Ethics pledge

"I agree to complete this quiz without unauthorized assistance from any person. I will not discuss or share the exam questions, answers or any exam-related material with anyone until 9:30 AM East Coast Time, on Tuesday, February 15, 2022."

_____Brendon Davis___2/14/2022_____
Signed

I. Protein structure

IA. Phi and psi are the angles between different planes defined by the backbone of part of a polypeptide. Given an amino acid at position i, phi is the angle between the planes defined by:

 <u>Plane 1</u>
 - The carboxyl carbon on the (i-1)th amino acid
 - The backbone nitrogen of the ith amino acid
 - The alpha carbon of the ith amino acid
 <u>Plane 2</u>
 - The backbone nitrogen of the ith amino acid
 - The alpha carbon of the ith amino acid
 - The carboxyl carbon of the ith amino acid

Similarly, psi is defined as the angle between:

 <u>Plane 1</u>
 - The alpha carbon of the ith amino acid
 - The carboxyl carbon of the ith amino acid
 - The backbone nitrogen of the (i+1)th amino acid
 <u>Plane 2 (same as Plane 2 above)</u>
 - The backbone nitrogen of the ith amino acid
 - The alpha carbon of the ith amino acid
 - The carboxyl carbon of the ith amino acid

Thus, for any given non-terminal amino acid at position i, you can define phi and psi based on these atoms. Each point on the Ramachandran plot represents one amino acid at a given position i with its phi angle plotted on the x-axis and its psi angle plotted on the y-axis.

IB.
(i) This is a loop structure that does not have either of the canonical secondary structures. Regions like this tend to be more flexible, so they may be in any of the allowed phi, psi angles that avoid steric hindrance and are thus favorable. This means that the angles for these amino acids will be found predominately in the two main clusters around [-60°< Φ,Ψ<-40°] and [Φ≃-120°, Ψ≃120°], as well as sometimes in the cluster near [Φ≃75°, Ψ≃30°]. Because the loops are more flexible, they may also have angles that are slightly outside of the normal range for the clusters expected for beta-sheets and alpha-helices.
(ii) This is a right-handed alpha helix, which is known to have ideal angles of around -60°< Φ,Ψ<-40°. This corresponds with the cluster of points just below the x-axis on left hand side of the plot.
(iii) This is a beta sheet, which is known to have ideal angles of Φ≃-120°, Ψ≃120°. This corresponds to the cluster of data points that is predominately in the top left of the Ramachandran plot.

IC. The most likely segment containing the represented element is segment ii (VEGIY). The highlighted point is outside of the clusters that are normally considered favorable. This is typically only possible at a glycine residue, where the lack of side chain (only a hydrogen)

means that there is reduced steric hindrance between the side chain and neighboring amino acids. This allows for more flexibility, which can create phi, psi angle combinations outside of the expected ones, as we see here.

ID. The most likely segment to contain a cis peptide bond would be segment i (TWEPF). While proline does not always have a cis peptide bond, the cis isomer is less unfavorable for proline than it is for other amino acids. This is due to the fact that the side chain of proline loops back to covalently attach to the backbone nitrogen, which can help to stabilize the cis form.

IE. The largest contribution to stability of protein tertiary structure is typically hydrophobic interactions. These hydrophobic interactions are driven by water interacting with itself as well as polar residues on the protein, which creates a hydration shell and causes the hydrophobic residues to sequester at the core of the protein. However, hydration is a temperature-dependent process. Hence, one explanation of why some proteins denature at cold temperatures is that the hydration of non-polar residues becomes thermodynamically favorable at low temperature because the enthalpy of hydration decreases with decreasing temperature. This means that it becomes favorable for the hydrophobic residues to be exposed to water, and thus the main driving force of protein folding is abolished, causing denaturation.

II. Nucleic Acid structure

IIA. In aqueous solution, the atoms that are hydrogen bonding between nucleotide bases can also form hydrogen bonds with water in the solvent. This means that these hydrogen bonding atoms would be stabilized if the duplex melted, so H-bonds between bases are only moderately increasing stability compared to the case where water molecules were making the H-bonds instead.

IIB. Firstly, there are Van der Waals interactions that come from base stacking in the duplex structure. Specifically, London attraction occurs where an instantaneous dipole in one base induces the same dipole in a nearby base, and these forces, though small on their own, can happen many times in the stacked bases to create an attractive, favorable force that stabilizes the duplex.

Additionally, there are ion-ion interactions between the negatively charged phosphates in the backbone of the two nucleotide strands in a duplex. Because these charges are negative on both backbones, they repel each other. Thus, this is an unfavorable force that decreases duplex stability and makes it easier to destabilize the structure.

IIC. This is a guanine-cytosine base pair (G on the left, C on the right). Shown in the drawing are H-bonds between the two bases, in which the $G-N_1$, $G-N_2$, and $C-N_4$ act as H-bond donors while the $G-O_6$, $C-N_3$, and C-O act as H-bond acceptors. In addition to what is shown, the following additional donors and acceptors on each of the grooves are possible:
   - Major groove donors: $C-N_4$
   - Major groove acceptors: $G-N_7$, $G-O_6$
   - Minor groove donors: $G-N_2$
   - Minor groove acceptors: $G-N_3$, C-O

IID. Compared to a B-form helix, the overall ability of these atoms to form additional H-bonds with molecules outside the DNA duplex would be decreased if this was part of an A-form helix. This is because the atoms capable of H-bonding are easily accessible in a B-form helix, particularly in the major groove. However, in an A-form helix, the major groove is narrow and deep, making it harder for molecules to access the H-bonding atoms inside. Partially compensating for this, the minor groove is slightly more accessible in an A-form helix, so the minor groove atoms would have slightly more H-bonding potential, but this does not make up for the severe loss of H-bonding potential for atoms in the major groove.

III. Interactions

IIIA.

a. I would *not* exploit this for either (i) or (ii). Both DNA and RNA have negative phosphate charges on the backbone, and these charges are uniform regardless of base sequence.

b. I would exploit this for (i). DNA duplexes are usually in B-form, meaning they have greater pitch, while RNA duplexes are usually A-form, which have less pitch. However, this is not helpful for recognizing a base sequence because the pitch is the same regardless of sequence.

c. I would *not* exploit this for either (i) or (ii). Both DNA and RNA duplexes experience base pair stacking, and this stacking occurs regardless of sequence.

d. I would exploit this for (ii) *if and only if* the protein also opened the duplex. AT/AU base pairs have two hydrogen bonds while CG base pairs have three H-bonds, so recognizing the atoms that form H-bonds between base pairs would help to define which pair you have. However, the H-bonds between pairs are not accessible in the duplex ground state, so this I would only exploit this if the protein was able to open the duplex (or if it bound to a duplex that had been partially opened by another protein). Additionally, this is not helpful for (i) because DNA and RNA have the same inter-base H-bond donors and acceptors. Uracil (RNA) and thymine (DNA) have the same atoms that form H-bonds with adenine.

e. I would exploit this for both (i) and (ii). The major groove is easily accessible in B-form DNA while it is not as accessible in A-form RNA duplexes. This means that if H-bond donors or acceptors are easily available on the major groove, this is likely a DNA strand. Additionally, this is exploitable for (ii) because the H-bond donor/acceptor order is different on the major groove for each of the four possible base pairs (AT, TA, CG, GC). This means that by having H-bonding atoms in a protein that H-bonded to a specific order of donors/acceptors on the major groove of DNA, the sequence of the DNA could be determined without opening it.

f. I would exploit this for (i) and possibly exploit this for (ii). In A-form RNA, the minor groove is relatively wide and shallow, which allows other molecules to access it, while in B-form DNA, the minor groove is relatively inaccessible. This means the accessibility of H-bonds in the minor groove can distinguish between RNA and DNA. As for (ii), the minor groove of A-form RNA does have unique H-bond acceptor/donor orders for the different base pairs, but they are fairly similar to each other. This means that the h-bonding atoms on the minor groove face could be used to recognize sequence information, but this would likely be slightly unreliable.

IIIB.

(i) For (a), all of the residues can form H-bonds through their backbones, since the carboxyl oxygen can be an H-bond acceptor and the backbone nitrogen can be an H-bond donor.
For (b), the residues that can form side chain H-bonds are 2, 4, 7, 8, 10, 11, 12. Additionally, the methionine at position 3 can form a relatively weak H-bond, as the sulfur in methionine has been shown to act as a very weak H-bond acceptor at times.

(ii) Solving for the standard free energy of binding, we can find that
$\Delta G^0_{bind}=RT*\ln(K_D)$. Plugging in numbers we get
$$\Delta G^0_{bind}=(8.314\times10^{-3})(298)*\ln(12.5\times10^{-9})=-45.09 \text{ kJ/mol}$$
Hence, the standard free energy of binding is about -45.09 kJ/mol.

(iii) For clarity I will use G as a shortening of ΔG^0_{bind}.
$$K_D=e^{G/(RT)}=e^{-40.09/((8.314E-3)(298))}=9.389\times10^{-8}$$
Hence, the K_D of this mutant form is about 938.9 µM.

(iv) This suggests that the given amino acid and the given DNA position interact together to help stabilize the overall interaction between the protein and DNA target. Given the K233A mutation and the sequence specificity of the interaction, it is likely that the wildtype interaction involves H-bonding between the lysine and the DNA atoms accessible in the major groove of the AT base pair. Energetically, the overall interaction is stabilized by 5 kJ/mol if this interaction exists, and the overall interaction is not stabilized by this amount if the interaction does not exist. This all-or-nothing behavior explains why either one of the mutations or both at the same time result in the same amount of destabilization.

IIID.

(i) We know that U_e, the potential energy of an electrostatic interaction, is equal to $1389(q_1q_2)/(\varepsilon r)$. Therefore,
$$U_e=1389(1*-1)/(80*6)=-2.89 \text{ kJ/mol}$$

(ii) Using the Boltzmann distribution and a given ratio n of non-interacting to interacting particles, we can say
$$n=e^\wedge(-\Delta U/(k_BT))$$
$$\Delta U=-k_BT*\ln(n)$$
I will define stable as being that more than 90% of all possible interactions are in an interacting state at 298°K, so n=1/9. Thus,
$$\Delta U=-(8.314\times10^{-3})(298)*\ln(0.9)=5.44 \text{ kJ/mol}$$
Since the magnitude of the potential energy of the interaction is less than the energy needed to maintain at least 90% of particles in the interacting state, this is *not* a stable interaction under this definition of stable.

(iii) First calculate the change in energy for the ion-ion interaction:
$$\Delta U_e=1389(1*-1)/(80*5.5)-(-2.89)=-0.267 \text{ kJ/mol}$$

Next calculate the change in energy of the bond stretching:
$$U_{bond}=0.5*500*(1.9-1.4)^2=62.5 \text{ kJ/mol}$$

This suggests it would be highly unfavorable to bring the two charges together by stretching a bond 0.5 Å. This is because the energy increase of the bond stretch is much greater than the energy decrease of the electrostatic interaction.

(iv) - For part (i): $U_e=1389(1*-1)/(2*6)=-115.75 \text{ kJ/mol}$
- For part (ii): The magnitude of the energy of this interaction (115.75 kJ/mol) is now much greater than the energy required to maintain at least 90% of the particles in the interacting state (5.44 kJ/mol). Thus, this is now a stable interaction.
- For part (iii): $\Delta U_e=1389(1*-1)/(2*5.5)-(-115.75)=-10.52 \text{ kJ/mol}$

As calculated before, U_{bond}=62.5 kJ/mol, so because the energy decrease of the electrostatic interaction (10.52) is still less than the energy increase caused by the bond stretch (62.5 kJ/mol), it remains unfavorable to stretch the bond to bring the two interacting ions closer together.