Investigation of routes and funnels in protein folding by free energy functional methods: Supplementary Material

Steven S. Plotkin and José N. Onuchic

Department of Physics, University of California, San Diego

SUPPLEMENTARY MATERIAL We include below some physically motivated arguments which indicate how ordering (folding) different parts of the protein heterogeneously lowers the thermodynamic barrier.

Consider making random energetic perturbations on the contact energies of an initially homogeneous idealized system with free energy barrier $F_{\rm HOMO}$ and folding rate $k_o \exp(-F_{\rm HOMO}/T)$. If the total native energetic variance is $\Delta E_{\rm N}^2$, the variance at the transition state is $\cong Q^{\ddagger}\Delta E_{\rm N}^2$. Approximating the transition state as an ensemble of states with uncorrelated energies (1), and considering only the effects of changing native interactions, the energy will always decrease twice as much as the entropy times the temperature under the influence of heterogeneity, and thus the free energy barrier decreases: $\delta F(T) = \delta E(T) - T\delta S(T) = -Q^{\ddagger}\Delta E_{\rm N}^2/T - Q^{\ddagger}\Delta E_{\rm N}^2/2T = -Q^{\ddagger}\Delta E_{\rm N}^2/2T$, which increases the rate as

$$k_f \approx k_{\rm HOMO} \exp\left(\frac{Q^{\ddagger}\Delta E_{\rm N}^2}{2T^2}\right) . \qquad (1)$$

This crude argument yields essentially the same result as a more detailed analysis (see eq. (6) of the text), thus even for an initially fully symmetric funnel, introducing arbitrary heterogeneity lowers the barrier.

Consider next the free energy change in making equal and opposite perturbations on the energies of two contacts with different formation probability, say $Q_1 > Q_2$ (2). The free energy change δF in perturbing a contact i's energy is $Q_i \delta\epsilon_i$, analogous to the free energy

change in a magnet due to the coupling of a spin with its local external field: $\delta F = -m_i \delta h_i$. Neglecting explicit changes in the entropy is acceptable (c.f. discussion before eq. (5)). So the change in the free energy barrier $\delta\Delta F^\dagger$ is roughly $Q_1\delta\epsilon - Q_2\delta\epsilon$ (letting $Q_1 = Q_2 \cong 0$ in the unfolded state). Thus if $Q_1 > Q_2$, $\delta\Delta F^\dagger < 0$ if $\delta\epsilon < 0$, i.e. if the more probable contact is made stronger (more negative), which in turn increases its probability of formation. So the barrier is lowered by increasing the dispersion in contact formation. For a large number of contacts the first order change in barrier height is $\delta\Delta F^\dagger = \sum_i^M Q_i \delta\epsilon_i$, which is a sum of a large number of random uncorrelated terms, so $\delta\Delta F^\dagger$ is Gaussianly distributed. The mean of this distribution is zero since:

$$\overline{\delta\Delta F^\dagger} = \sum_i^M \overline{Q_i \delta\epsilon_i} = M\overline{Q}\;\overline{\delta\epsilon} \tag{2}$$

and $\overline{Q} = (1/M)\sum_i^M Q_i = Q$ and $\overline{\delta\epsilon} = (1/M)\sum_i^M \delta\epsilon_i = 0$. The standard deviation

$$\sqrt{\overline{(\delta\Delta F^\dagger)^2}} = \sqrt{MQ\,(1-Q)}\,b \tag{3}$$

scales like $\sqrt{N}$ since $M = zN$. Similar arguments of the effects of heterogeneity on the barrier were considered in (3).

Another argument makes use of thermodynamic perturbation theory (4). Consider a Gō model with M contacts, whose configurational states are perturbed in energy by a random contribution $V_i \equiv \delta E_i$ so that the new energy of state i is $E_i = E_i^o + V_i$. Let the native energy be unchanged: $\delta E = 0$ in the native state. Then the change in free energy to second order in V is

$$\delta\Delta F(Q) = \overline{V} - \frac{1}{2T}\left\langle \left(V - \overline{V}\right)^2 \right\rangle \tag{4}$$

where

$$\overline{V} = \frac{1}{Z}\sum_{c\in Q}{}' V_c \exp(-E_c^o/T) = \langle \delta E\rangle_o' \tag{5}$$

is calculated by summing over all configurations c having Q native contacts. It is instructive to show that $\langle\delta E\rangle_o' = \delta\Delta F$ mentioned above. First note that the average occupation probability of contact i is

$$Q_i(Q) = \sum_{c\in Q}{}' \delta(i,c)\frac{\exp(-E_c^o/T)}{Z} = \langle\delta_i\rangle_o \tag{6}$$

where $\delta(i,c) = 1$ if contact i is made in configuration c, and $\delta(i,c) = 0$ otherwise. Next note that

$$\frac{\partial F}{\partial \epsilon_i} = \frac{\partial}{\partial \epsilon_i}(-T\ln Z) = \frac{1}{Z}\sum_c{}' \frac{\partial E_c}{\partial \epsilon_i}\mathrm{e}^{-E_c/T}\ , \tag{7}$$

and since the energy of conformation c is a sum of its contact energies: $E_c = \sum_{j\in c}\epsilon_j$, $\partial E_c/\partial\epsilon_i = \delta(i,c)$ and thus $\partial F/\partial\epsilon_i = Q_i$ as noted above. Finally then

$$\begin{aligned}\langle\delta E\rangle_o' &= \sum_{c\in Q}{}' \delta E_c \frac{\mathrm{e}^{-E_c/T}}{Z} \\ &= \sum_{c\in Q}{}' \sum_{j\in c_{\mathrm{N}}} \delta(i,c)\delta\epsilon_j \frac{\mathrm{e}^{E_c/T}}{Z} = \sum_{j\in c_{\mathrm{N}}}\langle\delta_j\delta\epsilon_j\rangle \\ &= \sum_{j=1}^{M}\langle\delta_j\rangle\,\delta\epsilon_j = \sum_{j=1}^{M} Q_j\delta\epsilon_j\end{aligned} \tag{8}$$

which was to be demonstrated. c_{N} in the derivation is the native configuration and thus the final sum is over native contacts. Again, the typical value for the first order perturbation scales like $\sqrt{N}$. On the other hand, the second order term in (4) is proportional to $\langle\delta V^2\rangle$ and scales like N. Thus the free energy change due to random perturbations in the native energies is negative in the thermodynamic limit.

That higher order terms do not reverse the trend in barrier height can be ensured by the Peierls-Bogoliubov inequality $F \le F_o + \langle V\rangle_o$ where F_o is the free energy in absence of the random component and V is the random part of the Hamiltonian averaged over the unperturbed states, which is just the first order term in eq. (4). Thus the transition state free energy (per volume) $F(Q^\dagger)/N$ is always less than unperturbed free energy $F_o(Q^\dagger)/N$ in the thermodynamic limit; since $F(0) \cong F_o(0)$ in the unfolded state, the barrier is always lowered.

The effects of *correlations* between energetic perturbations and contact probabilities, as well as between energies and loop lengths, are described in the body of the paper.

CONFIGURATIONAL ENTROPY LOSS TO FOLD TO A GIVEN TOPOLOGICAL STRUCTURE We can write a very general form for the change in entropy due to contact formation $\mathcal{S}_{\rm BOND}(\{Q_i(Q)\}|\{Q_i(0)\})$, to go from $\{Q_i(Q=0)\}=\{0\}$, to another state having $\{Q_i(Q)\}$, as

$$\mathcal{S}_{\rm BOND} = \sum_i \int_0^Q \mathcal{D}Q_i\,(Q')\;\; s_i\,(\ell_i, \{Q_j\,(Q')\})\;. \tag{9}$$

Here $s_i(\ell_i, \{Q_j(Q)\})$ is the entropy loss to form contact i having sequence separation ℓ_i, in the presence of the contact pattern $\{Q_j(Q)\}$, which is itself parameterized through Q. Each $s_i(\ell_i, \{Q_j(Q)\})$ in eq. (9) is functionally integrated along the M-dimensional path specified by $\{Q_i(Q)\}$. However the entropy as a function of the set $\{Q_i\}$ must be a state function, meaning that the value of the integral depends only on the end points and not on the path taken. By ensuring zero curl, the condition for path independence is obtained:

$$\frac{\partial s_j}{\partial Q_i}\,(\ell_j, \{Q_k\}) = \frac{\partial s_i}{\partial Q_j}\,(\ell_i, \{Q_k\}) \quad \text{for } \; i \neq j\;. \tag{10}$$

Then the entropy difference only depends on the initial and final states and can be rewritten as a regular integral:

$$\mathcal{S}_{\rm BOND}\,(\{Q_i(Q)\}|\{0\}) = \sum_i \int_0^{Q_i} dQ_i'\; s_i\left(\ell_i, \left\{Q_j'\right\}\right)\;. \tag{11}$$

We use an approximate formula for s_i by introducing an effective loop length $\ell_{\rm EFF}(\ell_i, \{Q_j\})$ into [1] $s_i(\ell_i, \{Q_j\}) = \ln(a/\ell_{\rm EFF})^{3/2}$, and we satisfy eq. (10) by using a Hartree style ansatz for the functional form of $\ell_{\rm EFF}$:

$$\ell_{\rm EFF}(\ell_i, \{Q_k\}) = f(\ell_i)\, g(\{Q_k\}) = f(\ell_i)\, g(\frac{1}{M}\sum_k Q_k) \tag{12}$$

---

[1] We ignore here possibly important changes in the power of the ideal chain exponent 3/2, since we have not found a simple way to incorporate an exponent dependent on $\{Q_i\}$ and to simultaneously satisfy eq. (10).

so that the loop length is decreased by a function of the mean of the contact density field, $g(Q)$. The condition $\ell_{\rm EFF}(\ell_i, Q=0) = \ell_i$ gives $f(\ell_i) = \ell_i$ and $g(0) = 1$. The condition that $\ell_{\rm EFF}(\ell_i, Q=1) \approx 1$ gives $g(1) \approx 1/\overline{\ell}$ (since $g(Q)$ cannot depend on ℓ_i), where $\overline{\ell} = (1/M)\sum_i \ell_i$. We approximate the effective loop length $\ell_{\rm EFF}(\ell_i, Q)$ at Q in the spirit of the Flory mean-field result (5,6) by dividing the total loop length ℓ_i by the approximate number of bonds in the loop $\cong \overline{\ell} Q$, so that

$$\ell_{\rm EFF}(\ell_i, Q) \approx \frac{\ell_i}{(\overline{\ell}-1)Q+1}, \tag{13}$$

which has the mean-field behavior for large ℓ_i and also has approximately the right limiting behavior as $Q \to 0$ and $Q \to 1$. The line of argument followed here is accurate only for weak dispersion in loop lengths; for larger values of $\delta\ell_i$ modifications must be made because of inaccuracies at large values of Q. The coefficient a in s_i is determined from eq. (11) by the condition that the total entropy loss to fold must be the unfolded entropy: $\mathcal{S}_{\rm BOND}(\{1\}|0) = -Ns_o$ where s_o is the entropy per monomer in the unfolded state. Substituting $\ell_{\rm EFF}(\ell_i, Q)$ and a into s_i, eq. (11) gives the contact entropy loss to form the state $\{Q_i\}$:

$$\mathcal{S}_{\rm BOND} = -\frac{3}{2} M \left(\langle \delta Q \, \delta \ln \ell \rangle + S_{\rm MF}(Q, \overline{\ell}) \right) \tag{14}$$

where

$$\begin{aligned} S_{\rm MF}\left(Q, \overline{\ell}\right) &= \frac{2}{3z} Q s_o + Q \frac{\overline{\ell} \ln \overline{\ell}}{\overline{\ell}-1} \\ &\quad - \frac{1}{\overline{\ell}-1} \left[1 + (\overline{\ell}-1)Q\right] \ln \left[1 + (\overline{\ell}-1)Q\right] \end{aligned} \tag{15}$$

is only a function of the mean field Q and mean loop length $\overline{\ell}$, and

$$\langle \delta Q \, \delta \ln \ell \rangle = \frac{1}{M} \sum_i (Q_i - Q) \left(\ln \ell_i - \overline{\ln \ell} \right) \tag{16}$$

is the correlation between the fluctuations in contact probability and log loop length. By eq. (14) the entropy is *raised* above that of a symmetrically ordering system when shorter ranged contacts have higher probability to be formed; this effect lowers the barrier. As a

check, when $Q_i = Q$, $\ell_i = \overline{\ell}$ and the limit $\overline{\ell}Q \gg 1$ is taken, the bond entropy in eq. (14) becomes

$$\begin{aligned} \mathcal{S}_{\text{BOND}}^{\text{MEAN FIELD}} &= -\frac{3}{2} M S_{\text{MF}}(Q, \overline{\ell} \to \infty) \\ &\cong \frac{3MQ}{2}\left(-\frac{2s_o}{3z} + \ln Q\right), \end{aligned} \tag{17}$$

which is the Flory result derived earlier in the mean-field limit (5,6).

REFERENCES

1. Derrida, B. (1981) *Phys Rev B* **24**, 2613–2626.

2. We thank H. Nymeyer for a helpful discussion of this argument.

3. Wolynes, P. G. (1997) *Proc Nat Acad Sci USA* **94**, 6170–6175.

4. Landau, L. D & Lifshitz, E. M. (1980) *Statistical Physics.* (Pergamon Press, Oxford), 3 edition.

5. Flory, P. J. (1956) *J Am Chem Soc* **78**, 5222–5235.

6. Plotkin, S. S, Wang, J, & Wolynes, P. G. (1996) *Phys Rev E* **53**, 6271–6296.