1. **Equilibrium binding of two species:** A protein (P) can bind with a short piece of DNA (D) to form the complex PD. Assume that there is fixed total amount of P and D, with concentrations $[P]_{tot}$ and $[D]_{tot}$ respectively. Find a general expression for the concentration of the complex, $[PD]$. We will show that when $[P]_{tot} \gg [D]_{tot}$, we recover the familiar form given in class.

   (a) Write down the dissociation constant K_d in terms of the free concentrations of P, D and the complex PD.

   (b) Write down the concentrations of total P and total D ($[P]_{tot}$, and $[D]_{tot}$) in terms of the free concentrations and the concentration of PD.

   (c) Use your results from **(b)** to express the expression for K_d from **(a)** in terms of $[P]_{tot}$, $[D]_{tot}$ and $[PD]$.

   (d) Solve the expression for K_d from **(c)** for $[PD]$.

   (e) Suppose $K_d = 10$ nM. Plot the fraction of DNA occupied, $f \equiv [PD]/[D]_{tot}$, against $[P]_{tot}$ for $[P]_{tot} = 1 - 100$ nM, for $[D]_{tot} = 1, 10, 100$ nM.

   (f) Show that for $[P]_{tot} \gg [D]_{tot}$, the fraction of DNA occupied is given approximately by $f \approx [P]_{tot}/(K_d + [P]_{tot})$.
   (Hint: It will be useful to use Taylor's expansion $\sqrt{1-x} \approx 1 - x/2$ for $x \ll 1$.)

2. **Protein-protein interactions:** Two transcription factors A and B can bind in solution to form the complex AB with a dissociation constant of $K_{AB} = 2$ μM. Transcription factor A is maintained at the level of 500 molecules per cell *in the free form*. Only the complex AB can bind to a target site in the genome of *E. coli*, characterized by an effective dissociation constant $K_{AB-DNA} = 6$ nM.

   (a) Write equilibrium constants K_{AB} and K_{AB-DNA} in terms of the relevant protein concentrations.

   (b) What should the cellular level of proteins B be in order for the target site be occupied 5% and 95% of the time?

   (c) What would be the corresponding synthesis rates of protein B be if B is not degraded and the cells are growing exponentially at 60 min per doubling. (You may neglect the occurrence of multiple targets within a cell.)

   (d) If instead, protein A can bind specifically to the target site by itself, with $K_{A-DNA} = 6$ nM, but protein B can bind to protein A only when A is already bound to the target site, what should the level of protein B be in order for the target site be occupied by

AB 5% and 95% of the time?

3. **Binding energy matrix:** Mnt is a dimeric transcription factor which binds to a 17bp DNA segment. The binding energy matrix $G_i(b)$ was measured by the Stormo lab and is reproduced below for the half site from position 10 to 17. The binding energies for the other half (position 1-8) can obtained as the *reverse complement* of those shown here. Position 9 is a neutral position which does not affect Mnt-DNA binding.

Table 1: Binding energy matrix

position	10	11	12	13	14	15	16	17
A	1.8	2.4	1.6	1.0	0	2.1	0.8	1.1
C	2.4	1.9	4.2	2.1	0.3	0	0	0
G	0	1.6	0	0	1.2	3.2	1.0	1.2
T	3.0	0	2.2	2.2	0.6	2.2	0.7	0.3

(The numbers are expressed in units of $k_B T \approx 0.6$ kcal/mole)

(a) Given that $G^{ns} - G^* \approx 16 k_B T$, find the effective dissociation constant $\widetilde{K}^*$ for the strongest binder in the presence of genomic DNA that is $5 \cdot 10^6$ bp in length.

   [Hint: approximate the genomic DNA as a random string of nucleotides with equal distribution of $\{A, C, G, T\}$.]

(b) Approximate the non-zero entries of the binding energy matrix by one parameter, ϵ, and find the smallest value of ϵ that would result in the same $\widetilde{K}^*$.

(c) The target sequence is located within the following segment of DNA

   5' - TCTACGATCCACTGTCGACTCGACTGCCGTAT - 3'

   Compute and plot the binding energy $G_j = \min(G_j^{sp}, G^{ns})$ as a function of the position j, the position in the sample sequence that the first position of the Mnt motif aligns to. Repeat the plot for $G^{ns} - G^* \approx 30 k_B T$. Attach your computer code (or show your method if performed otherwise) and comment on your findings.

4. **Multiple target sites.**
   [Those who have not had a course in Statistical Mechanics need not do this problem.]

   Suppose a transcription factor can bind to N_0 distinguishable target sites with the same specific binding free energy G_0, and N additional distinguishable background sites with the non-specific binding free energy G_{ns}. Suppose that there are M TF molecules in the cell, with $N \gg \{M, N_0\}$, and assume all TFs are associated with the DNA. Derive an expression for the probability $P_A(M; N, N_0)$ that a particular target site, site A, is occupied by the following procedure.

(a) Write down the Boltzmann weight $W(M - m, m)$ that $M - m$ proteins are bound to the background sites and m proteins are bound to the target sites.

(b) Given that m proteins are bound to the N_0 target sites, what is the probability, $f(m, N_0)$, that the site A is occupied?

(c) Argue that P_A is given by

$$P_A(M; N, N_0) = \frac{\sum\limits_{m=0}^{N_0} f(m, N_0)W(M - m, m)}{\sum\limits_{m=0}^{N_0} W(M - m, m)}$$

Simplify this expression for $M \gg N_0$ and cast it into the form $P_A = \dfrac{1}{1 + \widetilde{K}_A/M}$

[Hint: For $N \gg n$, $N!/(N - n)! \approx N^n$]

Find $\widetilde{K}_A$ in terms of the other variables introduced above. On which variable does it not depend for $M \gg N_0$?

(d) Would the dependence of $\widetilde{K}_A$ on N_0 affect the "programmability" of the binding affinity?

(e) Finally, we'll see just how large M needs to be before the above approximation becomes valid. Write the exact expression for $\widetilde{K}_A$ starting from the full form of P_A for $N_0 = 2$ and show that it is dependent on M.

For $N = 10^7$ bp, and $G^{ns} - G_0 = 16k_B T$, plot $\widetilde{K}_A$ as a function of M for $N_0 = 2$. Find the range of M where $\widetilde{K}_A$ approaches the M-independent limit.