNA repair mechanisms which are kinetically fast should develop inhibitors for proteins that require homology between strands for repair. (TB4)

A. True

B. False

(g) (5 points) Mutations in HNH primarily lead to HNH mimicking RuvC function.

A. True

B. False

(h) (5 points) dCas9 based transcription factors are ineffective due to their relative inability to target specific DNA sequences.

A. True

B. False

24. (2 points) KRAB is a transcriptional inhibitor. A researcher performs an experiment on the transcription of the gene inhibited by KRAB in three conditions–(i) absence of KRAB, (ii) presence of KRAB, and (iii) presence KRAB fused to dCas9 in the presence of sgRNA. Which of the following experimental techniques can the researcher use to compare the expression levels of this gene in each background?

A. qRT-PCR

B. Southern Blotting

C. ELISA

D. Proteomics Assays

25. (2 points) Which condition is the control, if any?

A. i

B. ii

C. iii

D. none

26. (2 points) Which of the following most likely lists the three conditions in the correct order of mRNA levels for this particular gene?

A. $i < ii < iii$

B. $i > iii > ii$

C. $i > ii > iii$

D. $i > ii = iii$

E. $i = ii < iii$

F. $ii < i = iii$

27. (2 points) Another researcher performs the same experiment but incorrectly mutates the RuvC domain so that it has some functionality. Which of the following experimental errors might he observe?

A. Non-target transcriptional activation

B. Non-target strand SSB formation

C. Target Strand transcriptional activation

D. Off-target transcriptional activation

28. (3 points) A scientist wishes to modify the experimental design protocol that was previously discussed. However, instead of using sgRNA as a guide the researcher uses individual crRNA and tracrRNA. Which of the following enzymes might be needed for the experiment to mimic the original?

A. A ribonuclease

B. A helicase

C. A protease

D. An exonuclease

29. (3 points) Suppose the above researcher did not include the particular enzyme. Which experimental condition might you expect this to be most similar to?

A. i

B. ii

C. iii

D. None, as the experiment would not work at all

30. (2 points) The first stage of the CRISPR immune response in bacteria is space acquisition, or the insertion of foreign DNA into a CRISPR locus. Select all of the following proteins that facilitate this process. Select ALL correct answers.

A. Cas1

B. Cas9

C. Cas12a

D. Cas3

E. Cas2

31. (3 points) Cas12a (Cpf1) is a Type _____ CRISPR system that generates _____ breaks following binding of the crRNA to a DNA sequence

A. III; Blunt end DSB

B. I; ssDNA nick

C. II; Staggered end DSB

D. II; Blunt end DSB

E. I; Staggered end DSB

32. (3 points) You are interested in modifying Cas9 with the aim of controlling and manipulating epigenetic modifications. Select all of the following Cas9 fusions/modifications that could be reasonable for your purpose. Linker is denoted with a double dash "–". Select ALL correct answers.

A. dCas9 – DNMT3a (DNA methyltransferase domain)

B. dCas9 – KRAB (complexes with 2 histone methyltransferases)

C. dCas9 – E2 (ubiquitin ligase for nearby protein degradation)

D. dCas9 – HDAC1 (histone deacetylase)

E. dCas9 – PARP (poly-ADP ribose polymerase; detects ssDNA breaks)

33. (3 points) Genome wide CRISPR Cas9 knockout screens are designed to find the relationship between a genotype and phenotype using libraries of single guide RNAs (sgRNAs). Identify all of the following statements that are true regarding the differences between Cas9 knockout screens and RNA interference (RNAi). Select ALL correct answers.

A. RNAi libraries sequences are shorter than synthesized sgRNAs, requiring a smaller library

B. At present, upwards of 19,000 human genes can be targeted with Cas9 knockout screens

C. RNAi knockout affects expression post-transcriptional, while Cas9 targeting would knockout genes pre-transcription

D. Recessive Cas9 knockout mutations need only one edited chromosome for knockout phenotype

E. Cas9 approaches have lower noise and less off target binding compared to RNAi knockouts

34. (5 points) Which of the following mechanisms could be effective for preventing Cas9 expression in off-target tissues. Select ALL correct answers.

A. ProCas9 - Cas9 is active only in the presence of a certain protease present in target tissue

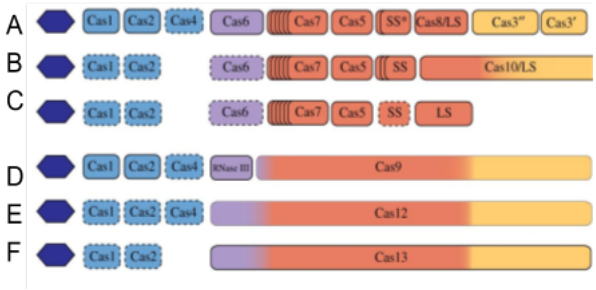
B. AcrIIA4 - protein that will bind and inactivate Cas9; expression in non-target tissues

C. SunTag - polypeptide array fused to Cas9 that recruits antibodies with transcription factors

D. Cas9 coding sequence and sgRNA under control of a tissue specific promoter

E. Ubiquitin - Cas9 fusion results for a low half life of Cas9 in target and non-target tissues

35. (5 points) Select ALL correct statements about the below diagram.



[Question notes: This question was thrown out after the image did not appear on many participants screen during the test. 0 points were awarded for all teams.]

A. [A] has a multiprotein effector complex and would belong to a Class 1, type I CRISPR system

B. [B] has a single protein effector (Cas10) that is homologous to Cas10d

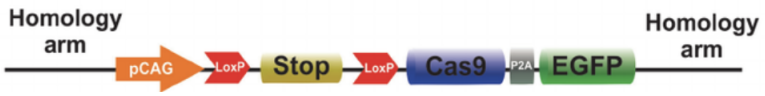
C. [C] has a multiprotein effector complex and has a type IV Cas protein

D. [D] has a Cas protein with a conformationally flexible domain that cleaves the non-target strand

E. [E] has a functionally unique domain architecture that does not require a tracr-RNA

F. [F] has been used as an RNA-guided RNAase in novel diagnostic techniques such as SHERLOCK

36. (5 points) You are interested in knocking out a gene in a specific tissue and create a Cre-dependent Cas9 targeting vector for insertion. Cre recombinase will excise DNA between LoxP sites, and is frequently used to create self-excising cassettes. A diagram of your vector is provided below. Select ALL of the following choices that are TRUE.



A. Successful genomic integration into the target locus will depend heavily on the sequences and length of the homology arm used

B. In off target tissues, fluorescent microscopy reveals green fluorescence / EGFP expression

C. pCAG could be replaced with a common promoter such as T7 (derived from the T7 bacteriophage) for expression in eukaryotic cells

D. Replacement of pCAG with a stronger promoter allows for Cas9 expression in absence of Cre recombinase

E. The "Stop" block likely corresponds to a transcriptional repressor that is excised following expression of Cre in the target tissue

37. (5 points) Use the following text and background knowledge to select ALL of the following statements that are TRUE.

> "Indeed, in the structure of the Cas9 R-loop complex, the first 11 nt of the protospacer on the displaced [1] **DNA strand are disordered**, suggesting that their movement is not highly restricted. It has also been speculated that Cas9 nickase-induced mutations at cytosines [2] **in the non-template strand might arise from their accessibility by cellular cytosine deaminase enzymes"**
>
> Alexis C. Komor1,2, yongjoo B. Kim1,2, michael S. Packer1,2, John A. Zuris1,2 David R. Liu1,2

A. [1] directly implies that the sgRNA binding to the protospacer is comparatively weak and allows for "sliding"

B. [2] suggests the presence of cellular cytosine deaminases that operate solely on ssDNA

C. Cas9 nickase-induced mutations at cytosines could still occur if a catalytically dead (dCas9) protein was used

D. Conformational entropy studies of the Cas9 R-loop complex would likely show a highly disordered domain that is stabiiized only by sgRNA binding

Research Literacy

In this section, you will be presented with abstracts from a few research works studying CRISPR-Cas9. You will be tested on how well you can apply your knowledge of the Cas9 editing system to understand the information presented in these abstracts.

38. **Real-space and real-time dynamics of CRISPR-Cas9 visualized by high-speed atomic force microscopy**

Mikihiro Shibata, Hiroshi Nishimasu, Noriyuki Kodera, Seiichi Hirano, Toshio Ando, Takayuki Uchihashi Osamu Nureki

> The CRISPR-associated endonuclease Cas9 binds to a guide RNA and cleaves double-stranded DNA with a sequence complementary to the RNA guide. The Cas9–RNA system has been

harnessed for numerous applications, such as genome editing. Here we use high-speed atomic force microscopy (HS-AFM) to visualize the real-space and real-time dynamics of CRISPR-Cas9 in action. HS-AFM movies indicate that, whereas apo-Cas9 adopts unexpected flexible conformations, Cas9–RNA forms a stable bilobed structure and interrogates target sites on the DNA by three-dimensional diffusion. These movies also provide real-time visualization of the Cas9-mediated DNA cleavage process. Notably, the Cas9 HNH nuclease domain fluctuates upon DNA binding, and subsequently adopts an active conformation, where the HNH active site is docked at the cleavage site in the target DNA. Collectively, our HS-AFM data extend our understanding of the action mechanism of CRISPR-Cas9.

(a) (2 points) You purify and conjugate apo-Cas9 to a fluorescent marker to visualize DNA binding. Do you expect DNA binding? If so, what differences are observed relative to Cas9-RNA?

Solution: You would expect DNA binding as Cas9 would associate with DNA containing the PAM sequence. (+1) Compared to Cas9 bound to an sgRNA, this binding would be non-specific since there may be binding to any region with a PAM. (+1). The likely model of Cas9 binding involves PAM binding followed by stabilization by formation of the phosphate lock loop.

(b) (1 point) What two domains are responsible for the bilobed structure of Cas9?

Solution: REC / recognition and NUC / nuclease (+1)

(c) (2 points) What DNA strands do the RuvC and HNH domains cleave, respectively? Provide 2 structural changes that Cas9 binding induces on the DNA.

Solution: HNH - Target ; RuvC - non-target (Accept: non-target and target) (+1 for both). Reasonable answers were accepted for changes to DNA including the most common: local melting / unwinding (+1)

39. **Multiplex Genome Engineering Using CRISPR/Cas Systems**

Le Cong1,2,*, F. Ann Ran1,4,*, David Cox1,3, Shuailiang Lin1,5, Robert Barretto6, Naomi Habib1, Patrick D. Hsu1,4, Xuebing Wu7, Wenyan Jiang8, Luciano A. Marraffini8, Feng Zhang1,†

Functional elucidation of causal genetic variants and elements requires precise genome editing technologies. The type II prokaryotic CRISPR (clustered regularly interspaced short palindromic repeats)/Cas adaptive immune system has been shown to facilitate RNA-guided site-specific DNA cleavage. We engineered two different type II CRISPR/Cas systems and demonstrate that Cas9 nucleases can be directed by short RNAs to induce precise cleavage at endogenous genomic loci in human and mouse cells. Cas9 can also be converted into a nicking enzyme to facilitate homology-directed repair with minimal mutagenic activity. Lastly, multiple guide sequences can be encoded into a single CRISPR array to enable simultaneous editing of several sites within the mammalian genome, demonstrating easy programmability and wide applicability of the RNA-guided nuclease technology.

(a) (1 point) Describe one way on which type I CRISPR systems differ from type II CRISPR systems? [Question notes: This was thrown out due to ambiguity over Class I / II and Type I / II CRISPR systems. All teams should have received no points]

(b) (2 points) Name 2 advantages of a Cas9 protein that cuts only one strand of DNA rather than create double stranded breaks. Mention DNA repair pathways in one answer.

Solution: A nickase will not generate DSBs (double stranded breaks) which are often repaired through NHEJ, which results in indels (+1). HDR has less chance of indels than NHEJ (+1) Alternative answers: fewer off target effects if multiplexed /multiple Cas9 nickases are used (+1); no loss of specificity relative to regular Cas9 (+1)

(c) (2 points) The nuclease Cas12a (Cpf1) is also a type II CRISPR system. Name 2 advantages of Cas12a relative to Cas9.

Solution: Possible answers (+1 each): Staggered cuts vs blunt ends in Cas9, different PAM, only requires crRNA, shorter crRNA for use in multiplexed genome editing; greater specificity, Cas12a can process crRNA

40. **Protein Engineering Strategies to Expand CRISPR-Cas9 Applications**
Lucas F. Ribeiro,1 Liliane F. C. Ribeiro,2 Matheus Q. Barreto,2 and Richard J. Ward3

To achieve this goal, dCas9 was used as a DNA-binding domain and fused to a deaminase. In 2016, Komor et al. created a so-called "base editor" to convert cytidines into uridine within a sequence of five nucleotides located between the protospacer and PAM [51]. A cytidine deaminase from Rat known as APOBEC1 (apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 1) was fused to the N-terminus of dCas9 using a linker in order to maintain deaminase activity. Iterative optimizations in the linker and chimera were performed, and each step of optimization produced a new "generation" of the construction. Therefore, the third generation of base editors (BE3) consisted of APOBEC1 linked to the N-terminus of dCas9 with the catalytic His840 restored and a uracil glycosylase inhibitor linked at Cas9 C-terminus (APOBEC–linker–dCas9(A840H)–UGI).

(a) (1 point) The mutation A840H restores endonuclease function in the HNH domain of a catalytically inactive dCas9 and is used in the third generation of base editors (BE3) referenced above. You experiment with the purified protein and realize that deaminase activity is maintained without the A840H mutation. Provide a brief explanation as to why the partial restoration of Cas9 endonuclease function is necessary for base editing.

Solution: Restoration of nickase activity recruits DNA repair enzymes and allows for deaminase activity to result in a permanent mutation (+1)

(b) DNA glycosylases are a family of enzymes that catalyze the first step of base excision repair (BER). Uracil DNA glycosylase (UDG) cleaves the N-glycosydic bond to generate an AP site. Predict the sequence of the DNA in the following 3 cases (consider each separately / non sequentially). The DNA sequence is given below:

5' ACCUGGAUGGG 3'
3' TGGGCCTGCCC 5'

i. (1 point) After treatment with UDG (assume 100% efficiency – denote any space missing a nucleotide with a dash)

Solution:
5' ACC–GGA–GGG 3' (accept AP site as a space)
3' TGGGCCTGCCC 5'
As noted in the description, Uracil DNA Glycosylase (UDG) will generate an abasic (AP) site by excising the nitrogenous base. Immediately following the creation of the abasic site, the backbone is left intact. However, AP endonucleases will usually convert AP sites to nicks. This answer was accepted as well.

ii. (2 points) After one round of replication in the ABSENCE of UDG (provide both strands)

Solution: All or nothing for both strands (+2)
5' ACCUGGAUGGG 3'
3' TGGACCTACCC 5'

5' ACCCGGACGGG 3'
3' TGGGCCTGCCC 5' There is no UDG present, so replication will proceed as normal with

DNA polymerase incorporating an adenine for any uracil nucleotides present. This generates two strands. Note that the bottom strand has no mutations, while the top would have permanent mutations following BER.

iii. (1 point) In the presence of UGI and UDG with complete efficiency of inhibition

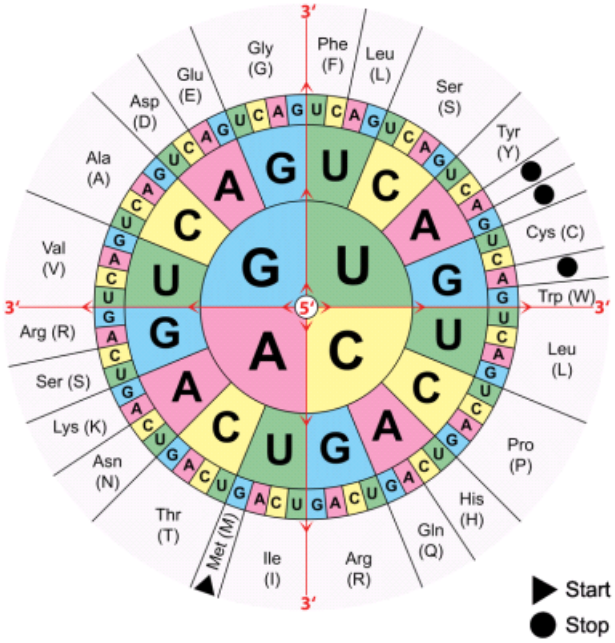
Solution:
All or nothing for both strands (+1)
5' ACCUGGAUGGG 3'
3' TGGACCTACCC 5'

5' ACCCGGACGGG 3'
3' TGGGCCTGCCC 5'
UGI will inhibit UDG, so the replicated strands will be the same as part ii. This question is intended to show how the addition of UGI will result in permanent mutations following deamination (to uracil).

Note: Refer to the abstract for parts of the following question.
Your PI asks you to modify a protein from S. hrunt so that it is catalytically inactive by mutating a threonine residue to a hydrophobic residue (T→X). You identify the following sequence located in frame with the coding strand 5' → 3'. Assume that the mutation can occur up to 5 bp upstream of the PAM sequence and that SpyCas9 (5' NGG 3') is used for both fusion proteins. Note that any base editing will occur only on the strand where the PAM sequence is located.

—1	4	7	10	13	16	19	22 —
5' GGG	ACA	TGG	CTA	ACA	CCT	GAT	ACG 3'
3' CCC	TGT	ACC	GAT	TGT	GGA	CTA	TGC 5'

For reference, the numbers above the sequence indicate base pair position–each number corresponds to the index of the first base in that codon.
Your PI gives you two tools: the base editor mentioned in the paper (BE3) and an adenine base editor (ABE7) from SpyCas9 that converts A•T to G•C up to to 5 bp upstream of the PAM sequence. You can use either one, or both in conjunction with each other to edit the DNA, which is given to you as a purified plasmid. Assume that you have generated a sgRNA that will allow Cas9 to target the entire nucleotide sequence above. Here's a codon table.



(c) (1 point) Provide the nucleotide index (in range 1-24) of the PAM sequence on the coding strand.

Solution: (7-9). Note that the PAM sequence is 3 nucleotides so (7-10) was not accepted.

(d) (1 point) Provide the nucleotide index (in range 1-24) of the threonine codon given that it is possible for the base editors to edit this codon.

Solution: (4-6). This must also be a 3 nucleotide range. (4-7) was not accepted.

(e) (3 points) To achieve the desired mutation using the above editors, what is a possible residue X can you mutate the specific threonine to? There are multiple possible answers–choose the residue which is closest in molar mass to threonine amongst these possibilities.

Solution: Valine / Val (+3). Valine is closest in molecular mass to threonine.

(f) (5 points) Propose a method to edit the target threonine to the residue in part e, using the provided base editors. Indicate which base editor(s) is (are) used, and the sequence 5' → 3' of the **target strand** following the edits.

Solution: ABE7 / BE3 (+2)
5' GGG **GTG** TGG GAT CCT CAT ACA ACG 3' (+3)
The target strand will have edits ONLY 5 bp upstream of the PAM sequence (7-9), so the bolded region corresponding to the threonine residue will have an edit on the target strand.

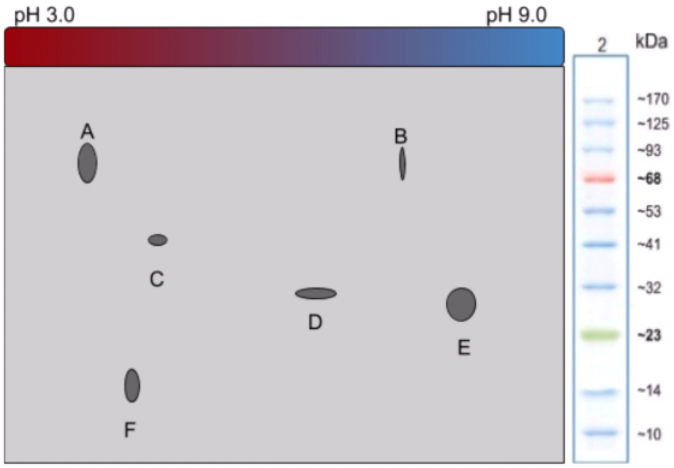
(g) (3 points) You perform the method in part f, expecting it to work flawlessly, but you notice that your method produces (an) additional mutation(s) following PCR that results in a nonfunctional protein. As your PI admonishes you, you figure out the cause of the mutation(s). Provide the nucleotide index(indices) of the mutated codon(s) along with the codon sequence(s) on the coding strand before and after mutation. Briefly state why such this would result in a nonfunctional protein.

Solution: 19-21 (+1)
GAT → AAC (+1)
Accept reasonable answers: Glu is charged with an acidic side chain that would function in active site (+1)
There is an additional PAM sequence on the coding strand 5' AGG 3' that Cas9 could have bound to. It's assumed in the problem that the sgRNA can target this region. Credit was awarded for an explanation that this mutation would not occur given proper sgRNA design.

(h) (2 points) To salvage your reputation, you search for a modification to your protocol that would prevent off target effects using the same base editors. Briefly detail this below along with why it would work.

Solution: You would change your sgRNA to target only upstream of the PAM sequence desired. (+1) This would prevent any ssDNA from forming downstream and thus prevent off target edits by cytidine deaminase (which will only catalyze conversion in ssDNA).

41. Now we will begin some analysis of experimental techniques. First, 2D gel electrophoresis. Two-dimensional gel electrophoresis (2-DE) is a popular and versatile method of protein separation that uses two independent biochemical characteristics of proteins. Difference gel electrophoresis (DIGE) is a form of 2D gel electrophoresis that stains protein samples with spectrally resolvable fluorescent dyes. Your PI, Dr. Rajesh, requests that you analyze the 2-DE gel of S. patel proteins.



Using the results SDS-PAGE gel above and your knowledge of the technique, answer the following questions:

(a) (1 point) What important chemical property of a protein can be calculated from its horizontal position in the gel?

Solution: pI or isoelectronic point. This is not pH (which ended up a very common answer). pH is a characteristic of a solution, not a protein in particular.

(b) (3 points) It is found through crystallographic measurements that proteins A and E interact quite strongly (more so than basic van der waals interactions) in the cytoplasm, but they can be separated from each other in a non-reducing environment. What interaction do you expect between these two proteins? Explain briefly

Solution: Electrostatic interactions. If they can be separated in nonreducing environments then the units are not interacting with covalent bonds like disulfide bridges. Looking at the positions of the spots, they seem to have very different isoelectric points, which means in solution they will end up having opposite charges–this leads to the electrostatic interaction.

(c) (3 points) Suppose the genes encoding these proteins were cloned into an organism which had the unique ability to synthesize proteins without the use of glutamate or aspartate–instead these residues are substituted for structural "equivalents" glutamine and asparagine. What might you expect to happen to the positions of each of these fluorescence spots on the gel? Why? (1 sentence explanation)

Solution: Everything should move right slightly, because every protein's pI will increase a little bit as a result of the changing of negatively charged amino acids to neutral amino acids. We are making all the proteins "less acidic" so we should see them all move toward the more basic region on the DIGE.

(d) (3 points) It is found that D forms a multimer in solution. When running in native PAGE, you see a dark band at a mass of approximately 96 kDa. What is a lower bound on the number of cysteine residues in protein subunit D? (TB5)

Solution: We can estimate that when we run in Native PAGE we see a trimer of D since the mass of D is 32 and the band on the PAGE is at 96. Thus three D subunits are held together, likely by disulfide bridges. The minimum number of cysteines in one D must be $\boxed{2}$ since one of the D subunits has to bind to each of the other two in the trimer, each with a disulfide bond.

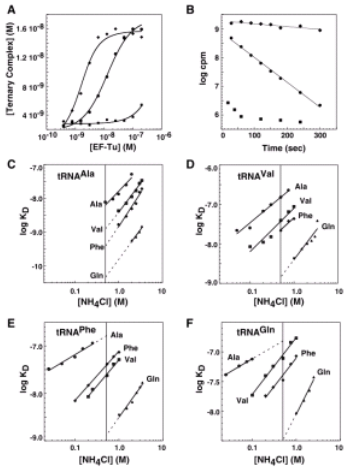
42. (Adapted from LaRiviere et al)
Elongation factor Tu (EF-Tu) is a guanine nucleotide binding protein that, when complexed with

guanosine 5'-triphosphate (GTP), binds elongator aminoacyl-tRNAs (aa-tRNAs) and participates in the early steps of codon-directed peptide bond formation catalyzed by the ribosome.

EF-Tu binds all aa-tRNAs for delivery to the ribosome during protein synthesis in a uniform manner. While this was previously thought to be accomplished via general specificity for the tRNA structure, EF-Tu is able to bind misacylated tRNAs over a wider range of affinities than the corresponding correctly acylated tRNA. In addition, EF-Tu may contribute to translational accuracy due to its ability to recognize, through binding affinity, various misacylated tRNAs.

To characterize the contribution of amino acid and tRNA body to EF-Tu binding, variants of E. coli tRNAAla, tRNAVal, tRNAGln, and yeast tRNAPhe, lacking modified nucleotides, were each amino-acylated with alanine, valine, glutamine, and phenylalanine, creating 4 cognate aa-tRNAs and 12 misacylated aa-tRNAs

The researchers designed a ribonuclease (RNase) protection assay to determine the K_d's for the 16 different aa-tRNAs to EF-Tu GTP. The results are shown below. The vertical line in plots C-F corresponds to [NH4Cl] = 0.5 M. Note that there may be shifted scales in pots C-F!



(a) (3 points) The formation of the EF-Tu ternary complex is shown by representative equilibrium binding curves in figure A. In prokaryotic translation, what 3 components are part of this ternary complex?

Solution: EF-Tu, GTP, tRNA

(b) (1 point) The formation of the ternary complex from the uncomplexed parts could be explained by a mechanism in which each of the 3 parts can bind to each other before combining to create the final ternary complex

A. True

B. False

Professor Rajesh synthesizes an antibiotic that disrupts the binding of complexed EF-Tu to the amino-acylated tRNA (aa-tRNA) in a competitive manner. He asks his grad student, Bob, to predict the effect of this on translation.

(c) (1 point) Addition of the antibiotic would not halt translation as bacteria have evolved exogenous mechanisms that allow for direct incorporation of the aa-tRNA into the ribosomal A (amino-acyl) site.

A. True

B. False

(d) (1 point) This antibiotic could bind to EF-Tu away from the active site and increase the Kd of aa-tRNA binding by inducing a conformational change in EF-Tu

A. True

B. False

(e) (1 point) A bacterial compensatory mechanism that increases the steady state concentration of aa-tRNAs could be effective in mitigating the effect of the antibiotic

A. True

B. False

(f) (1 point) Addition of the antibiotic would likely result in increased ribosomal stalling and subsequent dissociation

A. True

B. False

(g) (1 point) Mutations in the 30S subunit of the ribosome would could prevent the binding of the antibiotic and effectively mitigate the effect of the antibiotic

A. True

B. False

(h) (1 point) This antibiotic would likely affect translational fidelity (misincorporation of amino acids)

A. True

B. False

(i) (2 points) The chemical reagent NH_4Cl is used to generate the plots C-F. Select the choice(s) that is (are) True.

A. NH_4^+ ions likely disrupt the binding of an aa-tRNA to EF-Tu

B. A 10x increase in the concentration of NH_4Cl (0.1 to 1.0 M) would result in a roughly 10x decrease in binding affinity for the correctly acylated aa-tRNA in plot D

C. The linear relationship of $\log K_d$ vs $\log [NH_4Cl]$ for all binding plots and the similarity in slopes could suggest that NH4+ ions compete with EF-Tu in binding to the tRNA body

D. The predicted binding affinities for higher concentrations of NH4Cl beyond the data in the plot can be extrapolated since NH_4Cl is unlikely to affect the structure of EF-Tu at high concentrations.

(j) (3 points) Graph interpretation. Select the choice(s) that is (are) True.

A. At 0.5 M NH_4Cl, the misacylated val-$tRNA^{Phe}$ has a higher binding affinity than the correctly acylated tRNA

B. At 0.5 M NH_4Cl, the misacylated ala-$tRNA^{Val}$ has a lower binding affinity than the correctly acylated tRNA

C. At 0.5 M NH_4Cl, the misacylated val-$tRNA^{Gln}$ has a lower binding affinity than the misacylated phe-tRNAGln

(k) (5 points) The table shows the thermodynamic contributions of the amino acid and tRNA body in binding to EF-Tu. Select all of the following statements that are TRUE using Table 2 and the plots A-F

Table 2. Independent thermodynamic contributions of amino acid and tRNA body. $\Delta G^\circ = -RT\ln(1/K_d)$.

Cognate aa-tRNAs			Misacylated aa-tRNAs		
aa-tRNA	ΔG° (kcal/mol)	Sum (kcal/mol)	aa-tRNA	ΔG° (kcal/mol)	Sum (kcal/mol)
Phe-tRNA^{Phe}	−9.4		Ala-tRNA^{Phe}	−8.4	
		−19.8			−20.4
Ala-tRNA^{Ala}	−10.4		Phe-tRNA^{Ala}	−12.0	
Val-tRNA^{Val}	−9.4		Phe-tRNA^{Val}	−9.6	
		−18.8			−19.3
Phe-tRNA^{Phe}	−9.4		Val-tRNA^{Phe}	−9.7	
Phe-tRNA^{Phe}	−9.4		Phe-tRNA^{Gln}	−9.4	
		−20.0			−20.6
Gln-tRNA^{Gln}	−10.6		Gln-tRNA^{Phe}	−11.2	
Val-tRNA^{Val}	−9.4		Ala-tRNA^{Val}	−8.6	
		−19.8			−19.8
Ala-tRNA^{Ala}	−10.4		Val-tRNA^{Ala}	−11.2	
Ala-tRNA^{Ala}	−10.4		Ala-tRNA^{Gln}	−8.3	
		−21.0			−21.4
Gln-tRNA^{Gln}	−10.6		Gln-tRNA^{Ala}	−13.0	
Val-tRNA^{Val}	−9.4		Val-tRNA^{Gln}	−8.9	
		−20.0			−19.9
Gln-tRNA^{Gln}	−10.6		Gln-tRNA^{Val}	−11.0	

A. The data in table 2 show that for any pair of amino acids and tRNA bodies, the sum of the binding free energy of cognate and noncognate tRNAs are nearly identical

B. For each tRNA, binding affinity to EF-Tu is highest when Alanine (Ala) is esterified, followed by phenylalanine (Phe), then valine (Val), and Glutamine (Gln)

C. In general, for a given esterified amino acid, binding affinity to EF-Tu is greatest when the amino acid is acylated to $tRNA^{Ala}$ while the binding affinity when acylated to $tRNA^{Gln}$ is lowest.

D. The data presented is consistent with the hypothesis that near uniform binding affinity by the 4 correctly acylated tRNAs is consistent with a thermodynamic compensation of the amino acid and tRNA body such that a "tight" amino acid such as Gln is acylated to the "weak" $tRNA^{Gln}$

Now for some easier cool down questions!

43. (1 point) Which of the following levels of structure is most stabilized by hydrogen bonding?

A. Primary

B. Secondary

C. Teritary

D. Quaternary

44. (1 point) What two atoms are linked together in a peptide bond? Select both.

A. Carbon

B. Hydrogen

C. Oxygen

D. Nitrogen

E. Sulfure

F. Phosphorus

45. (1 point) When a protein is heated to very high temperatures it starts to _____ ?

A. Detonate

B. Denature

C. Deflagrate

D. Hydrolyze

46. (1 point) What is the relation between protein function and structure?

A. A proteins structure is uniquely defined by its function

B. A proteins function evolves to fit its structure

C. A protein is structured in a way that allows it to perform a particular function

D. Protein structure and function are independent concepts

47. (1 point) Scientists often add reducing agents to PAGE gels for separating proteins. What purpose does this serve?

A. To break disulfide bonds

B. To break peptide bonds

C. To break electrostatic attractions

D. To hydrolyze the protein

48. (1 point) Valine is a nonpolar amino acid.

A. True

B. False

49. (1 point) Proteins have to fold properly to carry out their function

A. True

B. False

50. (1 point) Proteins are synthesized by translating unprocessed mRNA

A. True

B. False

Solution: While we expected the answer False, it is true that mRNA in prokaryotes does not need to be processed. We did not accept both answers at UT, but scientifically speaking they are both correct.

51. (1 point) Tertiary structures in a protein form to protect a hydrophobic core from water

A. True

B. False

52. (1 point) Proteins with many polar amino acids can be inserted into the lipid membrane easily.

A. True

B. False

53. (1 point) What does CRISPR stand for?

A. Clustered Regularly Interspaced Short Palindromic Repeats

B. Clustered Randomly Inserted Short Palindromic Repeats

C. Clumped Randomly Introduced Selective Partial Repeats

D. Clustered Regularly Introduced Selective Palindromic Repeats

E. Clobbered Random Intronic Sequences Placed Restrictively

54. (1 point) What are the stages of CRISPR-mediated immunity (select the correct order out of the choices provided)?

I) spacer acquisition

II) crRNA processing

III) exonuclease activity

IV) apoptotic expression

V) interference

A. I,II,III
B. I,IV,V
C. I,II,IV
D. I,II,V

55. (1 point) Which of the following would NOT be found in Cas9?
A. Exonuclease Domain
B. REC1
C. HNH
D. RuvC

56. (1 point) Which of the following are risks of CRISPR-based gene editing? Select all that apply.
A. Off-target genomic edits
B. Introduction of mutations (insertions or deletions) at site
C. Spontaneous meiosis
D. Recruitment of bacteriophages
E. Genomic integration of S. pyogenes DNA

57. (1 point) What does the PAM stand for and what nucleic acid type is it found in?
A. PAM associated motif; DNA
B. Protospacer Adjacent Motif; RNA
C. Protospacer Adjacent Motif; DNA
D. PAM associated motif; RNA

58. (1 point) Which of the following can affect protein function?
A. Temperature
B. pH
C. Denaturing agents
D. Solvent identity (water, DMSO, etc)

59. (1 point) The central dogma of molecular biology describes which of the following transitions?

A. DNA – RNA – Protein

B. RNA – DNA – Protein

C. Protein – DNA – RNA

D. RNA – Protein – DNA

60. (1 point) Which of the following mutations could result in a truncated (shortened), nonfunctional protein? Select all that apply.

A. Stop codon introduction

B. Nonsense mutation

C. Missense mutation

D. Silent mutation

61. (1 point) You discover a synthetic protein hormone that has a half-life of 12 hours in blood serum. Assuming there is no production of this hormone in the body, what is the concentration of the protein in serum after 36 hours if it was initially 48 uM?

A. 48 uM

B. 24 uM

C. 12 uM

D. 6 uM

E. 3 uM

62. (1 point) The Nobel Prize in chemistry was recently awarded to _____ and _____ for their contributions to CRISPR-Cas9 genome editing.

A. Feng Zhang ; David Liu

B. Aviv Regev ; Eric Lander

C. Emmanuelle Charpentier ; Jennifer Doudna

D. David Sabatini ; Rodolphe Barrangou

E. Frances Arnold ; George Smith