J|A|C|S ARTICLES

Published on Web 01/26/2009

Theory for Protein Folding Cooperativity: Helix Bundles

Kingshuk Ghosh† and K. A. Dill*

Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94158

Received October 15, 2008; E-mail: dill@maxwell.ucsf.edu

Abstract: We present a theory for protein folding stability and cooperativity for helix bundle proteins. We treat the individual helices with a Schellman−Zimm−Bragg-like approach, using nucleation and propagation quantities, and we treat the hydrophobic and van der Waals contacts between the helices as a binding equilibrium. Predictions are in good agreement with experiments on both thermal and urea-induced transitions of (1) molecules that can undergo single helix-to-coil transitions for various chain lengths and (2) three-helix-bundle proteins A and α3C. The present model addresses a problem raised by Kaya and Chan that proteins fold more cooperatively than previous models predict. The present model correctly predicts the experimentally observed two-state cooperativities, $\Delta H_{van't\ Hoff}/\Delta H_{cal} \approx 1$, for helix-bundle proteins. The predicted folding cooperativity is greater than that of helix formation alone, or collapse alone, because of the nonlinear coupling between the tertiary interactions and the helical interactions.

I. Introduction

We present here a theory of protein stability and cooperativity. We focus on helix-bundle proteins. Protein folding involves both secondary structure formation and collapse. Historically, two types of models have been prominent in explaining conformational cooperativity in proteins and polymers. First, *helix−coil models* treat the sharp transition that some polymers undergo from a disordered random coil state to a single helix.[1−7] Such processes are dominated by the local interactions among nearest neighbors in the chain. Helix−coil experiments are typically well described in terms of the two parameters of helix−coil theory: σ, a nucleation parameter, and s, a helix propagation parameter.[8] Second, *polymer collapse theories* treat the sharp condensation collapse transition of hydrophobic polymers in water.[9−17] Collapse processes are dominated by nonlocal interactions—the solvent-mediated contacts among pairs of monomers that need not be adjacent in the chain sequence. While both types of models have given important insights, a deeper understanding of the folding cooperativity of proteins requires an approach that treats both local and nonlocal interactions within the same theoretical framework.

The underpinnings of cooperativity have recently been of considerable interest. Chan and co-workers have been at the forefront of protein folding cooperativity by comparing different theoretical models.[17−21] This question has also been the subject of experimental controversy in the matter of whether proteins undergo downhill folding with no barrier.[22−27] It is a challenge for experiments to determine the density of states, with the exception of some recent work using NMR[24] or FRET[27] studies. Hence, there remains a need for a microscopic statistical mechanical model of the density of states and folding cooperativity.

There is also an experimental challenge in understanding protein folding cooperativity because a given protein differs from the next in nonsystematic ways: different proteins have different chain lengths, different secondary structures, different numbers of hydrogen bonds and hydrophobic interactions, and different packing densities. There is no simple single variable (i.e., "knob") that can systematically vary the folding cooperativity

† Present address: Department of Physics, University of Denver, Denver, CO 80208.

(1) Schellmann, J. *J. Phys. Chem.* **1958**, *62*, 1485–1494.
(2) Zimm, B.; Bragg, J. *J. Chem. Phys.* **1958**, *28*, 1246−1247; **1959**, *31*, 526−535.
(3) Zimm, B. H.; Doty, P.; Iso, K. *Proc. Natl. Acad. Sci. U.S.A.* **1959**, *45* (11), 1601–1604.
(4) Lifson, S.; Roig, A. *J. Chem. Phys.* **1961**, *34*, 1963–1974.
(5) Polan, D.; Scheraga, H. *Theory of Helix-Coil Transitions in Biopolymers*; Academic Press: New York, 1970.
(6) Munoz, V.; Serrano, L. *Nat. Struct. Biol.* **1994**, *1*, 399–409.
(7) Munoz, V.; Serrano, L. *Curr. Opin. Biotechnol.* **1995**, *6*, 382–386.
(8) Scholtz, J. M.; Baldwin, R. L. *Annu. Rev. Biophys. Biomol. Struct.* **1992**, *21*, 95–118.
(9) Dill, K. *Biochemistry* **1985**, *24*, 1501–1509.
(10) Dill, K.; Stigter, D. *Adv. Protein Chem.* **1995**, *46*, 59–104.
(11) Finkelstein, A.; Shakhnovich, E. *Biopolymers* **1989**, *28*, 1667–1680.
(12) Saven, J. G.; Wolynes, P. G. *J. Mol. Biol.* **1996**, *257*, 199–216.
(13) Saven, J. G.; Wolynes, P. G. *Physica D* **1997**, *107*, 330–337.
(14) Grosberg, A. Yu.; Khokhlov, A. R. *Statistical Physics of Macromolecules*; AIP Press: New York, 1994.
(15) Pande, V. S.; Grosberg, A. Y.; Tanaka, T. *Rev. Mod. Phys.* **2000**, *72*, 259–314.
(16) Lucas, A.; Huang, L.; Joshi, A.; Dill, K. A. *J. Am. Chem. Soc.* **2007**, *129*, 4272–4281.
(17) Moghaddam, M. S.; Shimizu, S.; Chan, H. S. *J. Am. Chem. Soc.* **2005**, *127*, 303–316.
(18) Knott, M.; Chan, H. *Chem. Phys.* **2004**, *307*, 187–199.
(19) Chan, H.; Shimizu, S.; Kaya, H. *Methods Enzymol.* **2004**, *380*, 350–379.
(20) Chan, H. *Proteins: Struct., Funct., Genet.* **2000**, *40*, 543–571.
(21) Kaya, H.; Chan, H. S. *Proteins: Struct., Funct., Genet.* **2000**, *40*, 637–661.
(22) Kelly, J. W. *Nature* **2006**, *442*, 255–256.
(23) Zhou, Z.; Bai, Y. W. *Nature* **2007**, *445*, E16−E17.
(24) Sadqi, M.; Fushman, D.; Munoz, V. *Nature* **2006**, *442*, 317–321.
(25) Sadqi, M.; Fushman, D.; Munoz, V. *Nature* **2007**, *445*, E17−E18.
(26) Ferguson, N.; Sharpe, T. D.; Johnson, C. M.; Schartau, P. J.; Fersht, A. R. *Nature* **2007**, *445*, E14−E15.
(27) Huang, F.; Sato, S.; Sharpe, T. D.; Ying, L.; Fersht, A. R. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 123–127.

10.1021/ja808136x CCC: $40.75

of a protein. In contrast, our understanding of helix−coil processes was considerably advanced because of experiments that have systematically explored the effects of changing the chain length N and the propagation parameter s.[8,28,29] Also, unlike helix−coil or collapse processes in simple model polymers, the chain length of a protein is seldom the most important variable controlling its folding cooperativity.

However, there is a class of proteins in which, in principle, cooperativity can be studied systematically. These are helix-bundle proteins. Modern methods now allow for the synthesis of simple repeating helical sequences of various numbers and lengths that can be connected by loops into bundles. Experiments by Hecht,[30,31] Degrado,[32] and others,[33] for example, show that it is straightforward to design helix-bundle folds: you mainly need hydrophobic residues on the inside and polar ones (for solubility) on the outside. Moreover, foldable helix-bundle polymers have also been made using nonbiological backbones, such as in peptoids.[34,35] Despite these possibilities, however, as far as we know there are not yet systematic investigations of folding stability and cooperativity of experimental model helix-bundle systems.

Here, we develop an analytical theory for the equilibrium properties of helices and helix-bundle proteins. We first describe below our treatment of a single helix, since the helix-bundle models that follow rely upon it. Our helix-coil treatment here differs slightly from most earlier models, such as Schellman's model,[1] the Zimm−Bragg model,[2] or the Lifson−Roig model.[4] In these classical models, the only entropy that is treated explicitly is the "combinatoric entropy", which is the count of the number of different locations where helical residues can be located in the sequence. For example, three consecutive helical units can occur in a chain of five monomers in three ways, CCHHH, CHHHC, and HHHCC, where C represents a segment of chain that is in a coil configuration and H represents a helical bond. However, this combinatoric entropy is only one of the contributors to the entropy of a polymer chain. Another important entropy—the chain conformational entropy—is not treated explicitly in the classical helix−coil models. Our approach keeps the chain entropy explicit, because (1) it is essential for more complex helix-bundle treatment that follows, (2) it gives insights into the temperature dependences, and (3) it allows for predictions of observables such as the radius of gyration which are not otherwise available from helix−coil theories.

(28) (a) Huyghues-Despointes Beatrice, M. P.; Scholtz, J. M.; Baldwin, R. L. *Protein Sci.* **1993**, *2*, 80−85, 1604−1611. (b) Huyghues-Despointes Beatrice, M. P.; Klingler, T. M.; Baldwin, R. L. *Biochemistry* **1995**, *34*, 13267–13271.
(29) (a) von Dreele, P. H.; Poland, D.; Scheraga, H. A. *Macromolecules* **1971**, *4*, 396–407. (b) von Dreele, P. H.; Lotan, N.; Ananthanarayanan, V. S.; Andreatta, R. H.; Poland, D.; Scheraga, H. A. *Macromolecules* **1971**, *4*, 408–417. (c) Ananthanarayanan, V. S.; Andreatta, R. H.; Poland, D.; Scheraga, H. A. *Macromolecules* **1971**, *4*, 417–424. (d) Platzer, K. E. B.; Ananthanarayanan, V. S.; Andreatta, R. H.; Scheraga, H. A. *Macromolecules* **1972**, *5*, 177–187.
(30) Kamtekar, S.; Schiffer, J. M.; Xiong, H. Y.; Babik, J. M.; Hecht, M. H. *Science* **1993**, *262*, 1680–1685.
(31) Brunet, A. P.; Huang, E. S.; Huffine, M. E.; Loeb, J. E.; Weltman, R. J.; Hecht, M. H. *Nature* **1993**, *364*, 355–358.
(32) (a) Regan, L.; Degrado, W. F. *Science* **1988**, *241*, 976–978. (b) Bryson, J.; Betz, S.; Lu, H.; Suich, D.; Zhou, H.; Oneil, K.; Degrado, W. *Science* **1995**, *270*, 935–941.
(33) Johansson, J. S.; Gibney, B. R.; Skalicky, J. J.; Wand, A. J.; Dutton, P. L. *J. Am. Chem. Soc.* **1998**, *120*, 3881–3886.
(34) Burkoth, T.; Beausoleil, E.; Kaur, S.; Tang, D.; Cohen, F.; Zuckermann, R. *Chem. Biol.* **2002**, *9*, 647–654.
(35) Lee, Y. C.; Zuckermann, R. N.; Dill, K. A. *J. Am. Chem. Soc.* **2005**, *127*, 10999–11009.

Figure 1. Different conformations considered in the partition sum for a single helix. A1 denotes a complete coil, A2 is for a partially formed helix, and A3 denotes a fully formed helix.

In recent years, a key focus of helix−coil models has been on predicting how helical stability depends on the amino acid sequence of a protein. To treat such dependences, such theories typically utilize transfer matrices. For example, the Zimm−Bragg and Lifson−Roig models use the 1-dimensional Ising matrix method. Such matrix treatments are also useful for treating multiple noninteracting helical stretches within a longer chain; this was particularly important in the early years for the proof of principle of helix−coil cooperativity in poly(benzyl glutamate) chains having 1500 monomers.[2] However, for treating the lengths of helices found in globular proteins, which are typically less than about 20 monomers long, much simpler models are possible and, for applications of interest to us, more desirable. Here, we make this "single-helix" simplification.

We start with a Schellman-like partition function for a helix−coil process and then introduce terms into the statistical physics that account for the interactions among pairs of helices to understand two-helix-bundle (2hb) and three-helix-bundle (3hb) molecules.

II. Partition Function for a Single Helix

Consider a chain molecule that can have a maximum number N of helical bonds and for which z is the number of rotameric configurations accessible to each backbone virtual bond. M is the number of amino acid residues in the protein molecule. These quantities are related by $N = M - 4$ since residue i forms a helical hydrogen bond with residue $i + 4$. In this section, we consider a molecule that undergoes a transition from a coil to a single helix. We call this a one-helix molecule to distinguish it from two-helix and three-helix bundles below. We factorize the total partition function $q_{1\mathrm{tot}}$

$$q_{1\mathrm{tot}} = q_{\mathrm{p}}(N) q_{1\mathrm{c}} \tag{1}$$

into a product of two terms. The first term, q_{p}, is the total count of all the polymer chain conformations:

$$q_{\mathrm{p}}(N) = (z-1)^{N+2} \tag{2}$$

The factor of z is the total number of conformations of the virtual bonds in one helical turn, one conformation of which is helical and $z - 1$ of which are coil conformations. The second term, which accounts for the combinatorics, can be expressed as a sum of Boltzmann factors over all the helical and coil states of a chain that can form a single helix (see Figure 1):

$$\begin{aligned} q_{1\mathrm{c}}(N,T) &= 1 + \sigma[Ns + (N-1)s^2 + (N-2)s^3 + \ldots s^N] \\ &= 1 + \sigma \sum_{i=1}^{N} (N-i+1)s^i \\ &= 1 + \sigma \frac{s^{N+2} - (N+1)s^2 + (N)s}{(s-1)^2} \quad \text{for} \quad s \neq 1 \\ &= 1 + \sigma \frac{(N)(N+1)}{2} \quad \text{for} \quad s = 1 \end{aligned} \tag{3}$$

where s is the equilibrium constant for forming each helical bond relative to a coil unit and σ is the nucleation parameter, i.e., the equilibrium constant for forming the first H after a string

of C's. The first term (1) in eq 3 is the statistical weight for the coil state; i.e., it counts all the chain conformations that have all C's and no helices. The second term, $N\sigma s$ expresses that there are N different locations in the chain that can have a single helical bond. The third term, $(N-1)\sigma s^2$ expresses that there are $N-1$ locations in the chain sequence at which an HH pair can appear in a string of N monomers. The factor of σ accounts for the difficulty of nucleating the helix, i.e., forming the first H. The factor of s^2 is an equilibrium constant for having two consecutive H's in the helix. Each term in the partition function is made up from factors in this way. This model is essentially identical to the model originally introduced many years ago by Schellman,[1] which makes the "single-helix" approximation, except that we treat the temperature dependence a little differently, as indicated below. To make explicit the full temperature dependence, we express the helix-propagation equilibrium constant s, or unit partition function, in terms of its energetic and entropic parts:

$$s = \frac{e^{\varepsilon_{hb}/kT}}{z-1} \tag{4}$$

where $\varepsilon_{hb} > 0$ is the interaction energy increase upon breaking one helical bond, k is Boltzmann's constant, and T is the absolute temperature. This expression for s shows that (a) there is an entropic cost (the factor of $z-1$ in the unit equilibrium constant) when the chain forms a helical bond from among the possible coil conformations and (b) there is an energetic advantage for forming the helix. Helix−coil theories usually treat σ as being a temperature-independent equilibrium coefficient and thus as resulting from an entropy. Here, we treat it more broadly as a free energy, since helix nucleation appears to be thermally activated; hence

$$\sigma = \frac{1}{(z-1)^2} e^{-\varepsilon_{nuc}/kT} \tag{5}$$

The enthalpic barrier is ε_{nuc}, and the entropic component cost is $(z-1)^2$ because the first bond of a helix points in an arbitrary direction and the second and third virtual bonds must then be restricted to the correct orientation to form a helix. Now, to compute the properties of this helix−coil model, we need the probabilities of the various states. From the model, the probability that a chain has $i \neq 0$ helical contacts is

$$p_i = \frac{(N-i+1)\sigma s^i}{q_{1c}} \tag{6}$$

and $p_0 = {q_{1c}}^{-1}$. Several experimental properties are of interest for individual helices and for helix bundles, including the average fractional helicity Θ, the average energy $\langle E \rangle$, and the heat capacity C_v. We get these quantities from the model using standard expressions:[38]

$$\Theta = \frac{\langle i \rangle}{N} = \frac{1}{N}\frac{d \ln q}{d \ln s} \tag{7}$$

$$\langle E \rangle = kT^2 \frac{d \ln q}{dT} \tag{8}$$

$$C_v = \frac{d\langle E \rangle}{dT} = 2kT\frac{d \ln q}{dT} + kT^2\frac{d^2 \ln q}{dT^2} \tag{9}$$

where q is the partition sum of the system.

When computing $\langle E \rangle$ and C_v, we substitute the combinatoric part, q_{1c}, for q in eqs 8 and 9 since q_p does not depend on the temperature.

We are interested in how the helix−coil equilibrium is affected by the temperature and in how the helix−coil equilibrium is affected by denaturing and stabilizing solvents. For example, for denaturants, we adopt the standard expression[36−38] that the interaction strength ε is a linear function of the urea concentration:

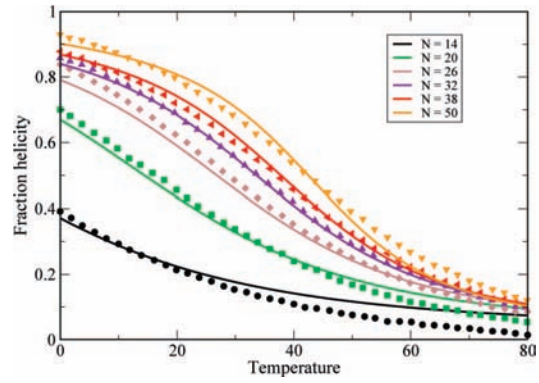

Figure 2. Thermal denaturation data from ref 39 vs theory. In the model, we use $z = 6.83$, $\varepsilon_{hb} = 1.14$ kcal/(mol of residue), and $\varepsilon_{nuc} = 1.08$ kcal/(mol of residue). Different colors are for different chain lengths.

$$\varepsilon_{hb} = \varepsilon_{hb}^0 - mc \tag{10}$$

where ε_{hb}^0 represents the interaction energy in the absence of urea and c is the concentration of urea. For some types of denaturant, higher order terms may be needed at high concentrations.

A. Predictions for the Single-Helix-to-Coil Process. While the main point of this paper is to treat helix-bundle proteins, described below, we first validate that this simple Schellman-like one-helix model adequately describes the transition between a single helix and its coil states. There is much more data now than when Schellman first developed this kind of model. Here is how we apply it. First, for a given peptide, we know the chain length N and the temperature T. A particular chain sequence will be characterized in our model by three parameters: a nucleation parameter ε_{nuc} that is averaged over the different types of sequence monomers, an average chain flexibility z, and an average helical turn energy ε_{ht}. This is the same number of parameters that would be used in other helix-coil models, such as the Zimm−Bragg model, when the temperature dependence is of interest. For a given monomer sequence, we take these three quantities and m to be fit parameters. Figures 2 and 3 show the model predictions are in good agreement with the experimentally observed temperature and urea denaturation for different lengths of a given peptide sequence.[39,40] The fits for the thermal denaturation data and urea denaturation data for different chain lengths were obtained from a single set of parameter values.

We also use this model to predict the specific heat vs temperature. Figure 6 compares the predictions with experimental data on Baldwin's peptide[41] using the same parameters that were used for the thermal and urea denaturation curves.

Because the model described above does not fit the heat capacity data accurately, we first explore an improved version

(36) Dill, K. A.; Alonso, D. O. V. *Biochemistry* **1991**, *30*, 5974–5985.
(37) Greene, R. F., Jr.; Pace, C. N. *J. Biol. Chem.* **1974**, *249*, 5388–5393.
(38) Dill, K. A.; Bromberg, S. *Molecular Driving Forces*; Garland Science: Oxford, U.K., 2002.
(39) Scholtz, J. M.; Qian, H.; York, E. J.; Stewart, J. M.; Baldwin, R. L. *Biopolymers* **1991**, *31*, 1463–1470.
(40) Scholtz, J. M.; Barrick, D.; York, E. J.; Stewart, J. M.; Baldwin, R. L. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 185–189.
(41) Scholtz, J. M.; Marqusee, S.; Baldwin, R. L.; York, E. J.; Stewart, J. M.; Santoro, M.; Bolen, D. W. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 2854–2858.

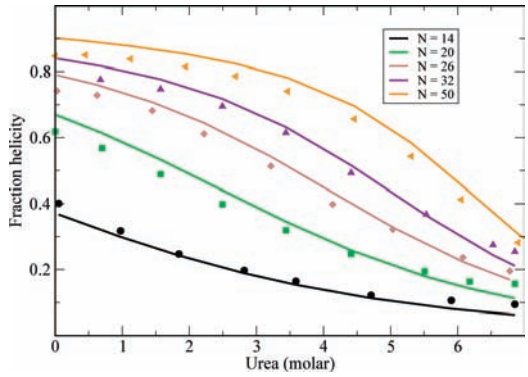

Figure 3. Urea denaturation from Schotlz et al.[40] In the model, we use $z = 6.83$, $\varepsilon_{hb} = 1.14$ kcal/(mol of residue), $\varepsilon_{nuc} = 1.08$ kcal/(mol of residue), and $m = 0.028$ kcal/M. Different colors are for different chain lengths.

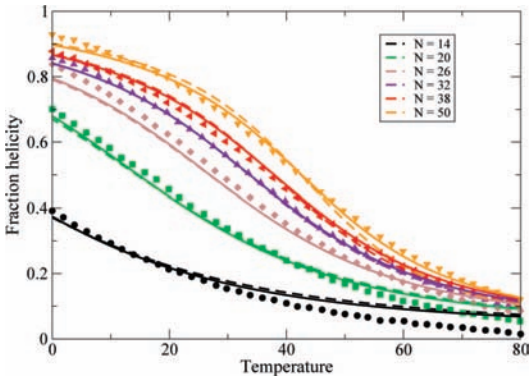

Figure 4. Thermal denaturation of peptides of different lengths from Scholtz et al.[40] Symbols represent experimental data, dashed lines denote best fit lines using the model with the single-helix approximation, and solid lines represent the best fit using the two-helical-segment model. In this model, we use $z = 7.43$, $\varepsilon_{hb} = 1.2$ kcal/(mol of residue), and $\varepsilon_{nuc} = 1.0$ kcal/(mol of residue).

before treating helix bundles. In particular, we now explore the two-helix approximation (2ha): a molecule can have a maximum of two helices anywhere in the chain.

B. Single-Helix Protein Again, Now in the Two-Helical-Segment Approximation. In 2ha, the partition sum is given by

$$q_{1c,2h}(N,T) = 1 + \sigma\sum_{i=1}^{M-4}(M-i-3)s^i + \sigma^2\sum_{j=1}^{M-9}\sum_{k=1}^{M-j-8}\frac{(M-j-k-7)(M-j-k-6)}{2}s^{j+k} \quad (11)$$

where s and σ have the same definitions as before. The first term represents the complete coil, the second term represents all conformations containing a maximum of a single helix, and the third term represents all conformations having two helical segments (j and k helical bonds each). The fractional helicity Θ and specific heat C_v are found by substituting $q_{1c,2h}$ into eqs 7 and 9.

Relative to the single-helix approximation, the two-helix approximation leads to a small change in the best fit parameters (see Figures 2 and 3). The best fit value of the hydrogen bond parameter changes by 6%, and the z parameter changes by 9%. While the predictions in Figures 4 and 5 are only slightly improved, the predicted specific heat improves more significantly using the two-helix approximation than the single-helix approximation (see Figure 6). This calculation shows the nature of the errors made by these approximations. We take the simpler single-helix approximation to be sufficient for the purpose of treating helix bundles below.

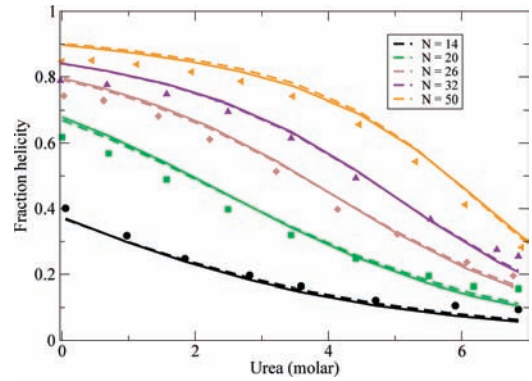

Figure 5. Urea denaturation of peptides of different lengths from Scholtz et al.[40] Symbols represent experimental data, dashed lines denote best fit lines using the model with the single-helix approximation, and solid lines represent the best fit using the two-helical-segment model. In this model, we use $z = 7.43$, $\varepsilon_{hb} = 1.2$ kcal/(mol of residue), $\varepsilon_{nuc} = 1.0$ kcal/(mol of residue), and $m = 0.029$ kcal/M.

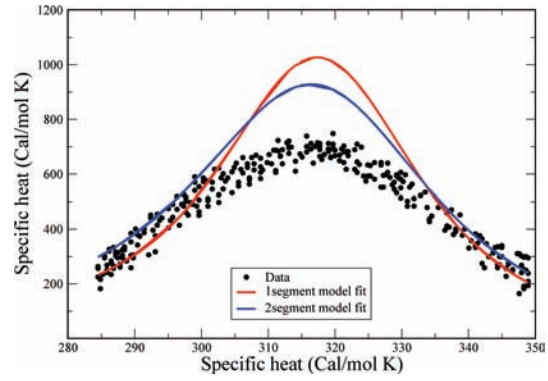

Figure 6. Heat capacity data from Scholtz et al.[41] These data were not used to obtain the best fit parameter values. The solid line is the theoretical prediction, while data points are shown in filled circles. The red curve is produced using single-helical-sequence approximation with the parameter values $z = 6.83$, $\varepsilon_{hb} = 1.14$ kcal/(mol of residue), and $\varepsilon_{nuc} = 1.08$ kcal/(mol of residue), whereas the blue curve was produced by using two-helical-sequence approximation with the parameter values $z = 7.43$, $\varepsilon_{hb} = 1.2$ kcal/(mol of residue), and $\varepsilon_{nuc} = 1.0$ kcal/(mol of residue).

III. Partition Function for Two-Helix-Bundle Proteins

In this section, we treat two-helix-bundle (2hb) molecules. Now, in addition to the local interactions within each helix, we also treat the nonlocal interactions that occur when the two helices are packed adjacent to each other. As before, we factorize the total partition function q_{2tot} for the two-helix bundle into its $(z-1)^{2(N_2+2)}$ chain conformations and its combinatoric factor q_{2c}:

$$q_{2tot}(N_2,T) = (z-1)^{2(N_2+2)}q_{2c} + q_{1tot}(N_1,T) \quad (12)$$

where q_{2c} is expressed as

$$q_{2c}(N_2,T) = [q_{1c}(N_2,T) - 1]^2 z^4 + \sigma^2\sum_{i=1}^{N_2/4}\sum_{j=1}^{N_2/4}s^{4(i+j)}(r^{\min(i,j)} - 1) \quad (13)$$

These equations are based on the following parsing of terms. q_{1c} is the partition sum for each individual helical sequence of the chain, given by eq 3. Hence, the first term in eq 13 (*two-*

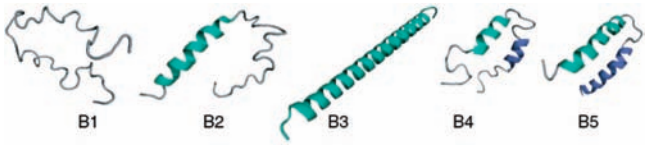

Figure 7. Different conformations considered in the partition sum for a two-helix-bundle protein. B1, B2, and B3 denote all the components of a single helix conformation, while B4 shows all possible configurations with two partially formed helices interacting with each other. B5 represents two fully formed helices, the native state of a two-helix-bundle protein.

independent-arm term) accounts for all the ways that each helical arm can have at least one helical turn, where the helices do not interact with each other. The factor of z^4 accounts for the minimum of three virtual bonds connecting the two helices and the fact that the second helix can orient in any possible direction compared to the three-bond linker. Subtracting unity from q_{1c} ensures that we count only noncoil states, i.e., all the states in which there is at least one helical bond. The second term in eq 13 (*two-helix-bundle term*) is a sum over all the states in which one helix is partially formed to any degree, the second helix is formed to any degree, and the two helices are in contact and interact with each other (see Figure 7). Each helix−helix contact has a contact energy ε_{hh} corresponding to an equilibrium constant $r = \exp(\varepsilon_{hh}/kT)$. This contact term operates between helical turns, and hence, the sum is over the number of helical turns, rather than over the number of helical bonds. We regard these as helix−helix interactions as primarily hydrophobic and packing interactions. We have also assumed that only the configuration with maximum helix−helix interactions contributes to the partition sum, and hence, we have the term $\min(i, j)$. The other configurations which are responsible for forming contacts of less than $\min(i, j)$ between two helices have been ignored because the statistical weight of those terms is negligible. Thus, the first component in eq 12 accounts for all combinations of two segments where helical turns nucleate into two helices (B4 and B5 in Figure 7). The final term (*single-helix term*), given in eq 12, accounts for the possibility that instead of a two-helix bundle, the whole chain simply forms a single long helix of any degree of helicity (including the all-coil state; see Figure 7). N_1 is the maximum number of helical bonds that can be formed if the chain has a complete single-helix conformation. Thus, $2N_2 + 10 = M$ and $N_1 = M - 4$, where for the two-helix conformations we assume a three-bond linker between the two helices. The expression of N_2 is derived from the fact that there are $M - 1$ bonds, three of which contribute to the linker and the rest of which are equally distributed in each helix, and the maximum number of helical bonds is always three less than the total number of bonds available because of i, $i + 4$ nature of the helical contacts. Thus, $N_2 = (M - 4)/2 - 3$, which gives $M = 2N_2 + 10$. As above, added denaturant will diminish the hydrophobic interactions approximately linearly:

$$\varepsilon_{hh} = \varepsilon_{hh}^{0} - mc \tag{14}$$

where ε_{hh}^{0} is the interhelical contact formation energy in the absence of denaturant. We use the same value of m as was used for the hydrogen bond strength dependence on the denaturant concentration.

IV. Partition Function for Three-Helix-Bundle Proteins

Finally, we treat three-helix bundles in a similar way. Following previous notation, we factorize the total partition function into its $(z - 1)^{3(N_3 + 2)}$ chain conformations and its combinatoric factor q_{3c}.

$$q_{3tot}(N_3,T) = (z-1)^{3(N_3+2)} q_{3c} + q_{2tot}(N_2,T) \tag{15}$$

q_{3c} is expressed as

$$q_{3c}(N_3,T) = [q_{1c}(N_3,T) - 1]^3 z^8 + \sigma^3 \sum_{i=1}^{N_3/4} \sum_{j=1}^{N_3/4} \sum_{k=1}^{N_3/4} s^{4(i+j+k)} (r^{3\min(i,j,k)} - 1) \tag{16}$$

The logic behind different terms follows much the same way as before, $q_{1c}(N_3, T)$ is the partition function for a single helix having a total of N_3 bonds, given by eq 3. Hence, the first term in eq 16 (*three-independent-arm term*) accounts for all conformations having three fully or partially formed helices throughout the chain where the helices do not interact with each other. The factor of z^8 accounts for the minimum of three virtual bonds connecting the helices, and the second and third helices can orient in any possible direction with respect to the three-bond linker. Subtracting unity from q_{1c} ensures that we count only noncoil states, i.e., all the states in which there is at least one helical bond. The second term in eq 16 (*three-helix-bundle term*) is a sum over all the states in which one helix is partially formed to any degree, the second and third helices are formed to any degree, and the three helices are in contact and interact with each other. Each helix−helix contact has a contact energy ε_{hh} (hydrophobic and packing interaction), corresponding to an equilibrium constant $r = \exp(\varepsilon_{hh}/kT)$ as before. This contact term operates between helical turns, and hence, the sum is over the number of helical turns, rather than over the number of helical bonds. It is important to note that the interaction is still two-body, and hence, we get a factor of 3 in the exponent of r. Once again, we assume the most dominant contribution to the partition sum arises due to three helices (with i, j, and k turns) having a $\min(i, j, k)$ number of contacts, and the configurations with fewer contacts have been ignored. Thus, the first component in eq 15 accounts for all combinations of three segments where helical turns nucleate into three helices (see Figure 8). The final term (*two-helix term*) in eq 15 includes all possible conformations that the protein molecule could adopt if it were in a two-helix-bundle conformation (see C1, C2, C3, C4, and C5 in Figure 8). This has already been derived explicitly in eq 12 and accounts for the single-helix term as well as a complete coil state. There are a maximum of $3N_3$ helical bonds, with each helix having N_3 bonds for a three-helix configuration. N_3 is related to the total number of amino acids M by $M = 3N_3 + 16$ using the same argument used for a two-helix-bundle protein and assuming each helix has a three-bond linker spacing as before. N_2 is the number of bonds in each helix for a two-helix-bundle conformation and is related to the total number of amino acids M as before, $M = 2N_2 + 10$.

A. Results. Here we compare the predictions from the model with experiments on two three-helix-bundle proteins. Because of the extensive experimental data available, we study the thermally induced and denaturant-induced unfolding of protein A[42] (see Figures 9 and 10), with a single set of parameters, and the thermal denaturation for three different values of the denaturant concentration of the protein α3C[43] (see Figure 11). For these cases, although the agreement is not perfect, it is quite good. A deeper test of the theory would be if our model helix−coil parameters were known for the individual helices from independent experiments, but they are not in these cases, as far as we know. Our parameter values are given in Table 1.

(42) Dimitriadis, G.; Drysdale, A.; Myers, J. K.; Arora, P.; Radford, S. E.; Oas, T. G.; Smith, D. A. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 3809–3814.

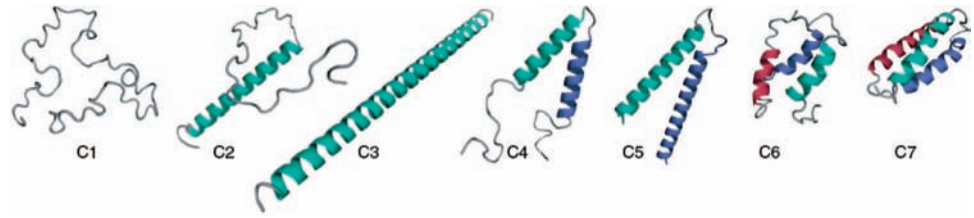

Figure 8. Different conformations considered in the partition sum for a three-helix-bundle protein. C1, C2, C3, C4, and C5 show all the components for a full two-helix-bundle partition sum. C6 denotes structures where the helices in three helical arms are partially formed, and C7 is the three-helix-bundle native structure with completely formed helices.

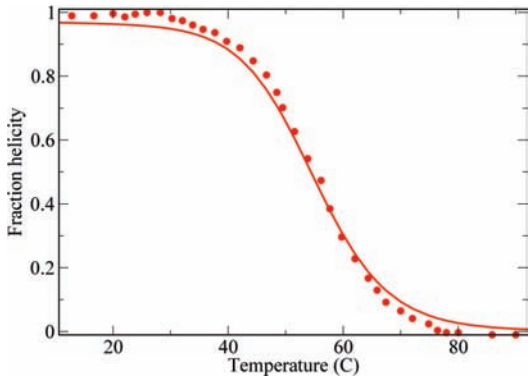

Figure 9. Protein A thermal denaturation data.[42] The values of the parameters used are $z = 3.73$, $\varepsilon_{hb} = 0.73$ kcal/(mol of residue), $\varepsilon_{hh} = 2.38$ kcal/(mol of residue), $m = 0.04$ kcal/M, and $\varepsilon_{nuc} = 6.33$ kcal/mol.

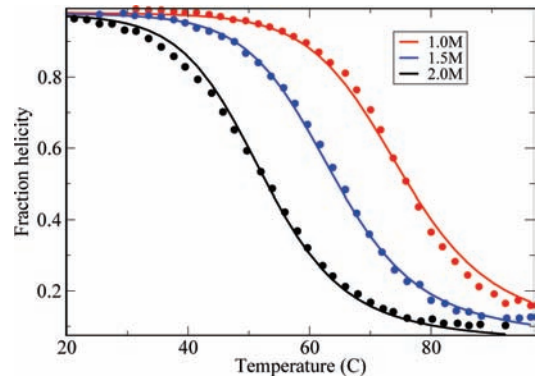

Figure 11. α3C thermal denaturation data.[43] The values of the parameters used are $z = 3.43$, $\varepsilon_{hb} = 0.64$ kcal/(mol of residue), $\varepsilon_{hh} = 1.31$ kcal/(mol of residue), $m = 0.04$ kcal/M, and $\varepsilon_{nuc} = 3.6$ kcal/mol.

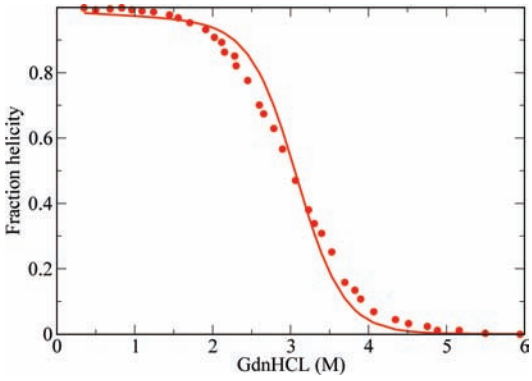

Figure 10. Protein A GdnHcl denaturation.[42] The values of the parameters used are $z = 3.73$, $\varepsilon_{hb} = 0.73$ kcal/(mol of residue), $\varepsilon_{hh} = 2.38$ kcal/(mol of residue), $m = 0.04$ kcal/M, and $\varepsilon_{nuc} = 6.33$ kcal/mol.

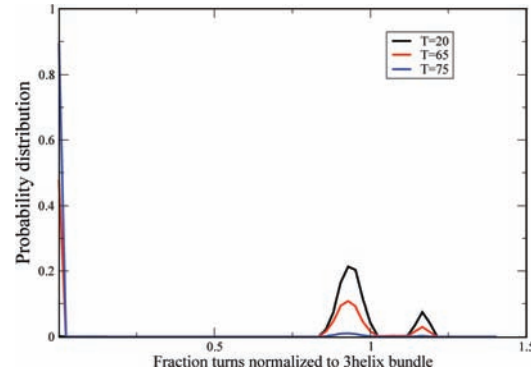

Figure 12. Protein A folds with a two-state transition. The blue curve shows the population of different states at a high-temperature ($T = 75$ °C) denaturing condition with a major peak near the coil-like conformation regime. The distribution of states at low temperature ($T = 20$ °C) is shown in black, which shows a peak near the three-helix-bundle native state with a very small peak near the two-helix-bundle conformation as well. The red curve, predicting the distribution of states at intermediate temperature near melting ($T = 65$ °C), shows an approximate two-state behavior with two peaks near (i) the coil-like state and (ii) the three-helix-bundle native state and with a slight population near the two-helix-bundle state. These curves were generated using the parameter values $z = 3.73$, $\varepsilon_{hb} = 0.73$ kcal/(mol of residue), $\varepsilon_{hh} = 2.38$ kcal/(mol of residue), $m = 0.04$ kcal/M, and $\varepsilon_{nuc} = 6.33$ kcal/mol.

V. Nature of the Cooperativity in Helix-Bundle Folding

Kaya and Chan and their colleagues[21] have argued that previous models of protein folding underpredict the high cooperativities that are observed in experiments. Small single-domain globular protein molecules tend to fold in a two-state manner; i.e., at the transition midpoint, there is a negligible population of intermediate states. Experiments reflect this either when specific measurements are made of the individual chain populations or through the observation that the ratio of the van't Hoff enthalpy (H_{vH}) to the calorimetric enthalpy (H_{cal}) is experimentally found to be very close to 1. Theoretical models, in contrast, tend to predict a value of this ratio that is smaller than 1. As noted by Kaya and Chan,[21] even Go models, which are nonphysical models that are designed to be highly cooperative, give ratios of this quantity that are too small.

To explore the cooperativity predicted by our model for three-helix bundles, we computed conformational populations at the transition midpoint, and we computed the ratio of enthalpies. We find that this model predicts two-state cooperativity. For proteins A and α3C, we evaluate the density of states at three different temperatures. At high temperatures, the model predicts a peaked unimodal distribution around the denatured states. At low temperatures, it predicts a peaked unimodal distribution around the native state; however, there is a very small peak at the two-helix conformation as well predicted by our mean-field

(43) Bryson, J. W.; Desjarlais, J. R.; Handel, T. M.; Degrado, W. F. *Protein Sci.* **1998**, *7*, 1404–1414.

Table 1. Values of Fitted Parameters

protein name	z	ε_{hb} [kcal/(mol of residue)]	ε_{hh} [kcal/(mol of residue)]	ε_{nuc} (kcal/mol)	$\sigma(T = 298\ K)$	m (kcal/M)
Ala-Glu-Ala-Ala-Lys (one helix)	6.83	1.14		1.08	0.17	0.028
protein A (three helices)	3.73	0.73	2.38	6.33	0.00003	0.04
α3C (three helices)	3.43	0.64	1.31	3.6	0.002	0.04

model. Also, at the midpoint temperature, the model predicts a predominantly bimodal distribution but with a small population of the two-helix-bundle intermediate conformation. Our model predicts a high cooperativity for protein A and moderately high cooperativity for α3C (see Figures 12 and 13).

In addition, we tested the prediction for the calorimetric behavior. We define ΔH_{cal} as the enthalpy difference between the fully unfolded state and the native state. The van't Hoff enthalpy H_{vH} is defined as[21,44]

$$\Delta H_{vH} = k \frac{\partial \log K(T)}{\partial(1/T)} \tag{17}$$

where we take $K(T)$ to be

$$K(T) = \frac{\langle i \rangle - i_u}{i_{ns} - \langle i \rangle} \tag{18}$$

and where $\langle i \rangle$, which serves as an order parameter, is the average number of helical turns in the protein at temperature T, i_u is the average number of helical turns in the unfolded state, and i_{ns} is the number of helical turns in the native state. We define δ as the ratio of these two enthalpies:

$$\delta = \frac{\Delta H_{vH}}{\Delta H_{cal}} \tag{19}$$

For any system, it must be true that $\delta \leq 1$. $\delta = 1$ only for a two-state transition.[44] Thus, the quantity δ is a measure of how well a model captures the two-state cooperativity observed in proteins. In Table 2, we show predictions of the model for these

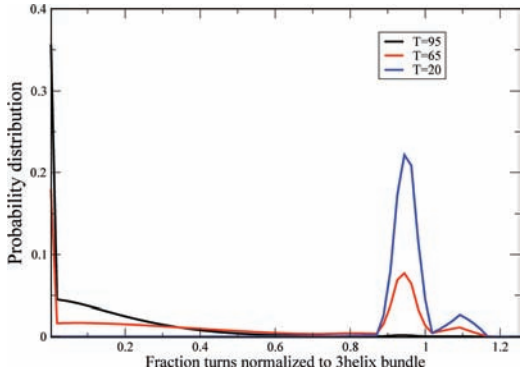

Figure 13. Protein α3C folds with a two-state transition. The black curve shows the population of different states at a high-temperature ($T = 95$ °C) denaturing condition with a major peak near the coil-like conformation regime. The distribution of states at low temperature ($T = 20$ °C) is shown in blue, which shows a peak near the three-helix-bundle native state with a very small peak near the two-helix-bundle conformation as well. The red curve, predicting the distribution of states at an intermediate temperature near melting ($T = 65$ °C), shows an approximate two-state behavior with two peaks near (i) the coil-like state and (ii) the three-helix-bundle native state with a slight population at the two-helix-bundle state as well. These curves were generated using the parameter values $z = 3.43$, $\varepsilon_{hb} = 0.64$ kcal/(mol of residue), $\varepsilon_{hh} = 1.31$ kcal/(mol of residue), $m = 0.04$ kcal/M, and $\varepsilon_{nuc} = 3.6$ kcal/mol.

Table 2. Cooperativities of Different Proteins and Their Component Helices

protein name	δ(one helix at T_m)	δ(three helices at T_m)	T_m (K)
protein A (2.2 M)	0.36	0.91	327.3
α3C (2.0 M)	0.33	0.72	326

two three-helix-bundle proteins. In the same table, we also show the cooperativities of the individual helices (if the full protein molecule assumes a completely long helix) taken alone computed at the experimentally observed melting temperature, predicted from the model. Our model predicts that protein A should have near two-state cooperativity, but that α3C is less cooperative. We are not aware of calorimetric data for these particular proteins. We also find that the individual component helices of these proteins are not by themselves sufficiently cooperative to account for the folding cooperativity of the full protein. Hence, this model indicates that the process of both forming helices and the association and packing of multiple helices together is more cooperative than the helix formation process alone. This indicates how protein folding may be so highly cooperative.

VI. Conclusions

We study a simple analytical model for a single-polymer chain transition from a large denatured ensemble to either a single helix or compact helix-bundle conformations. First, at the single-helix level, this model is among the simplest possible versions of classical helix−coil theory, and it performs well on a substantial body of data on thermally induced and solvent-induced denaturation vs chain length. We find that the single-helix approximation is adequate for short polymers, except when predicting the heat capacity, in which case the two-helix approximation is needed. Because of the simplicity of this model, we can then also treat helix-bundle proteins analytically. Even though we use mean-field treatment, this helix-bundle model gives a good accounting of the experimental data on denaturation and cooperativity. The model predicts two-state cooperativity for helix-bundle proteins. The two-state behavior arises because the cooperativity in the helix−coil process is enhanced by the process of the tertiary packing of helices upon each other. This model for cooperativity may generalize to the folding of other small globular proteins and provides a simple tractable model of protein stability.

Acknowledgment. We thank Hue Sun Chan, Adam Lucas, Bruce Kerwin, and David Brems for helpful discussions. This research was supported by a University of California discovery grant from Amgen and NIH Grant GM 34993.

JA808136X

(44) Lumry, R.; Biltonen, R.; Brandts, J. F. *Biopolymers* **1966**, *4*, 917–944.