Problem 1. (20 points)
A protein contains an α-helix with the following amino acid sequence:

ADEWLQKVHAMFINR

A. (5 points) How many full turns are in this α-helix?

One turn contains 3.6 residues; therefore 15 residues form 4 full turns (15/3.6 = 4.2).

B. (4 points) Which of the following mutations in the original sequence is likely to most strongly destabilize the helix? *Explain your reasoning*.

1. D, E, K, R → A
2. E, M → K
3. D or E → P
4. F, L, M, W, I → Q

The answer is 2: E, M → K. Here is why:

1. D, E, K, R → A

This will remove the charges on the helix by replacing those residues with Ala, but this won't destabilize it strongly.

2. E, M → K

This will result in positive charges placed close in space to each other (at almost similar positions on consecutive turns) – their electrostatic repulsion **will destabilize the helix**

3. D or E → P

Proline in the first turn of the α-helix should not significantly perturb it because it doesn't need to form H-bonds.

4. F, L, M, W, I → Q

This will replace all hydrophobic residues with polar ones but won't destabilize the helix.

The amino acid sequence is repeated here for your convenience

ADEWLQKVHAMFINR

C. (6 points: 2+2+2) List the residues that form backbone hydrogen bonds with the tryptophan, with the lysine, with the phenylalanine. For each of these residues indicate what chemical group is involved in each particular hydrogen bond (i.e. what group/atom is the donor and what is the acceptor).

Tryptophan: forms one H-bond: between its CO and the NH of V.

Lysine: forms 2 H-bonds: (1) between its NH and CO of residue E and (2) between its CO and NH of M.

Phenylalanine: forms one H-bond: between its NH and the CO group of V.

Obviously, in all these cases, the NH group serves as a donor and O in CO group as an acceptor of a hydrogen bond.

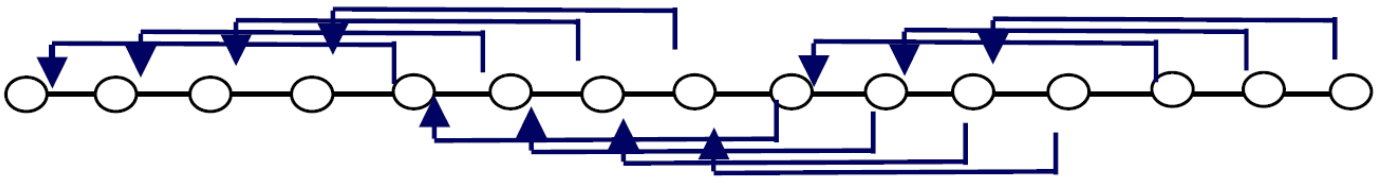
D. (5 points) How many hydrogen bonds between the backbone atoms are in this helix? *Explain how you arrived at your answer*.

The answer is **11**.

Below I list a few ways to get this answer. There can be other ways.

Perhaps the simplest way to figure this out is to use the notion that an H-bond connects residues ***i*** and ***i***+4, and start counting the bonds, for example, from the N terminus in the direction of the C terminus. The first residue is connected to residue 5: this gives one H-bond. The second residue is H-bonded with residue 6: second bond. And so on, unless you reach residue 11, which is the final count because it's connected to residue 15. Residue 12 does not form an H-bond in the direction toward the C terminus, and its H-bond directed backwards has already been counted. This gives 11 bonds.

You can also draw the H-bond connections (between residues ***i*** and ***i+4***) and count them. An example is shown here, where the acceptors (CO) are indicated by an arrow.



You can also figure this out using the following simple considerations. 1 residue can be involved in maximum 2 H-bonds, therefore 15 residues can make up to 2x15=30 H-bonds. In the α-helix, 4 residues at the N-terminus and 4 at the C-terminus make only 1 bond per residue. This makes the total number of H-bonds 30-2x4=22. When calculating this number, each H-bond was counted twice: one time for the donor residue and one time for the acceptor. The real number of H-bonds is then 22/2=11.

Problem 2. (19 points)

When examining ligand binding properties of a protein P, you found that at 40 μM concentration of ligand A 40% of the protein was bound. By contrast, 80% of the protein was bound at 80 μM concentration of ligand B. Both ligands bind to the same site on the protein, hence their binding is mutually exclusive. Answer the following questions. Explain your assumptions and show your calculations/reasoning.

A. (11 points: 5+5+1 for spelling out at least one assumption) Which ligand has higher affinity for the protein? Calculate the dissociation constants to support your answer.

Assumptions: (1) binding is noncooperative; (2) one ligand binding site per protein (i.e. stoichiometry is 1:1); (3) $[L_{total}] >> [P_{total}]$, so that $[L] = [L_{total}]$.

For A: $K_{dA} = \frac{(1-\theta)}{\theta}[\mathrm{L}] = \frac{(1-0.4)}{0.4}[\mathrm{L}] = \frac{3}{2}[\mathrm{L}] = 60\ \mu\mathrm{M}.$

For B: $K_{dB} = \frac{(1-\theta)}{\theta}[\mathrm{L}] = \frac{(1-0.8)}{0.8}[\mathrm{L}] = \frac{1}{4}[\mathrm{L}] = 20\ \mu\mathrm{M}.$

These data indicate that B has higher affinity for the protein than A

B. (8 points) Based on your answers to question A, calculate the ratio of the amounts of the protein in complex with ligand A and with ligand B when both ligands are present in the same solution at equal concentrations. (Hint: answering this question might require some thinking)

Note that a simple equation like $\theta = \frac{[L]}{[L]+K_d}$ doesn't work here because it was obtained assuming that the protein is partitioned between only 2 states: free protein and ligand-bound. Now we have 3 states of the protein, including A-bound and B-bound states. Assume, as above, that $[A] = [A_{total}]$, $[B] = [B_{total}]$. Besides, we know that $[A_{total}] = [B_{total}]$ but we do not know the actual value of the ligand concentration.

The simplest way to answer the question is as follows. We can relate [PA] and [PB] to [P] through the binding equilibria and the respective K_d values as follows: $[PA] = [P][A_{total}]/K_{dA}$ and $[PB] = [P][B_{total}]/K_{dB}$. From these two equations (after dividing the first equation by the second) we get $[PA]/[PB] = [P]/[P] \times [A_{total}]/[B_{total}] \times K_{dB}/K_{dA}$. Taking into account that $[A_{total}] = [B_{total}]$, we get $[PA]/[PB] = K_{dB}/K_{dA} = 20/60 = 1/3$. Note that this solution doesn't depend of what the actual concentrations of the ligands are, as long as they are equal (in this case).

A complete consideration should include all states of the protein: $[P_{total}] = [P] + [PA] + [PB]$. Substituting here [PA] and [PB] from the above equations gives: $[P_{total}] = [P] + [P][A]/K_{dA} + [P][B]/K_{dB} = [P](1 + [A]/K_{dA} + [B]/K_{dB})$. Solving it for [P] gives: $[P] = [P_{total}]/(1 + [A]/K_{dA} + [B]/K_{dB})$. From here

$[PA] = [P][A]/K_{dA} = [P_{total}][A]/K_{dA}/(1 + [A]/K_{dA} + [B]/K_{dB})$

$[PB] = [P][B]/K_{dB} = [P_{total}][B]/K_{dB}/(1 + [A]/K_{dA} + [B]/K_{dB})$

And their ratio: $[PA]/[PB] = [A]/[B] \times K_{dB}/K_{dA} = K_{dB}/K_{dA} = 20/60 = 1/3$.

One can also use the result we obtained for competitive inhibition of enzymes, for example:

$[PA] = [P_{total}][A]/\{K_{dA}(1 + [B]/K_{dB}) + [A]\} = [P_{total}][A]/K_{dA}/(1 + [B]/K_{dB} + [A]/K_{dA})$

Note that these results give $\theta_A = [PA]/[P_{total}] = [A]/(K_{dA} + [A] + [B]\,K_{dA}/K_{dB})$ which differs from the result of single-ligand binding: $\theta_A = [PA]/[P_{total}] = [A]/(K_{dA} + [A])$

Problem 3. (15 points)

A particular chemical reaction, S ↔ P, is characterized by the difference in the free energy $\Delta G'^{o}_{S\to P} = -20$ kJ/mol at standard conditions. However, the uncatalyzed reaction is very slow because of the high activation barrier, $\Delta G^{\ddagger} = 50$ kJ/mol. Therefore, you want to design an enzyme to speed up the reaction.

A. (5 points) What should be the activation barrier, $\Delta G^{\ddagger}_{cat}$, for the catalyzed reaction in order to achieve a 5000-fold increase in the reaction rate at standard conditions?

Using the relationship between the rate of the reaction and the activation barrier, $k = \frac{k_B T}{h}\exp\left(-\frac{\Delta G^{\mp}}{RT}\right)$, we get

$\frac{k_{catalyzed}}{k_{uncatalyzed}} = \exp\left(-\frac{\Delta G^{\mp}_{cat} - \Delta G^{\mp}_{uncat}}{RT}\right) = \exp\left(-\frac{\Delta\Delta G^{\mp}}{RT}\right)$ = **5000**. From here:

$\Delta\Delta G^{\ddagger} = \Delta G^{\ddagger}_{cat} - \Delta G^{\ddagger}_{uncat} = -RT\ln(k_{catalyzed}/k_{uncatalyzed}) = -RT\ln(5000) = -8.5172\times RT = -21.086$ kJ/mol.

(or $\Delta G^{\ddagger}_{uncat} - \Delta G^{\ddagger}_{cat} = RT\ln(k_{catalyzed}/k_{uncatalyzed}) = RT\ln(5000) = 8.5172\times RT = 21.086$ kJ/mol.)

Therefore $\Delta G^{\ddagger}_{cat} = \Delta G^{\ddagger}_{uncat} - 21.086$ kJ/mol = **28.914 kJ/mol**.

B. (5 points) What will be the change in the activation barrier and in the rate of the backward reaction, P → S?

The activation barrier for the backward reaction will change by the same amount ($\Delta\Delta G^{\ddagger} = -21.086$ kJ/mol) because this reduction in the activation barrier affects both the forward and backward reactions similarly:

$\Delta G^{\ddagger}_{uncat(B\to A)} = \Delta G^{\ddagger}_{uncat(A\to B)} + \Delta G_{B\to A} = \Delta G^{\ddagger}_{uncat(A\to B)} - \Delta G_{A\to B} = 70$ kJ/mol.

$\Delta G^{\ddagger}_{cat(B\to A)} = \Delta G^{\ddagger}_{cat(A\to B)} + \Delta G_{B\to A} = 28.914$ kJ/mol + 20 kJ/mol = 48.914 kJ/mol.

$\Delta\Delta G^{\ddagger}_{(B\to A)} = \Delta G^{\ddagger}_{cat(B\to A)} - \Delta G^{\ddagger}_{uncat(B\to A)} = 48.914$ kJ/mol - 70 kJ/mol = - 21.086 kJ/mol = $\Delta\Delta G^{\ddagger}$.

Because $k_{catalyzed}/k_{uncatalyzed}$ depends on the energy only through $\Delta\Delta G^{\ddagger}$, **the rate of the backward reaction will also increase 5000 fold**.

C. (5 points) Estimate the K_d for substrate binding to the enzyme, assuming that all the energy gained from the binding is used to speed up the reaction.

$K_d = \exp(\Delta G/RT)$. If $\Delta G = \Delta\Delta G^{\ddagger} = -21.097$ kJ/mol, then $K_d = \exp(\Delta\Delta G^{\ddagger}/RT) = 2\times10^{-4}$ M = 200 μM

Problem 4. (15 points)

Both myoglobin and hemoglobin utilize heme to bind oxygen, the tertiary structure of myoglobin is very similar to that of the individual subunits in hemoglobin. And yet, the biological functions of the two proteins are very different. Based on what you learned, answer the following questions.

A. (5 points) What are the principal structural differences between these two proteins that are responsible for the difference in their function?

Myoglobin is a **monomer**, therefore, conformational changes caused by O_2 binding to one molecule of myoglobin **do not affect other myoglobin molecules** in solution. Hemoglobin is a **tetramer** (quaternary structure) in which the monomers/subunits **interact with each other**, such that conformational changes caused by O_2 binding to one **subunit are transmitted to other subunits through inter-subunit interactions**, thus affecting O_2 binding to the other subunits. These allosteric effects are responsible for the cooperativity of O_2 binding to hemoglobin.

B. (5 points) What is the principal difference in the mode of oxygen binding to hemoglobin and to myoglobin?

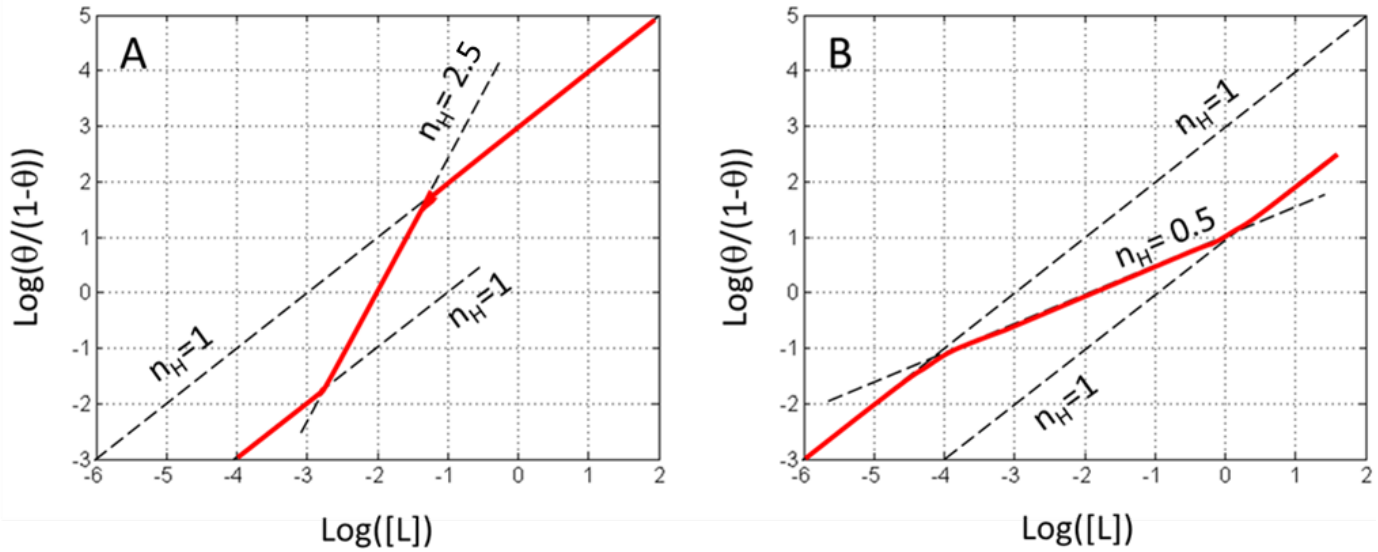
O_2 binding to myoglobin is **non-cooperative**. O_2 binding to hemoglobin is **cooperative**.

C. (5 points) All the following molecules/ions, when present in the blood, affect the ability of hemoglobin to bind oxygen: CO, CO_2, H^+, and 2,3-bisphosphoglycerate (BPG). What is the principal difference in the mechanism of how they affect oxygen binding to hemoglobin?

CO binds directly to the heme, and in addition to blocking the heme in the subunit where CO is bound, this binding stabilizes/causes a transition of hemoglobin to the R state. Because this binding is tight, much tighter than O_2 binding, the equilibrium is shifted towards the R-state. As the result, hemoglobin gets "locked" in the R state: it binds O_2 with high affinity and very efficiently, *but* does not release it in the tissues.
All other molecules/ions listed above bind elsewhere in the structure, but not to the heme, and this binding stabilizes the T state, thus weakening O_2 binding. Thus CO acts as positive heterotropic modulator while the other compounds act as negative heterotropic modulators.

Problem 5. (19 points) (see also next page)

Ligand binding to hypothetic proteins A and B is characterized by the Hill plots shown below. The ligand concentrations are in molar units, and log means $\log_{10}$. Answer the following questions.



A. (5 points) Based on the shape of the binding curves explain how the ligand concentration affects the affinity of protein A and protein B. What conclusions about possible cooperativity of binding can you draw from these plots?

Both proteins show **cooperative** ligand binding, but the cooperativity is different. A shows **positive cooperativity** ($n_H > 1$) while B exhibits **negative cooperativity** ($n_H < 1$).

This determines the shape of the binding curve:
A: as ligand concentration increases, because of cooperativity ligand binding causes a transition from low-affinity to a high-affinity binding mode, and eventually at high ligand concentration protein A binds the ligand with high affinity.
B: as ligand concentration increases, because of cooperativity ligand binding causes a transition from high-affinity to a low-affinity binding mode, and eventually at high ligand concentration protein B binds the ligand with low affinity.

B. (4 points: 2+2) Based on these plots, what is the minimal number of ligand binding sites on the protein molecule for each of the proteins? *Explain your reasoning*.

The number of binding sites is equal or greater than n_H. For protein A, $n_H = 2.5$, so the minimal integer number $\geq n_H$ is 3.
For protein B ($n_H = 0.5$), the minimal number of ligand binding sites is 2. The minimal integer $\geq n_H$ is 1, but at least two binding sites are required for cooperativity.

So the answers are: **3 sites for A and 2 sites for B.**

C. (4 points) Which of the two proteins has higher affinity for the ligand at the midpoint of the binding curve (i.e. where $\theta = 0.5$)? *Explain your reasoning*.

The answer: at the midpoint of the binding curve protein A binds ligand tighter than protein B.

A reminder: Hill plot: $\log\frac{\theta}{1-\theta} = n_H \log[\mathrm{L}] - \log K_d$. At θ = 0.5, the left-hand side is 0, therefore $K_d = ([\mathrm{L}]_{0.5})^{n_H}$.

So, the easiest way to answer this question is to determine the value of $[L]_{0.5}$ and then convert it into the respective K_d value.

For both proteins $[L]_{0.5} = 10^{-2}$ M. Qualitatively, because the value of $[L]_{0.5}$ is less than 1, $([L]_{0.5})^{2.5} < ([L]_{0.5})^{0.5}$. Quantitatively: using the experimental values of n_H and $[L]_{0.5}$ we get for protein A:
$K_d = (10^{-2}\ \mathrm{M})^{2.5} = 10^{-5}\ \mathrm{M}^{2.5}$ while for protein B the K_d is significantly greater: $K_d = (10^{-2}\ \mathrm{M})^{0.5} = 10^{-1}\ \mathrm{M}^{0.5}$.
Therefore we can conclude that protein A binds ligand tighter at the midpoint of the binding curve.

Alternatively, we can compare the ΔG^o values associated with binding. Substituting $K_d = ([\mathrm{L}]_{0.5})^{n_H}$ into the equation for ΔG^o gives: $\Delta G^o = RT \ln(K_d) = RT \ln(([L]_{0.5})^{nH}) = n_H RT \ln([L]_{0.5})$. Because $\ln([L]_{0.5})$ is the same for both proteins, we obtain: $\Delta G^o_A/\Delta G^o_B = n_{HA}/n_{HB} = 2.5/0.5 = 5$. And because $\ln([L]_{0.5}) = -2\ln(10)$ is negative both ΔG^o_A and ΔG^o_A are negative, therefore $\Delta G^o_A < \Delta G^o_B$.

The same conclusions can be obtained from the intercept of the middle-part (or its extension) of the graph with the Y-axis (log[L]=0): according to the Hill equation, the ordinate of this point gives $-\log K_d$. We thus get for $-\log K_d$ the values of approximately 3 (A) and 1 (B).

Note: just the fact that the slope (n_H) in the Hill plot is greater for A than for B means higher cooperativity, *not* necessarily higher binding affinity.

D. (6 points: 1.5*4) What are the K_d values at the very low and very high ligand concentrations for each protein.

At both the low and high ligand concentrations the binding is non-cooperative (the Hill coefficient n_H=1), thus $K_d = [L]_{0.5}$.

For protein A, the n_H=1 lines intercept with the $\log\frac{\theta}{1-\theta} = 0$ axis at $\log[L]_{0.5} = -1$ at low ligand concentration and at $\log[L]_{0.5} = -3$ at high ligand concentration. This converts into $K_d = 10^{-1}$ M and $K_d = 10^{-3}$ M, respectively, corresponding to transition from low-affinity to high-affinity state of protein A as [L] increases.

For protein B, the n_H=1 lines intercept with the $\log\frac{\theta}{1-\theta} = 0$ axis at $\log[L]_{0.5} = -3$ at low ligand concentration and at $\log[L]_{0.5} = -1$ at high ligand concentration, i.e. in opposite order as for protein A. This converts into $K_d = 10^{-3}$ M and $K_d = 10^{-1}$ M, respectively, corresponding to transition from high-affinity and low-affinity state of protein B.

Problem 6. (12 points)

You need to study a macromolecular system structurally and you are deciding which method to use.

A. (3+3+3 points) What are the features of the macromolecular system that can make it unsuitable for the structural methods listed below?

X-ray crystallography

Structural/conformational flexibility that could prevent crystallization, low solubility, poor stability for days.

Nuclear Magnetic Resonance

Large size of the molecule, low solubility, poor stability for days/weeks

CryoEM

Small size of the molecule (< 100 kDa), structural/conformational flexibility will affect class averaging.

B. (3 points) What structural method(s) that we covered in this course can be helpful if you want to characterize the unfolding of a protein?

Circular dichroism can monitor loss of secondary structure.

SAS can sense/monitor transition from compact/folded to extended/unfolded state.

NMR: chemical shifts reflect differences in local environment in a folded protein. When protein unfolds these differences disappear and the spectra become less spread.