

## Titration curve of certain residue in protein

Ye Mei\*

*State Key Laboratory of Precision Spectroscopy,  
East China Normal University, Shanghai 200062 China*

---

\*Electronic address: ymei@phy.ecnu.edu.cn

## I. INTRODUCTION

Prediction of the titration states of residues in protein is critical to the simulation of proteins.

## II. THEORY

Free energy of molecule A in solution with a concentration  $[A]$  can be written as

$$G_A = G_A^0 + kT \ln \frac{[A]}{C_0},$$

in which  $\Delta G_0$  is the free energy of molecule A at the standard state  $C_0$ , i.e. 1 mol/L. The free energy change for a reaction



can be written as

$$\Delta G = \Delta G_C - \Delta G_A - \Delta G_B = \Delta G_0 + kT \ln \frac{[C] C_0}{[A] [B]}.$$

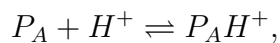
At equilibrium, the free energy change is zero, we have

$$\Delta G_0 = -kT \ln \frac{[C] C_0}{[A] [B]}, \tag{1}$$

in which  $[A] [B] / [C] C_0$  is called the dissociation constant  $K_a$ . So,

$$\Delta G_0 = kT \ln K_a.$$

Titration of a residue in a real protein can be written as



with

$$K_a = \frac{[P_A] [H^+]}{[P_A H^+] C_0}$$

The fraction of the deprotonated species is calculated as

$$\begin{aligned}
f_{[P_A]} &= \frac{[P_A]}{[P_A] + [P_A H^+]} \\
&= \frac{1}{1 + \frac{[P_A H^+]}{[P_A]}} \\
&= \frac{1}{1 + \frac{[P_A][H^+]}{C_0 K_a [P_A]}} \\
&= \frac{1}{1 + \frac{1}{C_0 K_a} [H^+]} \\
&= \frac{1}{1 + \frac{1}{K_a} 10^{-pH}}
\end{aligned} \tag{2}$$

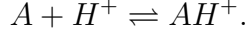
We can check the asymptotic behavior of this equation. At strong acidic condition ( $pH = -\infty$ ),  $f_{\{P_A\}} = 0$ , indicating that the residue is 100 percent protonated. While at an extremely basic condition ( $pH = \infty$ ),  $f_{\{P_A\}} = 1$ . This residue is 100 percent deprotonated. The  $pK_a$  can be determined by the  $pH$  of the state when  $[P_A] / [P_A H^+] = 1$  (the isoelectric point)

$$\begin{aligned}
pK_a &= -\log K_a \\
&= -\log \frac{[P_A]}{[P_A H^+]} - \log \frac{[H^+]}{C_0} \\
&= pH \quad \text{when } [P_A] = [P_A H^+]
\end{aligned}$$

The  $pK_a$  of each residue in a dipeptide has been determined by experiment. However, when this residue is located in a certain protein, its  $pK_a$  is different from that in a dipeptide. The difference is called the  $pK_a$  shift. Instead of measuring the  $pK_a$  for a residue in a protein, we are more interested in calculating/measuring the titration curve, which is the fraction of the deprotonated state as a function of pH. From Eq. 2,  $f_{[P_A]}$  can be easily calculated if we know  $K_a$  or equivalently the standard free energy change of protonation in Eq. 1. The standard free energy can be calculated from the partition functions as

$$\begin{aligned}
\Delta G_0 &= -kT \ln \frac{Q_{P_A H^+}}{Q_{P_A} Q_{H^+}} \\
&= -kT \ln \frac{\int \exp(-E_{P_A H^+}/kT) dR_H dR_o}{Q_{H^+} \int \exp(-E_{P_A}/kT) dR_o}
\end{aligned}$$

Generally, the absolute value of  $\Delta G_0$  is hardly computable. A relative protonation free energy  $\Delta\Delta G$  is preferred and is more reliable. Theoretically, the reference state can be any state you like. But the protonation free energy of the dipeptide at  $pK_a$  is often used. The reference protonation process can be written as



The reference free energy change is

$$\begin{aligned}\Delta\Delta G_0 &= \Delta G_0 - \Delta G_0^{ref} \\ &= -kT \ln \frac{\int \exp(-E_{P_A H^+}/kT) dR_H dR_o}{Q_{H^+} \int \exp(-E_{P_A}/kT) dR_o} \frac{Q_{H^+} \int \exp(-E_A/kT) dR_o}{\int \exp(-E_{AH^+}/kT) dR_H dR_o} \\ &= -kT \ln \frac{\int \exp(-E_{P_A H^+}/kT) dR_H dR_o \int \exp(-E_A/kT) dR_H dR_o}{\int \exp(-E_{P_A}/kT) dR_H dR_o \int \exp(-E_{AH^+}/kT) dR_H dR_o} \\ &= -kT \ln \frac{\int \int \exp \left[ - \left( E_{P_A H^+}^{bond} + E_{P_A H^+}^{QM} + E_{P_A H^+}^{ele} \right) / kT \right] dR_H \exp \left( -E_{P_A H^+}^{other} / kT \right) dR_o}{\int \int \exp \left[ - \left( E_{AH^+}^{bond} + E_{AH^+}^{QM} + E_{AH^+}^{ele} \right) / kT \right] dR_H \exp \left( -E_{AH^+}^{other} / kT \right) dR_o} \\ &\quad \cdot \frac{\int \exp(-E_A/kT) dR_o}{\int \exp(-E_{P_A}/kT) dR_o},\end{aligned}\tag{3}$$

where  $R_H$  is the coordinates of the specific  $H$  atom and the other degrees-of-freedom (DoF) are denoted as  $R_o$ .  $E^{bond}$  and  $E^{ele}$  are the bonded energy and electrostatic interaction energy related to this  $H$  atom, respectively.  $E^{QM}$  is the energy correction that *may* be required if the molecular mechanical Hamiltonian cannot well capture the energy of the system, such as the missing of charge transfer effect. The sum of the remaining energy term is denoted as  $E^{other}$ , which does not explicitly depend on the position of this specific  $H$  atom. Eq. 3 is not computable before some approximations are adopted. First, we assume that the total energy can be well described by the MM Hamiltonians for both the state interested in and the reference state. Therefore,

$$E_{P_A H^+}^{QM} = E_{AH^+}^{QM} = Const,$$

and they can be removed from the integral. Second, the bonded terms involving hydrogen atoms are usually constrained in the simulations. Therefore,  $E^{bond} = 0$ . Now, the relative protonation free energy can be simplified as

$$\Delta\Delta G_0 = -kT \ln \frac{\int \int \exp \left( -E_{P_A H^+}^{ele} / kT \right) dR_H \exp \left( -E_{P_A H^+}^{other} / kT \right) dR_o}{\int \int \exp \left( -E_{AH^+}^{ele} / kT \right) dR_H \exp \left( -E_{AH^+}^{other} / kT \right) dR_o} \cdot \frac{\int \exp(-E_A/kT) dR_o}{\int \exp(-E_{P_A}/kT) dR_o}.$$

Note that  $E_A = E_{AH^+}^{other}$  and  $E_{P_A} = E_{P_AH^+}^{other}$ , we have

$$\begin{aligned}
\Delta\Delta G_0 &= -kT \ln \frac{\int \int \exp\left(-E_{P_AH^+}^{ele}/kT\right) dR_H \exp\left(-E_{P_A}/kT\right) dR_O}{\int \exp\left(-E_{P_A}/kT\right) dR_O} \\
&\quad \cdot \frac{\int \exp\left(-E_A/kT\right) dR_O}{\int \int \exp\left(-E_{AH^+}^{ele}/kT\right) dR_H \exp\left(-E_A/kT\right) dR_O} \\
&= -kT \ln \left\langle \int \exp\left(-E_{P_AH^+}^{ele}/kT\right) dR_H \right\rangle_{P_A} + kT \ln \left\langle \int \exp\left(-E_{AH^+}^{ele}/kT\right) dR_H \right\rangle_A
\end{aligned}$$

Finally, if we assume  $E^{ele}$  and  $E^{other}$  are decoupled from each other, the double integral can be integrated out independently for  $R_H$  and  $R_O$ , the contribution from other DoF cancel out between the protonated state and the deprotonated state. And now we have

$$\begin{aligned}
\Delta\Delta G_0 &= -kT \ln \frac{\int \exp\left(-E_{P_AH^+}^{ele}/kT\right) dR_H}{\int \exp\left(-E_{AH^+}^{ele}/kT\right) dR_H} \\
&= \Delta G_{P_AH^+}^{ele} - \Delta G_{AH^+}^{ele}.
\end{aligned}$$

Therefore,

$$-kT \ln 10 \cdot pK_a = \Delta G_{P_AH^+}^{ele} - \Delta G_{AH^+}^{ele} - kT \ln 10 \cdot pK_a^{ref}.$$

At a certain pH, the free energy difference between the deprotonated and the protonated state can be written as

$$\Delta G_{P_A \rightarrow P_AH^+} = \Delta G_{P_AH^+}^{ele} + kT(pH - pK_a^{ref}) \ln 10 - \Delta G_{AH^+}^{ele}.$$