Derivations for Effective Free Energy Differences

Ian Kenney

May 16, 2018

1 Maximum likelihood estimation of the free energies of states

We start with the likelihood function, which defines the likelihood of seeing our measured results given a set of free energy values for the states in our network:

$$L(\{\bar{\Delta}_{ij}\}|\{g_i\}) = \prod_{ij} \exp\left(\frac{-1}{2\sigma_{ij}^2}((g_j - g_i) - \bar{\Delta}_{ij})^2\right) \tag{1}$$

We want to use this function to find the most likely $\{g_i\}$ that fits our data. We can do this by taking the gradient in g-space, $\boldsymbol{\nabla}_g$, and setting it equal to the null vector. Since the likelihood function is monotonic, we can maximize the natural logarithm of (1), as this is will greatly simplify the mathematics. The log-likelihood is given by:

$$\ln L = \sum_{ij} \frac{-1}{2\sigma_{ij}^2}((g_j - g_i) - \bar{\Delta}_{ij})^2 \tag{2}$$

We define the gradient operator as:

$$\boldsymbol{\nabla}_g \equiv \sum_k \hat{\boldsymbol{g}}_{\boldsymbol{k}} \frac{\partial}{\partial g_k} \tag{3}$$

Taking the gradient of (2) and setting equal to the null vector, we find:

$$\mathbf{0} = \boldsymbol{\nabla}_g \ln L = \sum_{ij} \frac{-1}{2\sigma_{ij}^2} \boldsymbol{\nabla}_g((g_j - g_i) - \bar{\Delta}_{ij})^2 \tag{4}$$

$$= \sum_{ij} \frac{-1}{\sigma_{ij}^2}((g_j - g_i) - \bar{\Delta}_{ij}) \left[\hat{\boldsymbol{g}}_{\boldsymbol{j}} - \hat{\boldsymbol{g}}_{\boldsymbol{i}}\right] \tag{5}$$

$$= \sum_{ij} \frac{1}{\sigma_{ij}^2}((g_j - g_i) - \bar{\Delta}_{ij}) \left[\hat{\boldsymbol{g}}_{\boldsymbol{i}} - \hat{\boldsymbol{g}}_{\boldsymbol{j}}\right] \tag{6}$$

From here, we can compute $\{g_i\}$ numerically using `scipy.optimize.root`, which uses the hybrid Powell method to find the roots of $\boldsymbol{\nabla}_g \ln L$. We provide the root function the jacobian, whose elements are:

$$J_{nm} = \frac{\partial}{\partial g_n} \left(\boldsymbol{\nabla}_g \ln L\right)_m = \frac{\partial}{\partial g_n} \sum_{ij} \frac{1}{\sigma_{ij}^2} \left[(g_j - g_i) - \bar{\boldsymbol{\Delta}}_{\boldsymbol{ij}}\right] (\delta_{mi} - \delta_{mj}) \tag{7}$$

$$= \sum_{ij} \frac{1}{\sigma_{ij}^2}(\delta_{nj} - \delta_{ni})(\delta_{mi} - \delta_{mj}) \tag{8}$$

$$= \sum_{ij} \frac{1}{\sigma_{ij}^2} \left[\delta_{nj}\delta_{mi} - \delta_{nj}\delta_{mj} - \delta_{ni}\delta_{mi} + \delta_{ni}\delta_{mj}\right] \tag{9}$$

$$\tag{10}$$

2 Macroscopic pKas for a single ligand species

We would like to use the calculated effective energy differences to determine macroscopic pKas (i.e. the equilibrium constant that describes the populations of various macrostates). We will use proton binding as an example but note that this derivation can apply to any binding ligand as long as it is the sole binding species.

We define the macroscopic pKa as:

$$pKa_m \equiv -\log\left(\frac{[C^-][H^+]}{[CH]c_0}\right) \tag{11}$$

Where $[C^-]$ is the concentration of protonated complexes, $[H^+]$ is the concentration of protons, $[CH]$ is the concentration of protonated complexes, and c_0 is the standard state concentration of hydrogen. Using the properties of logarithmic functions:

$$\begin{aligned} pKa_m &\equiv -\log\left(\frac{[C^-][H^+]}{[CH]c_0}\right) \\ &= -\log\left(\frac{[C^-]}{[CH]}\right) + pH \\ &= -\log\left(\frac{P(C^-)}{P(CH)}\right) + pH \\ &= -\ln\left(\frac{P(C^-)}{P(CH)}\right)/\ln(10) + pH \\ &= \frac{\beta\Delta G_{C^-,CH}}{\ln 10} + pH \end{aligned}$$

where $\Delta G_{C^-,CH}$ is the effective free energy difference of protonation, which is:

$$\Delta G_{N-1,N} = \beta^{-1}\ln\left[\frac{\sum_s \exp(-\beta\Delta G_s)\delta_{N_s,N-1}}{\sum_s \exp(-\beta\Delta G_s)\delta_{N_s,N}}\right] \tag{12}$$

where the $\delta_{N_s,N}$ and $\delta_{N_s,N-1}$ will pick out states that contain the correct number of protonated residues. For sake of clarity, we select as a reference energy, the energy any deprotonated microstate. The motivation for this choice becomes clear when considering the free energy of (microscopic) protonation/deprotonation. To protonated/deprotonate a residue, the free energy differnce is:

$$\Delta G_{prot} = -\Delta G_{dep} = \beta^{-1}\ln(10)(pH - pKa) \tag{13}$$

Consider a system of three protonatable residues. We will represent the protonation state by a string of 1s and 0s. For example, 111 represents that state where all residues are protonated (N=3) while 000 represents the state where no residues are protonated (N=0). These two examples, of course are macrostates with no degenerate microstates. We have two other macrostates of interest, N=1 and N=2. The N=1 macrostate has a three fold degeneracy (001,010,100) and the N=2 macrostate has a three fold degeneracy (011,101,110). In order to go from 001 to 010, we would be required to deprotonate residue three and then protonate residue two. The energy difference between these two microstates is then:

$$\Delta G_{001,010} = \beta^{-1}\ln(10)\left[(-pH + pKa_{000,001}) + (pH - pKa_{000,010})\right] = \beta^{-1}\ln(10)\left[pKa_{000,001} - pKa_{000,010}\right] \tag{14}$$

A similar argument can be made for free energy difference between 001 and 100. Because of this behavior, we see that the energy difference between two microstates is only dependent on the temperature and the difference of the pKas of each state. In the case of moving between microstates from different macrostates, there is a first order pH dependence, as well as a linear combination of pKas that is a function of the connectivity of the graph. We can select any reference state we would like. In our case, we select any ΔG

from the numerator of the argument of the natural logarithm. By doing this, all terms in the numerator have no pH dependence and are all just functions of the connectivity, which we will call η_i.

We then find that (12) can be written as:

$$\begin{aligned}
\Delta G_{N-1,N} &= \beta^{-1} \ln \left[\frac{\sum_s \exp(-\eta_s) \exp(-\ln(10) pH) \delta_{N_s,N-1}}{\sum_s \exp(-\eta_s) \delta_{N_s,N}} \right] \\
&= \beta^{-1} \ln \left[\frac{\exp(-\ln(10) pH) \sum_s \exp(-\eta_s) \delta_{N_s,N-1}}{\sum_s \exp(-\eta_s) \delta_{N_s,N}} \right] \\
&= -\beta^{-1} \ln(10) pH + \beta^{-1} \ln \left[\frac{\sum_s \exp(-\eta_s) \delta_{N_s,N-1}}{\sum_s \exp(-\eta_s) \delta_{N_s,N}} \right]
\end{aligned}$$

Plugging this into our expression for the pKa_m, we find that the pH dependence drops out and our result is:

$$pKa_m = \ln \left[\frac{\sum_s \exp(-\eta_s) \delta_{N_s,N-1}}{\sum_s \exp(-\eta_s) \delta_{N_s,N}} \right] / \ln(10)$$

We conclude that the macroscopic pKas are independent of pH. More generally, we have shown that the macroscopic equilibrium constants are independent of the ligand concentration.