Supplementary material

1. Theory Introduction

TK-SA system treating the protein folding as a chemical reaction, utilization TK-SA model, the reaction thermodynamics and the partition function to calculate the reaction Gibbs free energy $_{\Delta}G_{qq}$ caused by the charge distribution to measure the effect of surface charge distribution on the protein stability. If the value of $_{\Delta}G_{qq}$ is positive, the distribution of surface charge plays a negative role on the protein overall stability; otherwise it plays a positive role[1-18].

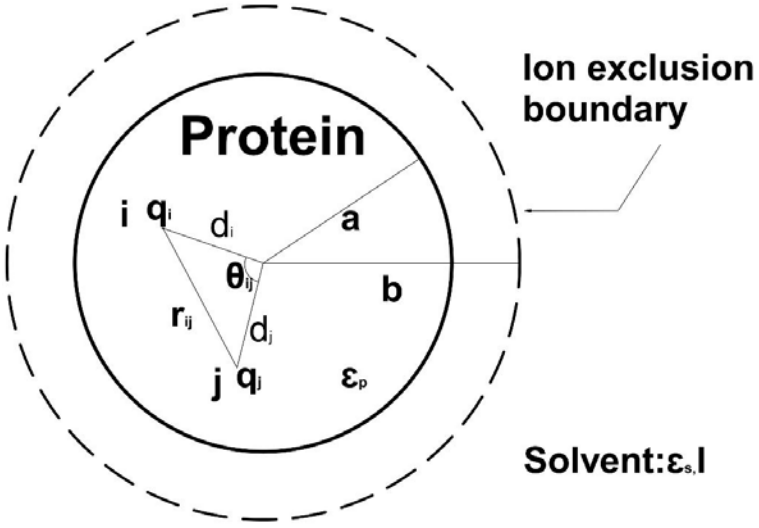

Figure S1. Diagrammatic sketch of TK model. a is the radius of the protein, b is the extent radius, i and j represent two chargeable residues, q_i and q_j are charge number when they are protonated, r_{ij} is the distance between point i and j, ε_p is the dielectric constant of protein, ε_s is the dielectric constant of solution, I ionic strength of the solution. di and dj is the distance between the center of the protein and the poison of i or j.

In TK-SA model, the interaction energies between the surface charges of point i and j are

calculated as:

$$E_{ij} = \varepsilon^2 \left(\frac{A_{ij} - B_{ij}}{2b} - \frac{C_{ij}}{2a} \right) \left(1 - SA_{ij}\right) \quad \text{(Eq.1)}$$

Where b is the radius of the sphere representing the protein, a is the radius of the sphere from which solvent ions are excluded, calculated as :

$$a = b + 1.4 \quad \text{(Eq.2)}$$

SA_{ij} is the average solvent accessibility of group i and calculated as:

$$SA_{ij} = \frac{\left(SA_i + SA_j\right)}{2} \quad \text{(Eq.3)}$$

SA_i and SA_j are the solvent accessibility of group i and j respectively, calculated as :

$$SA_i = \frac{ASA_{\text{protein}}}{ASA_{\text{tri}peptide}} \quad \text{（Eq.4）}$$

ε is the unit charge whose value is 1.602×10-19C

For A_{ij}:

$$A_{ij} = \frac{b}{D_i r_{ij}} \quad \text{（Eq.5）}$$

Where the value of D_i is 4F/m, r_{ij} represents the distance of point i and point j

For B_{ij}:

$$B_{ij} = \frac{1}{D_i} \sum_{n=0}^{\infty} \frac{(n+1)(D - D_i)}{(n+1)D + nD_i} \rho_{ij}^n p_n\left(\cos\theta_{ij}\right) \quad \text{（Eq.6）}$$

Where the value of D is 78.5 F/m, $p_n\left(\cos\theta_{ij}\right)$ represents Legendre polynomials

$$\rho_{ij}^n = \frac{r_i \times r_j}{b^2} \quad \text{（Eq.7）}$$

Where r_i and r_j are the distance from i and j to the protein center respectively.

For C_{ij}:

$$C_{ij}=\frac{1}{D}\left\{\frac{x}{1+x}+\sum_{n=1}^{\infty}\frac{2n+1}{2n-1}\left[\frac{D}{(n+1)D+nD_i}\right]^2\times\frac{x^2\sigma_{ij}^n p_n(\cos\theta_{ij})}{\frac{K_{n+1}(x)}{K_{n-1}(x)}+\frac{n(D-D_i)}{(n+1)D+nD_i}\left(\frac{b}{a}\right)^{2n+1}\frac{x^2}{4n^2-1}}\right\}$$

（Eq.8）

Where

$$\sigma_{ij}=\frac{r_i\times r_j}{a^2}$$ （Eq.9）

$$K_n(x)=\sum_{s=0}^{n}\frac{2^s n!(2n-s)!}{s!(2n)!(n-s)!}x^s$$ （Eq.10）

$$x=ka$$ （Eq.11）

κ is the Debye-Huckel parameters which is proportional to the square root of the ionic strength of the system:

$$\kappa^2=\left(\frac{4\pi N\varepsilon^2}{1000DkT}\right)\sum_i Z_i^2 C_i$$ （Eq.12）

Where N is Avogadro whose value is 6.02×10^{23}, k represents Boltzmann constant whose value is 1.38J/K, the value of T,D and ε are 298.15K, 78.5 F/m and 1.602×10-19C respectively.

$$2I=\sum_i Z_i^2 C_i$$ （Eq.13）

Where I is the ionic strength, the value of which is 0.1M.

The denaturation and renaturation of protein is essentially the change of a certain conformation caused by some factors. In view of the ‘two-state’ theory of protein denaturation,

the protein folding in certain circumstances can be simplified into a chemical reaction:

$$U\left(\text{unfolded}\right) \rightleftharpoons N\left(native\right)$$

According to the Gibbs criterion, the process can automatically carry forward when the value of Gibbs function decreases; when the value is constant, it is in equilibrium; and the value increases during reverse spontaneously. Therefore, the protein stability can be studied by the change of Gibbs free energy.

$$\Delta G_{\text{qq}} = -RT\ln\left(\frac{Z_U}{Z_N}\right) \quad \text{（Eq.14）}$$

Where R is gas constant whose value is 8.314472J/(K•mol),T represents absolute temperature:

$$T = \text{t} + 273.15 \quad \text{（Eq.15）}$$

Where t is the ambient temperature when calculating.

Z_N and Z_U represent the partition functions for the native and unfolded states, respectively.

$$Z_N = \sum_{\chi} \exp\left(-\frac{G_N(\chi)}{RT} - \nu(\chi)(\ln 10)\, pH\right) \quad \text{（Eq.16）}$$

$$Z_U = \sum_{\chi} \exp\left(-\frac{G_U(\chi)}{RT} - \nu(\chi)(\ln 10)\, pH\right) \quad \text{（Eq.17）}$$

Where $\nu(\chi)$ represents the number of protonated ionizable groups in the protonation state. $G_N(\chi)$ and $G_U(\chi)$ are the native and unfolding protein energy respectively in a certain protonation state. It is generally considered that in unfolded state the charged amino acids of the dissociated protein do not form charge force, so as not to contribute to the protein stability.

$$G_N(\chi) = -RT(\ln 10)\sum_{i=1}^{n}(q_i + x_i)\,pK_{\text{int},i} + \frac{1}{2}\sum_{i,j=1}^{n} E_{ij}(q_i + x_i)(q_j + q_j)$$

（Eq.18）

$$G_U(\chi) = -RT(\ln 10)\sum_{i=1}^{n}(q_i + x_i)\,pK_{\text{int},i} \qquad \text{（Eq.19）}$$

According to the expressions, the protein energy in native state is mainly affected by the charge interaction force between terrible points. q_i is the charge of titrable point in unprotonated state. The value of q_i is -1 when the amino acids are ASP, GLU and Ctr, the value is 0 when amino acids are ARG, LYS, HIS and Ntr.

X_i is used to describe N protonated sites, the value is 0 when site i is not protonated, otherwise the value is 1. $pk_{int,i}$ is used to represent the intrinsic pka value. The pka values of ASP,LYS,GLU,ARG,HIS,N terminal and C-terminal are 4.0,10.6,4.5,12.0,6.3,7.7,3.6 respectively.

E_{ij} is the charge interaction between charge i and j, shown as equation Eq.1

Therefore, when the value of ΔG_{qq} is negative, the reaction goes to folding direction, that is the surface charge distribution plays an active role on the protein stability. The reaction goes to the reverse direction when the value is positive which means that the distribution of the surface charge has a negative effect on the protein stability.

References

(1) Kirkwood, J. G. *The Journal of Chemical Physics* **1934**, *2*, 351.
(2) Tanford, C.; Kirkwood, J. G. *Journal of the American Chemical Society* **1957**, *79*, 5333.
(3) Lee, B.; Richards, F. M. *Journal of Molecular Biology* **1971**, *55*, 379.
(4) Matthew, J.; Hanania, G.; Gurd, F. *Biochemistry* **1979**, *18*, 1919.
(5) Richards, F. M. *Methods in Enzymology* **1985**, *115*, 440.
(6) Serrano, L.; Horovitz, A.; Avron, B.; Bycroft, M.; Fersht, A. R. *Biochemistry* **1990**, *29*, 9343.
(7) Bashford, D.; Karplus, M. *The Journal of Physical Chemistry* **1991**, *95*, 9556.

(8) Havranek, J. J.; Harbury, P. B. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 11145.

(9) Sanchez-Ruiz, J. M.; Ibarra-Molero, B.; Loladze, V. V.; Makhatadze, G. I. *Biochemistry* **1999**, *38*, 8138.

(10) Tsai, J.; Taylor, R.; Chothia, C.; Gerstein, M. *Journal of Molecular Biology* **1999**, *290*, 253.

(11) Elcock, A. H. *Journal of Molecular Biology* **2001**, *312*, 885.

(12) Sanchez-Ruiz, J. M.; Makhatadze, G. I. *Trends in biotechnology* **2001**, *19*, 132.

(13) Ibarra-Molero, B.; Sanchez-Ruiz, J. M. *The Journal of Physical Chemistry B* **2002**, *106*, 6609.

(14) Makhatadze, G. I.; Loladze, V. V.; Ermolenko, D. N.; Chen, X. F.; Thomas, S. T. *Journal of Molecular Biology* **2003**, *327*, 1135.

(15) Makhatadze, G. I.; Loladze, V. V.; Gribenko, A. V.; Lopez, M. M. *Journal of Molecular Biology* **2004**, *336*, 929.

(16) Strickler, S. S.; Gribenko, A. V.; Keiffer, T. R.; Tomlinson, J.; Reihle, T.; Loladze, V. V.; Makhatadze, G. I. *Biochemistry* **2006**, *45*, 2761.

(17) Gribenko, A. V.; Patel, M. M.; Liu, J.; McCallum, S. A.; Wang, C. Y.; Makhatadze, G. I. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 2601.

(18) Schweiker, K. L.; Makhatadze, G. I. *Methods Mol Biol* **2009**, *490*, 261.