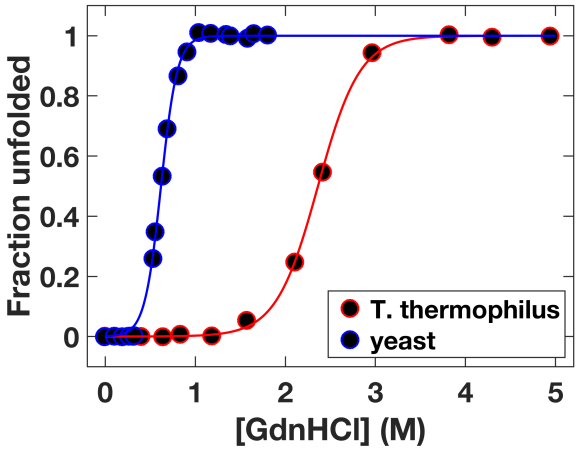
# Solutions to problem set 3

## Problem 1

**Two-state protein folding, continued.**

a) Looking at the data, one can quickly see that that rough midpoint (i.e. the point where 50% are folded and 50% denatured) of the transition is $\approx 0.5$ M GdnHCl for the yeast PGK, while it is $\approx 2.5$ M GdnHCl for the *T. thermophilus* variant. The fact that one needs a lot higher concentration of denaturant to unfold the *T. thermophilus* PGK variant, suggests that it is more stable, i.e. has a more negative $\Delta G_0$, than the yeast variant. This also seems overall plausible, given that thermophilic organisms often have more stable proteins than organisms that grow at lower temperatures, to avoid thermal denaturation.

b) Below is a plot of the data with the fits as solid lines. I have co-plotted both curves in the same coordinate system to facilitate the comparison.



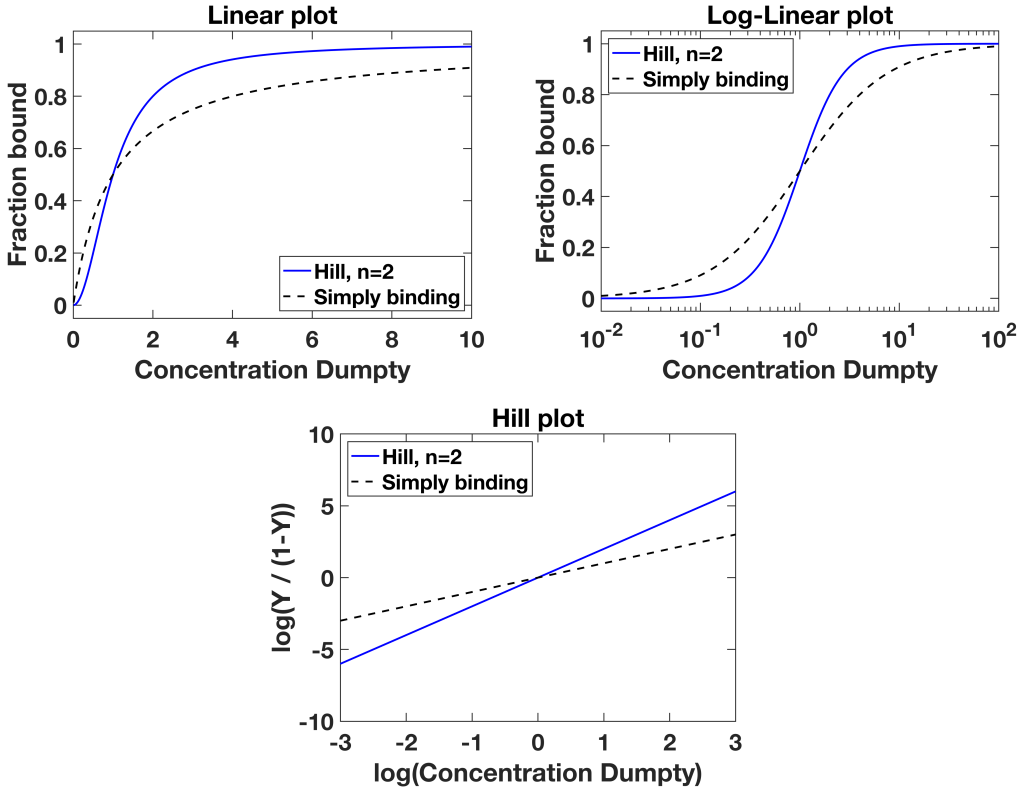

c) From the fit, I find for the *T. thermophilus* variant:
$\Delta G_0 = -6.09$ kcal/mol, $m = 2.58$ kcal/mol/M, and $[GdnHCl]_{1/2} = 2.36$ M,
and for the yeast variant:
$\Delta G_0 = -4.11$ kcal/mol, $m = 6.63$ kcal/mol/M, and $[GdnHCl]_{1/2} = 0.62$ M.
Note that the midpoint concentration corresponds to the concentration when the overall $\Delta G_{fold}$ is zero and is, therefore, given by $[GdnHCl]_{1/2} = \Delta G_0/m$.

d) In the paper by Nojima, *et al.* (see Table 1), they report for the *T. thermophilus* variant:
$\Delta G_0 = -6.32 \pm 0.18$ kcal/mol, $m = 2.70$ kcal/mol/M, and $[GdnHCl]_{1/2} = 2.34$ M,
and for the yeast variant:
$\Delta G_0 = -3.63 \pm 0.20$ kcal/mol, $m = 5.86$ kcal/mol/M, and $[GdnHCl]_{1/2} = 0.62$ M.
These values are very close the results obtained from the fit above. The fact that they differ slightly are likely due to the fact that we only fitted to the fluorescence intensity data (black data points) and not to the circular dichroism data (open symbol data points).
The fitted values mostly confirm our initial estimate based on reading off the midpoint from the graphs: *T. thermophilus* PGK is more stable in the absence of denaturant than yeast PGK, based on its more negative $\Delta G_0$. The difference in $\Delta G_0$ is somewhat less than what we might have expected from the midpoints alone, though, due to the fact that for the yeast protein the dependence of the free energy on denaturant is steeper (larger $m$ value) than for *T. thermophilus* PGK .

## Problem 2

**Humpty-Dumpty binding cooperativity.**

a) Sketches of a Hill binding curve with $n = 2$ (and binding constant = 1, in the "arbitrary" concentration units of the plots) are shown below. The linear plot is fine, but often not very informative. Most commonly, you will see the plot of fraction bound vs. log-concentration, where the "sigmoidal" shape of the binding curve is readily apparent. In the old days, before it was easy to do non-linear fits on a computer, one would often transform the data to in the end fit a straight line. This is what the Hill plot of $\log(Y/(1-Y))$ vs. $\log([Dumpty])$ accomplishes. The slope is equal to the Hill coefficient and the concentration where the x-axis is intercepted is the (log of) the midpoint concentration.




b) From the data, you can conclude that the binding of Dumpty to Humpty is cooperative. The data suggest that there are 2 binding sites that are bound in a "all-or-nothing" fashion; however, there could be more binding sites that binding with less than perfect cooperativity.

c) From the crystal structure with 4 binding sites, one can conclude that the binding exhibit cooperativity, but that the binding is not completely "all-or-nothing", since the Hill coefficient is less than 4.

## Problem 3

**Free vs. bound ligand.** This problem follows essentially Problem 6.3 or 6.4, respectively, from the first or second edition of *Physical Biology of the Cell* by Phillips *et al.*

a) To derive an expression for the fraction bound as a function of $[L]_{tot}$ and $[R]_{tot}$, we can start with the definition

$$Y = \frac{[RL]}{[R]_{tot}} = \frac{[RL]}{[R] + [RL]} = \frac{[L]/K_d}{1 + [L]/K_d} \tag{1}$$

where we have used the definition of the dissociation constant

$$K_d = \frac{[L] \cdot [R]}{[RL]} \Rightarrow [RL] = \frac{[L] \cdot [R]}{K_d} \tag{2}$$

in the last step in equation 2.

We now need to find expression of $Y$ in terms of $[L]_{tot}$, $[R]_{tot}$, and $K_d$ only. Total ligand concentration is the sum of free and bound ligand:

$$[L]_{tot} = [L] + [RL] \Rightarrow [L] = [L]_{tot} - [RL] \tag{3}$$

Similarly, total receptor is the sum of free and bound receptor:

$$[R]_{tot} = [R] + [RL] \Rightarrow [R]_{tot} = \frac{K_d[RL]}{[L]} + [RL] \tag{4}$$

In the last step, we have again used the definition of the dissociation constant. Now we plug in the last expression in equation 4 into 5:

$$[R]_{tot} = \frac{K_d[RL]}{[L]_{tot} - [RL]} + [RL] \tag{5}$$

The last expression can be rearranged to be a quadratic equation for $[RL]$, by multiplying through with the denominator and rearranging terms

$$[R]_{tot}[L]_{tot} - [R]_{tot}[RL] = K_d[RL] + [L]_{tot}[RL] - [RL]^2 \tag{6}$$

Now, we bring this into the standard form for quadratic equations:

$$[RL]^2 - ([R]_{tot} + [L]_{tot} + K_d)[RL] + [R]_{tot}[L]_{tot} = 0 \tag{7}$$

This can be solved by quadratic completion (known in German as the "p-q-formula") and gives:

$$[RL] = \frac{([R]_{tot} + [L]_{tot} + K_d) - \sqrt{([R]_{tot} + [L]_{tot} + K_d)^2 - 4[R]_{tot}[L]_{tot}}}{2} \tag{8}$$

Note that the other solution (with a + in front of the square root) gives physically meaningless results (you obtain negative dissociation constants from the fit). Finally, we recall that $Y = \frac{[RL]}{[R]_{tot}}$, the starting point of our derivation, and write

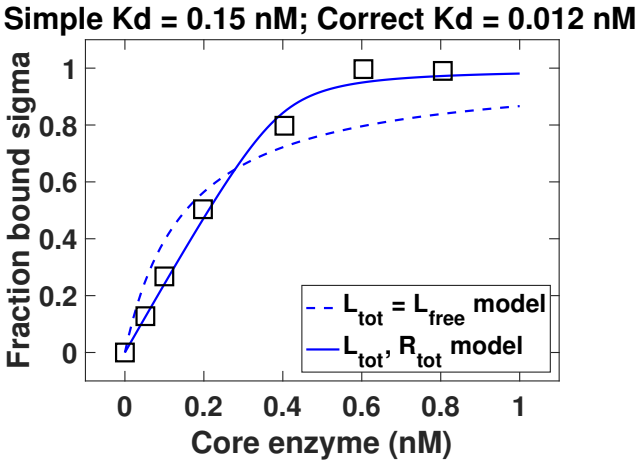
$$Y = \frac{([R]_{tot} + [L]_{tot} + K_d) - \sqrt{([R]_{tot} + [L]_{tot} + K_d)^2 - 4[R]_{tot}[L]_{tot}}}{2[R]_{tot}} \tag{9}$$

As desired, this gives us an expression of $Y$ as a function of $[L]_{tot}$, $[R]_{tot}$, and $K_d$ only. This expression can be used to fit $K_d$ to the data, see the separate matlab code for part c.

b) Briefly, in the limit where $[L]_{tot} \gg [R]_{tot}$, we also have $[L]_{tot} \gg [RL]$ and thus $[L]_{tot} \approx [L]$ and we do not need to distinguish free and bound ligand concentration and can use equation 2.

c) Inspecting the data of Maeda *et al.*, *Nucleic Acids Research* (2000), for core enzyme binding to the $\sigma^{70}$ subunit, we see that the core enzyme concentrations (the ligand, in this case) are in the range of 0 to 0.8 nM, while the receptor concentration ($\sigma^{70}$) is at a concentration of 0.4 nM; clearly we do not have $[L]_{tot} \gg [R]_{tot}$ and need to fit our modified model.

If we fit the standard expression (dashed line below), we obtain a relatively poor fit and an apparent dissociation constant of 0.15 nM. Fitting the correct model, we obtain a much better fit (solid line) and a significantly lower dissociation constant of 0.012 nM, corresponding to higher affinity. This is to be expected: Given the high receptor concentration, at the same total ligand concentration the free ligand concentration is always going to be lower than it would be for a lower receptor concentration. Therefore, the true affinity needs to be higher than it would have to be in the $[L]_{tot} \gg [R]_{tot}$ case to see e.g. half maximal binding at a certain concentration.




The matlab code used for fitting the two models and plotting the results is available on the course website.