Effect of Catenation on Protein Folding Stability

Huan-Xiang Zhou

Department of Physics and Institute of Molecular Biophysics, Florida State University, Tallahassee, FL 32306

SUPPORTING INFORMATION

Dimeric protein structures that are ready for catenation

A large number of dimeric proteins have intertwined interfaces similar to that of the M340E/L344K dimer mutant of the p53 tetramerization domain. Two such structures are shown here. Figure S1A shows the H-NS dimerization domain (Protein Data Bank entry 1ni8), and Figure S1B shows the dimerization domain of hepatocyte nuclear factor-1α (PDB entry 1g39).

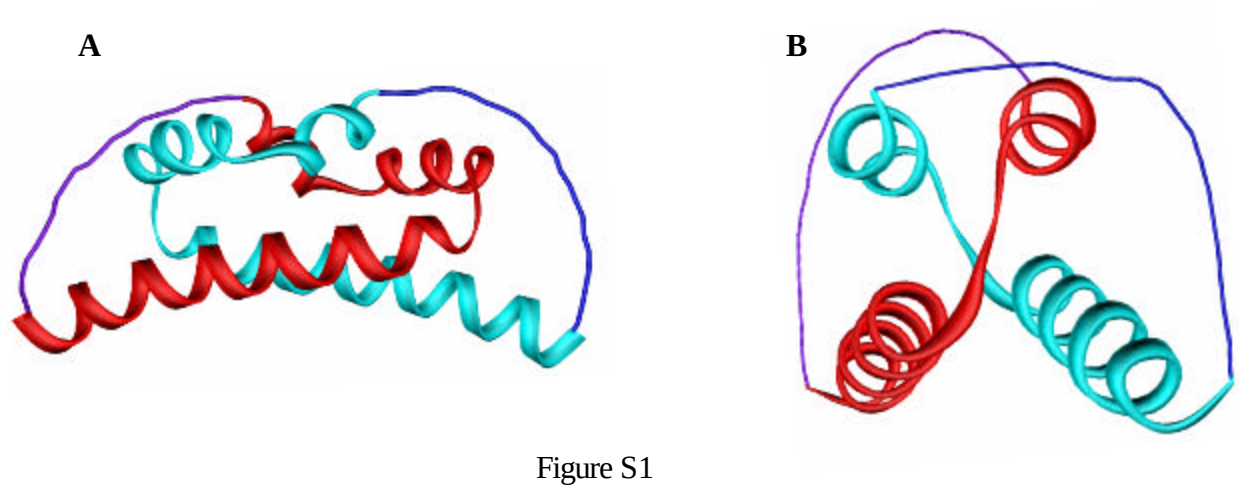

Figure S1

Theory for the Effective Concentration C_{cat}

Here eqs 2-4 of the main text are formally derived. This derivation follows closely the development of an earlier theory for the effect of backbone cyclization. First, the folding equilibrium of a linear dimeric protein is studied. Then the effect of catenation is examined.

Folding stability of a linear dimeric protein. First consider the linear variant of the first subunit. Let the vector from the N-terminal to the C-terminal be $\mathbf{r}_A$ and the remaining degrees of freedom be

collectively represented by $\mathbf{X}_A$ (see Figure 1A of main text). If the energy function of the molecule is $E_A(\mathbf{X}_A, \mathbf{r}_A)$, then the partition function in the unfolded state is

$$Z_{u,A}^{\ lin} = \int_u \exp[-bE_A(\mathbf{X}_A, \mathbf{r}_A)]d\{\mathbf{X}_A\}d^3\mathbf{r}_A$$

where the subscript "u" signifies that the integration is restricted to the phase space of the unfolded state. The partition function $Z_{u,B}$ for the linear variant of the second subunit in the unfolded state can be similarly written. If the energy function of the linear dimeric protein in the folded state is $E(\mathbf{X}_A, \mathbf{X}_B, \mathbf{r}_A, \mathbf{r}_B)$, then the corresponding partition function is

$$Z_f^{\ lin} = \int_f \exp[-bE(\mathbf{X}_A, \mathbf{X}_B, \mathbf{r}_A, \mathbf{r}_B)]d\{\mathbf{X}_A\}d\{\mathbf{X}_B\}d^3\mathbf{r}_A d^3\mathbf{r}_B$$

The folding equilibrium constant is then

$$K^{lin} = VZ_f^{\ lin} / Z_{u,A}^{\ lin} Z_{u,B}^{\ lin} \qquad \text{(S1)}$$

where V is the volume of the bulk solution.

Later use will be made of the probability densities for the end-to-end vectors $\mathbf{r}_A$ and $\mathbf{r}_B$. In the unfolded state,

$$P_{u,A}(\mathbf{r}_1)P_{u,B}(\mathbf{r}_2) = (Z_{u,A}^{\ lin} Z_{u,B}^{\ lin})^{-1} \int_u d(\mathbf{r}_A - \mathbf{r}_1)d(\mathbf{r}_B - \mathbf{r}_2)$$
$$\times \exp\{-b[E_A(\mathbf{X}_A, \mathbf{r}_A) + E_B(\mathbf{X}_B, \mathbf{r}_B)]\}d\{\mathbf{X}_A\}d\{\mathbf{X}_B\}d^3\mathbf{r}_A d^3\mathbf{r}_B$$

Rearrangement leads to

$$\int_u \exp\{-b[E_A(\mathbf{X}_A, \mathbf{r}_1) + E_B(\mathbf{X}_B, \mathbf{r}_2)]\}d\{\mathbf{X}_A\}d\{\mathbf{X}_B\} = Z_{u,A}^{\ lin} Z_{u,B}^{\ lin} P_{u,A}(\mathbf{r}_1)P_{u,B}(\mathbf{r}_2) \qquad \text{(S2)}$$

In the folded state, in principle the distributions of $\mathbf{r}_A$ and $\mathbf{r}_B$ are coupled. In analogy to eq S2, one has

$$\int_f \exp\{-bE(\mathbf{X}_A, \mathbf{X}_B, \mathbf{r}_1, \mathbf{r}_2)]\}d\{\mathbf{X}_A\}d\{\mathbf{X}_B\} = Z_f^{\ lin} P_f(\mathbf{r}_1, \mathbf{r}_2)$$

In practice, $\mathbf{r}_A$ and $\mathbf{r}_B$ fluctuate around well-defined mean displacements ($\mathbf{d}_1$ and $\mathbf{d}_2$ in Figure 1A). To a good approximation, their distributions can be represented by delta functions:

$$P_f(\mathbf{r}_1, \mathbf{r}_2) = d(\mathbf{r}_1 - \mathbf{d}_1)d(\mathbf{r}_2 - \mathbf{d}_2)$$

Effect of catenation. Now consider the situation where the N- and C-terminals of each linear subunit are connected by a peptide linker (see Figure 1B, upper branch). For the folded state, in the simplest case, the linkers and the linear dimeric protein do not interfere with each other, except that the end-to-

end distances of the linkers are restricted to those between the N- and C-terminals. Then the partition function for the catenated protein in the folded state can be written as

$$\begin{aligned} Z_{\mathrm{f}}^{\mathrm{cat}} &= \int_{\mathrm{f}} p_{\mathrm{A}}(\mathbf{r}_{\mathrm{A}})\, p_{\mathrm{B}}(\mathbf{r}_{\mathrm{B}}) \exp[-bE(\mathbf{X}_{\mathrm{A}}, \mathbf{X}_{\mathrm{B}}, \mathbf{r}_{\mathrm{A}}, \mathbf{r}_{\mathrm{B}})] d\{\mathbf{X}_{\mathrm{A}}\} d\{\mathbf{X}_{\mathrm{B}}\} d^3\mathbf{r}_{\mathrm{A}} d^3\mathbf{r}_{\mathrm{B}} \\ &= Z_{\mathrm{f}}^{\mathrm{lin}} \int_{\mathrm{f}} p_{\mathrm{A}}(\mathbf{r}_{\mathrm{A}})\, p_{\mathrm{B}}(\mathbf{r}_{\mathrm{B}}) P_{\mathrm{f}}(\mathbf{r}_{\mathrm{A}}, \mathbf{r}_{\mathrm{B}})] d^3\mathbf{r}_{\mathrm{A}} d^3\mathbf{r}_{\mathrm{B}} \\ &= Z_{\mathrm{f}}^{\mathrm{lin}} p_{\mathrm{A}}(\mathbf{d}_1)\, p_{\mathrm{B}}(\mathbf{d}_2) \end{aligned} \tag{S3}$$

where $p_{\mathrm{A}}(\mathbf{r})$ and $p_{\mathrm{B}}(\mathbf{r})$ are the probability densities for the end-to-end vectors of the two peptide linkers.

A primary effect of catenation in the unfolded state is to keep the two circularized chains interlocked so they cannot move away from each other. The Boltzmann factor for the unfolded state can be approximated by

$$\exp[-bE(\mathbf{X}_{\mathrm{A}}, \mathbf{X}_{\mathrm{B}}, \mathbf{r}_{\mathrm{A}}, \mathbf{r}_{\mathrm{B}})] = \exp\{-b[E_{\mathrm{A}}(\mathbf{X}_{\mathrm{A}}, \mathbf{r}_{\mathrm{A}}) + E_{\mathrm{B}}(\mathbf{X}_{\mathrm{B}}, \mathbf{r}_{\mathrm{B}})]\} H(\mathbf{X}_{\mathrm{A}}, \mathbf{X}_{\mathrm{B}}, \mathbf{r}_{\mathrm{A}}, \mathbf{r}_{\mathrm{B}})$$

where H is 1 if the dimer of circularized chains has the catenane topology and 0 otherwise. The partition function in the unfolded state is thus

$$\begin{aligned} Z_{\mathrm{u}}^{\mathrm{cat}} &= \int p_{\mathrm{A}}(\mathbf{r}_{\mathrm{A}}) p_{\mathrm{B}}(\mathbf{r}_{\mathrm{B}}) \exp\{-b[E_{\mathrm{A}}(\mathbf{X}_{\mathrm{A}}, \mathbf{r}_{\mathrm{A}}) + E_{\mathrm{B}}(\mathbf{X}_{\mathrm{B}}, \mathbf{r}_{\mathrm{B}})]\} \\ &\times H(\mathbf{X}_{\mathrm{A}}, \mathbf{X}_{\mathrm{B}}, \mathbf{r}_{\mathrm{A}}, \mathbf{r}_{\mathrm{B}}) d\{\mathbf{X}_{\mathrm{A}}\} d\{\mathbf{X}_{\mathrm{B}}\} d^3\mathbf{r}_{\mathrm{A}} d^3\mathbf{r}_{\mathrm{B}} \end{aligned} \tag{S4}$$

The folding equilibrium constant is

$$K^{\mathrm{cat}} = Z_{\mathrm{f}}^{\mathrm{cat}} / Z_{\mathrm{u}}^{\mathrm{cat}} \tag{S5}$$

Treatment of the unfolded catenane. The collective degrees of freedom, **X**, of a subunit can be further divided into a vector **R** representing the position, a vector w representing the orientation, and **X**' representing internal degrees of freedom. The energy functions $E_{\mathrm{A}}(\mathbf{X}_{\mathrm{A}}, \mathbf{r}_{\mathrm{A}})$ and $E_{\mathrm{B}}(\mathbf{X}_{\mathrm{B}}, \mathbf{r}_{\mathrm{B}})$ do not depend on the position and orientation vectors. To simplify, it may be further assumed that the H function only depends on the position and orientation vectors of the two unfolded chains:

$$H(\mathbf{X}_{\mathrm{A}}, \mathbf{X}_{\mathrm{B}}, \mathbf{r}_{\mathrm{A}}, \mathbf{r}_{\mathrm{B}}) = h(\mathbf{R}_{\mathrm{BA}}, \mathrm{w}_{\mathrm{A}}, \mathrm{w}_{\mathrm{B}})$$

where $\mathbf{R}_{\mathrm{BA}} = \mathbf{R}_{\mathrm{B}} - \mathbf{R}_{\mathrm{A}}$. Then eq S4 becomes

$$\begin{aligned} Z_{\mathrm{u}}^{\mathrm{cat}} &= \int p_{\mathrm{A}}(\mathbf{r}_{\mathrm{A}}) p_{\mathrm{B}}(\mathbf{r}_{\mathrm{B}}) \exp\{-b[E_{\mathrm{A}}(\mathbf{X}_{\mathrm{A}}, \mathbf{r}_{\mathrm{A}}) + E_{\mathrm{B}}(\mathbf{X}_{\mathrm{B}}, \mathbf{r}_{\mathrm{B}})]\} d\{\mathbf{X}_{\mathrm{A}}\} d\{\mathbf{X}_{\mathrm{B}}\} d^3\mathbf{r}_{\mathrm{A}} d^3\mathbf{r}_{\mathrm{B}} \\ &\times \int h(\mathbf{R}_{\mathrm{BA}}, \mathrm{w}_{\mathrm{A}}, \mathrm{w}_{\mathrm{B}}) d^3\mathbf{R}_{\mathrm{BA}} d^3\mathrm{w}_{\mathrm{A}} d^3\mathrm{w}_{\mathrm{B}} / (8\pi^2)^2 V \end{aligned}$$

$$= Z_{u,A}^{\ lin} Z_{u,A}^{\ lin} \int p_A(\mathbf{r}_A)\, p_B(\mathbf{r}_B) P_{u,A}(\mathbf{r}_A) P_{u,B}(\mathbf{r}_B) d^3\mathbf{r}_A d^3\mathbf{r}_B$$
$$\times \int h(\mathbf{R}_{BA}, w_A, w_B) d^3\mathbf{R}_{BA} d^3 w_A d^3 w_B / (8\pi^2)^2 V \quad \text{(S6)}$$

Inserting eqs S3 and S6 into eq S5, one finds

$$K^{cat} = K^{lin} q_{cycl,A} q_{cycl,B} C_{il} \quad \text{(S7)}$$

where $q_{cycl,\,A}$ and $q_{cycl,\,B}$ are the enhancements in folding stability by the backbone cyclization of the subunits,

$$q_{cycl,A} = p_A(\mathbf{d}_1) / \int p_A(\mathbf{r}) P_{u,A}(\mathbf{r}) d^3\mathbf{r} \quad \text{(S8a)}$$

$$q_{cycl,B} = p_B(\mathbf{d}_2) / \int p_B(\mathbf{r}) P_{u,B}(\mathbf{r}) d^3\mathbf{r} \quad \text{(S8b)}$$

and C_{il} is the effective concentration due to the interlocking of the two unfolded chains,

$$C_{il}^{-1} = \int h(\mathbf{R}_{BA}, w_A, w_B) d^3\mathbf{R}_{BA} d^3 w_A d^3 w_B / (8\pi^2)^2 \quad \text{(S9)}$$

Eq 2 is essentially eqs S8a and S8b while eq 4 is equivalent to eq S7. When eq S9 is specialized to two interlocked rigid circular rings, eq 3 is obtained.