Supporting Information

Itoh and Sasai

SI Text

Derivation of Energy, Entropy, and Free Energy Functions.
The partition function at a fixed (x, y) is obtained by

$$Z(x,y) = \nu^{N-x-y} \sum_{\text{conf}\in\{(x,y)\}} \exp(-\beta H), \qquad [1]$$

where $\sum_{\text{conf}\in\{(x,y)\}}$ is summation of conformations under the constraint of (x, y), H is Hamiltonian, $\beta = 1/k_{\mathrm{B}}T$ is the inverse of temperature T, N is the total number of residues, and ν is the number of nonnative configurations each residue can take. The order parameters of the native-likeness, x and y, are numbers of residues which take the native-like configuration in domain I and domain II, respectively. $Z(x, y)$ can be rewritten as

$$\begin{aligned} Z(x,y) &= \nu^{N-x-y} \sum_{\text{conf}\in\{(x,y)\}} 1 \Bigg/ \left(\nu^{N-x-y} \sum_{\text{conf}\in\{(x,y)\}} 1 \Big/ Z(x,y) \right) \\ &= \frac{1}{\left\langle \exp(\beta H) \right\rangle_{x,y}} \cdot \nu^{N-x-y} \sum_{\text{conf}\in\{(x,y)\}} 1 , \end{aligned} \qquad [2]$$

where $\langle \cdots \rangle_{x,y}$ is the average taken by using $Z(x, y)$.

The free energy at a fixed (x, y) is

$$\begin{aligned} F(x,y) &= -k_{\mathrm{B}}T \ln Z(x,y) \\ &= k_{\mathrm{B}}T \ln \left\langle \exp(\beta H) \right\rangle_{x,y} - Tk_{\mathrm{B}} \ln \left[\nu^{N-x-y} \sum_{\text{conf}\in\{(x,y)\}} 1 \right] \qquad [3] \\ &= \left\langle H \right\rangle_{x,y} + Tk_{\mathrm{B}} \ln \left\langle \exp[\beta(H - \left\langle H \right\rangle_{x,y})] \right\rangle_{x,y} - Tk_{\mathrm{B}} \ln \left[\nu^{N-x-y} \binom{n_{\mathrm{I}}}{x} \binom{n_{\mathrm{II}}}{y} \right]. \end{aligned}$$

Here, $\sum_{\mathrm{conf}\in\{(x,y)\}} 1 = \sum_{\mu_{\mathrm{I}}(x)} 1 \sum_{\mu_{\mathrm{II}}(y)} 1$ with $\mu_{\mathrm{I}}(x) = \{m_k \mid \Sigma_{k=1}^{n_{\mathrm{I}}} m_k = x \ (1 \le k \le n_{\mathrm{I}})\}$ and $\mu_{\mathrm{II}}(x) = \{m_k \mid \Sigma_{k=n_{\mathrm{I}}+1}^{N} m_k = y \ (n_{\mathrm{I}}+1 \le k \le N)\}$ (m_k = 0,1), and thus $\sum_{\mu_{\mathrm{I}}(x)} 1 = \binom{n_{\mathrm{I}}}{x}$ and $\sum_{\mu_{\mathrm{II}}(y)} 1 = \binom{n_{\mathrm{II}}}{y}$, where n_{I} and n_{II} are numbers of residues in the N-terminal domain (domain I) and the C-terminal domain (domain II), respectively, $n_{\mathrm{I}} + n_{\mathrm{II}} = N$.

The free energy function can be decomposed into energy and three entropy terms as

$$F(x,y) = E(x,y) - T[S_{\mathrm{e}}(x,y) + S_{\mathrm{c}}(x;n_{\mathrm{I}}) + S_{\mathrm{c}}(y;n_{\mathrm{II}})]. \qquad [4]$$

$E(x,y)$ is energy at a fixed (x,y),

$$E(x,y) = \langle H \rangle_{x,y} \qquad [5]$$

$S_{\mathrm{e}}(x,y)$ is the energy dependent entropy at a fixed (x,y),

$$S_{\mathrm{e}}(x,y) = -k_{\mathrm{B}} \ln \langle \exp[\beta(H - E(x,y))] \rangle_{x,y}. \qquad [6]$$

From the inequality $\langle e^{\beta H} \rangle_{x,y} \ge e^{\beta E(x,y)}$, we find $S_{\mathrm{e}}(x,y) \le 0$. $-TS_{\mathrm{e}}(x,y)$ can be expanded as

$$-TS_{\mathrm{e}}(x,y) = \sum_{n=2}^{\infty} \frac{\beta^{n-1}}{n!} \langle H^n \rangle_{x,y}^{c}, \qquad [7]$$

where $\langle H^n \rangle_{x,y}^{c}$ is the nth order cumulant of H under the constraint of (x,y). $-TS_{\mathrm{e}}(x,y)$ decreases as temperature increases and $-TS_{\mathrm{e}}(x,y) \approx \frac{1}{2k_{\mathrm{B}}T}(\langle H^2 \rangle_{x,y} - \langle H \rangle_{x,y}^2)$ at high temperature.

$S_{\mathrm{c}}(x;n_{\mathrm{I}})$ and $S_{\mathrm{c}}(y;n_{\mathrm{II}})$ express chain entropies of domain I and domain II, respectively, which arise from the total number of conformations that each domain can take under the

constraint of (x, y), where

$$S_c(x;n) = k_B \ln\left\{ \nu^{n-x} \binom{n}{x} \right\}. \qquad [8]$$

Notice that $S_c(x; n_I)$ and $S_c(y; n_{II})$ do not depend on the form of Hamiltonian, so that the effects of the domain-domain interactions on entropy are solely expressed in $S_e(x, y)$.

We decompose the Hamiltonian into the intra-domain parts, H_I and H_{II}, and the inter-domain part V as $H = H_I + H_{II} + V$, where H_I is a sum of terms belonging to domain I, H_{II} is a sum of terms belonging to domain II, and V is a sum of terms of interactions between a residue in domain I and a residue in domain II.

$$\begin{aligned} k_B T \ln\langle \exp(\beta H)\rangle_{x,y} &= k_B T \ln\left[\frac{\nu^{N-x-y} \sum_{\text{conf}\in\{(x,y)\}} 1}{Z_0(x,y)} \cdot \frac{Z_0(x,y)}{Z(x,y)} \right] \\ &= k_B T \ln\langle \exp[\beta(H_I + H_{II})]\rangle_{0\,x,y} + U(x,y), \end{aligned} \qquad [9]$$

with $Z_0(x,y) = \nu^{N-x-y} \sum_{\text{conf}\in\{(x,y)\}} \exp\{-\beta(H_I + H_{II})\}$, where $\langle\cdots\rangle_{0x,y}$ is the average taken by using $Z_0(x,y)$.

$$U(x,y) \equiv F(x,y) - F_0(x,y) = k_B T \ln\langle \exp(\beta V)\rangle_{x,y}, \qquad [10]$$

is difference in free energy between the connected two-domain protein and the separated two non-interacting domains, where $F_0(x,y) = -k_B T \ln Z_0(x,y)$ is the free energy function of separated domains. $U(x,y)$ can be described as

$$\begin{aligned} U(x,y) &= \langle V\rangle_{x,y} + T k_B \ln\left\langle \exp[\beta(V - \langle V\rangle_{x,y})]\right\rangle_{x,y} \\ &= [E(x,y) - E_0(x,y)] - T[S_e(x,y) - S_{e0}(x,y)]. \end{aligned} \qquad [11]$$

The Constrained Partition Function of Wako-Saito-Muñoz-Eaton (WSME) Hamiltonian.

The partition function at the constraint $(x, y) = (n, m)$ is calculated from the generating function:

$$Q(\lambda, \mu) \equiv \sum_{x,y} Z(x, y)\lambda^x \mu^y \,, \qquad [12]$$

as

$$Z(n, m) = \frac{1}{n!m!} \frac{\partial^n}{\partial \lambda^n} \frac{\partial^m}{\partial \mu^m} Q(\lambda, \mu)\Big|_{\lambda=\mu=0} \,. \qquad [13]$$

$Q(\lambda, \mu)$ of the WSME Hamiltonian,

$$H = -\sum_{i<j} \varepsilon_{i,j} \Delta_{i,j} \prod_{k=i}^{j} m_k \,.$$

can be obtained by means of the transfer matrix method without introducing any further approximation (1,2). $Q(\lambda, \mu)$ is expressed as

$$Q(\lambda, \mu) = P_{N-1}(1; \lambda, \mu) e^{-S_0/k_B} \,, \qquad [14]$$

and $P_{N-1}(1; \lambda, \mu)$ is obtained by calculating the following the recurrence equations:

$$\begin{aligned}
&P_0(1; \lambda, \mu) = 1, \\
&P_0(l; \lambda, \mu) = w_{N,N-l+2}(\lambda, \mu), \quad l \neq 1, \\
&P_k(1; \lambda, \mu) = P_{k-1}(1; \lambda, \mu) + P_{k-1}(2; \lambda, \mu), \\
&P_k(l; \lambda, \mu) = w_{N-k,N-k-l+2}(\lambda, \mu) P_{k-1}(l; \lambda, \mu) + P_{k-1}(l+1; \lambda, \mu), \\
&l = 1, 2 \cdots N-k-1, \; k = 1, 2 \cdots N-1,
\end{aligned} \qquad [15]$$

with

$$w_{j,i}(\lambda, \mu) = \exp[\sum_{r=i}^{j-1} \sum_{s=r+1}^{j} \alpha_{r,s} \Delta_{r,s} + \sum_{r=i}^{j} \sigma_r / k_B] \lambda^{X(j,i)} \mu^{Y(j,i)} \,. \qquad [16]$$

Here, $\alpha_{r,s} = -\varepsilon_{r,s}/k_B T$, with $\varepsilon_{r,s}$ being the energy gain by a native contact, σ_r is the entropic cost for a residue to take the native configuration, and $S_0 = \sum_r \sigma_r$, where we give the expression for the generic case that $\varepsilon_{r,s}$ and σ_r are dependent on residue types and positions, respectively, and in this paper we discuss the case of $\varepsilon_{r,s} = \varepsilon$ and $\sigma_r = \sigma$. $X(j,i)$ and $Y(j,i)$ are defined as

$$X(j,i) = \sum_{k=i}^{j} \eta_{\mathrm{I}}(k) \text{ with } \eta_{\mathrm{I}}(k) = \begin{cases} 1 & \text{for } 1 \le k \le n_{\mathrm{I}} \\ 0 & \text{otherwise} \end{cases},$$

and

$$Y(j,i) = \sum_{k=i}^{j} \eta_{\mathrm{II}}(k) \text{ with } \eta_{\mathrm{II}}(k) = \begin{cases} 1 & \text{for } n_{\mathrm{I}} + 1 \le k \le N \\ 0 & \text{otherwise} \end{cases}.$$

References

1. Go N, Abe H (1981) Noninteracting local-structure model of folding and unfolding transition in globular proteins. I. Formulation. *Biopolymers* 20:991-1011.
2. Bruscolini P, Pelizzola A (2002) Exact solution of the Muñoz-Eaton model for protein folding. *Phys Rev Lett* 88:258101.