## Instructions:

- Turn in your results in some kind of clear form (handwritten, typed, pdf).
- Feel free to work in groups.
- I have tried to be consistent in my language in my prompts if I want something specific.
  - *Sketch:* Hand draw a plot. Axes should be labeled conceptually, with key features indicated and explained.
  - *Generate a plot:* Plot using software. Label axes, use appropriate significant figures, etc.
  - *Calculate:* Actually calculate numbers using math. Report units and uncertainty, as appropriate.
  - *Describe/Argue:* Use any combination of writing, sketching, plotting, and calculation to argue for an interpretation.
  - Finally, I will sometimes specify the sort of explanation I want.
    - *Molecular*: Describe what the atoms and molecules are doing in space and time. Depending on the context of the question, this might also involve explaining the result in terms of atomic properties like hydrophobicity.
    - *Energetic*: Describe in energetic terms (entropy, free energy, statistics).
    - *Mathematical*: Answer in terms of how the functions behave. For example, if I asked "mathematically, why does $Kx/(1+Kx)$ saturate with increasing $x$" the answer would be: "because as $x \to \infty$, $Kx \gg 1$ and the function tends to 1."
- Some of the questions may require you to play with math and/or ideas we did not explicitly discuss in class. This is intentional. Many times in science you will be faced with a paper that uses an approach you are not familiar with. Learning how to gather enough information to critically evaluate their findings, as well as understand any mathematical models employed, is important.
- Some questions are listed as **GRAD STUDENT** questions. Undergrads are free to do those questions, but it is not required.

# 1 Simple Statistical Thermodynamics

$$\Delta G^\circ = -RTln(K_{eq})$$

$$K_{eq} = \frac{[product(s)]}{[reactant(s)]}$$

$$S = -R \sum_{i=1}^{i \leq W} p_i ln(p_i)$$

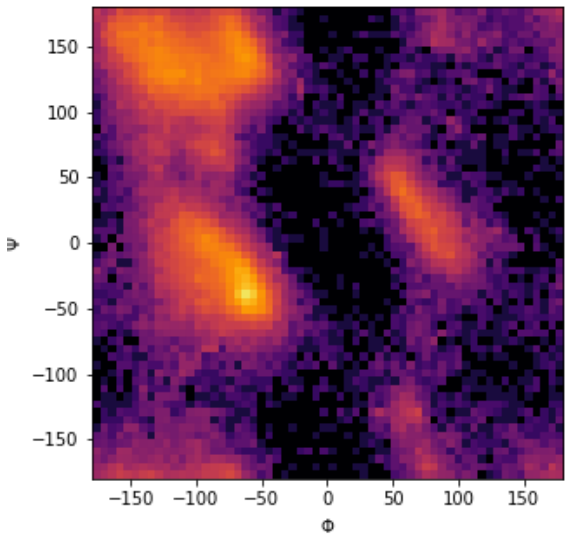
1. Consider a unimolecular reaction, $A \overset{K_{eq}}{\rightleftharpoons} B$, in which $\Delta G^\circ$ favors the formation of $B$. You start the reaction with $[A] = A_0$ and $[B] = 0$.
   (a) Sketch a curve of $[B]$ vs. time ($t$) for the un-catalyzed reaction. Explain why the curve has this shape in terms of the behavior of the molecules.
   (b) As $t \to \infty$, what value does $[B]$ approach, and why?
   (c) You add an enzyme which catalyzes the reaction. Sketch a new curve for this reaction on the same graph as before. As $t \to \infty$, do the curves reach the same value? Why or why not?

2. Now consider a different unimolecular reaction, $C \overset{K_{eq,C\to D}}{\rightleftharpoons} D$, with $\Delta G^\circ_{C\to D} = 10\ kJ \cdot mol$ (that is, it mildly *disfavors* the formation of $D$).
   (a) You start with $[C] = 10\ \mu M$, $[D] = 0\ \mu M$. As $t \to \infty$, what is the concentration of $C$ and $D$ in solution?
   (b) Would the addition of an enzyme that catalyzes this reaction change the outcome? If not, what would addition of the enzyme change about this process?
   (c) Now, imagine that you add a second enzyme that catalyzes the conversion of $D \overset{K_{eq,D\to E}}{\rightleftharpoons} E$, where $\Delta G^\circ_{D\to E} = -100\ kJ \cdot mol^{-1}$ (it strongly favors formation of $E$).
      i. As $t \to \infty$, what (qualitatively) happens to the concentrations of $C$ and $D$?
      ii. Calculate the $[D]_{t\to\infty}$.

3. You are studying a protein that favors the folded state by $-32.3\ kJ \cdot mol^{-1}$. You introduce a mutation that makes a new ion pair worth $-11\ kJ \cdot mol^{-1}$ formed only in the folded state.
   (a) What fraction of protein molecules are in the folded state for the wildtype protein?
   (b) What fraction of the protein molecules are in the folded state for the mutant protein?

4. An unfolded protein occupies many conformations; a folded protein occupies relatively few. Understanding how an amino acid chain finds its native structure out of the astronomical number of possibilities requires some notion of what conformations an unfolded protein occupies. It is, however, difficult to charactrerize this directly by techniques like crystallography, NMR, or cryo-EM. To get around this problem, Fitzkee and colleagues took tens of thousands of experimentally determined protein structures and extracted the "coil" regions. These regions are the floppy, loopy regions that are not part of helices ($\alpha$, $3_{10}$, or $\pi$), $\beta$-sheets, or $\beta$-turns. They then calculated a histogram of the backbone dihedral angles ($\Phi$ and $\Psi$) for every amino acid in the dataset.[1] This is (arguably) a good model for the likely conformations taken a residue in an unfolded protein. A heat map of the histogram follows, where the colors range from black (low probability) to yellow (high probability).

[1] Fitzkee et al. (2005) *Proteins*. https://doi.org/10.1002/prot.20394. If you need a refresher on these angles, see https://proteinstructures.com/Structure/Structure/Ramachandran-plot.html.




(a) Assume that, in the unfolded state, each amino acid samples $\Phi, \Psi$ angles with the same probability they are observed in protein coil regions. Calculate the entropy change for taking an amino acid from this unfolded distribution to a single folded conformation. For simplicity, you may assume that the folded conformation occupies a single grid square on the histogram. (The grid is $60 \times 60$. The probability values are given in *coil.csv*.)

(b) What would the entropy change upon folding be if, rather than sampling the distribution above, an unfolded amino acid could take any pair of $\Phi, \Psi$ values with equal probability. As above, you may assume that the folded conformation occupies a single grid square, and that there are $60 \times 60$ possible pairs of $\Phi$ and $\Psi$.

(c) How might this help explain how proteins find their native structure?

(d) What other factors might need to be accounted for to properly determine the entropy change on protein folding?

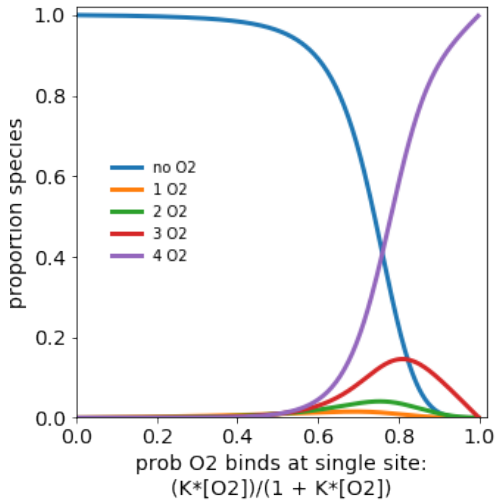
5. Many proteins have multiple binding sites for ligands. The most famous example might be hemoglobin, with its four $O_2$ binding sites.

(a) Generate a plot of the probability that a hemoglobin tetramer has 0, 1, 2, 3, or 4 $O_2$ molecules bound vs. the probability that an individual subunit binds $O_2$. You'll want to use this expression, discussed in the Simple Statistical Thermodynamics for Biochemists packet.

$$P(n,k) = \binom{n}{k} p^k (1-p)^{n-k}$$

(b) Describe what $n$, $k$, $\binom{n}{k}$, and $p^k(1-p)^{n-k}$ mean when modeling hemoglobin $O_2$ binding.

(c) The plot below shows the measured proportion of hemoglobin tetramers with 0, 1, 2, 3, or 4 $O_2$ molecules bound vs. the probability that an individual subunit binds $O_2$.




Does your plot reproduce this plot? Can you rationalize any differences between your predicted and this observed plot?

6. Explain, in molecular terms:

   (a) What is heat?

   (b) Why do both increasing the temperature and decreasing the temperature of a system lead to energy being converted to heat?

   (c) What is enthalpy?