J|A|C|S
ARTICLES
Published on Web 03/16/2007

Statistical Mechanics of Helix Bundles Using a Dynamic Programming Approach

Adam Lucas,*,† Liang Huang,‡ Aravind Joshi,‡ and Ken A. Dill§

Contribution from the Department of Mathematics, Saint Mary's College of California, Moraga, California 94575-3517, Department of Computer & Information Science, University of Pennsylvania, Philadelphia, Pennsylvania 19104, and Department of Pharmaceutical Chemistry, University of California at San Francisco, San Francisco, California 94143-2240

Received October 11, 2006; E-mail: arl3@stmarys-ca.edu

Abstract: Despite much study, biomolecule folding cooperativity is not well understood. There are quantitative models for helix-coil transitions and for coil-to-globule transitions, but no accurate models yet treat both chain collapse and secondary structure formation together. We develop here a dynamic programming approach to statistical mechanical partition functions of foldamer chain molecules. We call it the ascending levels model. We apply it to helix-coil and helix-bundle folding and cooperativity. For 14- to 50-mer Baldwin peptides, the model gives good predictions for the heat capacity and helicity versus temperature and urea. The model also gives good fits for the denaturation of Oas's three-helix bundle B domain of protein A (F13W*) and synthetic protein α_3C by temperature and guanidine. The model predicts the conformational distributions. It shows that these proteins fold with transitions that are two-state, although the transitions in the Baldwin helices are nearly higher order. The model shows that the recently developed three-helix bundle polypeptoids of Lee et al. fold *anti-cooperatively*, with a predicted value of $\Delta H_{vH}/\Delta H_{cal} = 0.72$. The model also predicts that two-helix bundles are unstable in proteins but stable in peptoids. Our dynamic programming approach provides a general way to explore cooperativity in complex foldable polymers.

1. Modeling Cooperativity in Helix-Bundle Molecules

Typical proteins fold cooperatively. They undergo sharp equilibrium transitions from denatured to native states as a function of temperature or denaturing solvents. What is the physical basis for the cooperativity? Various factors have been explored. First, protein cooperativity has been explored through polymer helix-coil transition experiments and theory, beginning in the late 1950s.[1−7] Helix formation is driven by *local interactions* (i.e., by the helical propensities among near neighbors in the chain sequence). The Zimm−Bragg theory, for example, predicts the conformational populations based on two parameters: σ, the statistical weight for nucleating a helical region of the chain, and s, the statistical weight (or equilibrium constant) for propagating one unit of helix. However, by itself, helix-coil theory is not sufficient to explain protein folding cooperativity because many proteins that fold cooperatively have no helices at all.

Second, protein folding also involves chain collapse, and therefore other models treat folding as a solvent-induced polymer collapse process.[8−17] Collapse models correctly predict that protein stabilities should go through a maximum with temperature[18] reflecting their basis in the hydrophobic interactions between the nonpolar amino acids and water. Such models also predict denaturation by solvents[19] and pH and salts.[20] However, collapse theories, too, are limited. They do not treat specific protein architectures, such as the helices or sheets.

Recently, Chan and colleagues have shown that even the most cooperative models, based on nonphysical potentials, such as Gō models, do not predict cooperativities as high as those that are observed in protein folding experiments.[16,17,21] It remains a major challenge to understand the physical origins of cooperativity in protein folding, and it remains a major challenge in polymer statistical mechanics to treat both local and nonlocal (solvation-driven) interactions within a single framework.

† Saint Mary's College of California.
‡ University of Pennsylvania.
§ University of California at San Francisco.

(1) Poland, D.; Scheraga, H. *Theory of Helix-Coil Transitions in Biopolymers*; Academic Press: New York, 1970.
(2) Zimm, B.; Bragg, J. *J. Chem. Phys.* **1958**, *28*, 1246−1247.
(3) Schellmann, J. *J. Phys. Chem.* **1958**, *62*, 1485−1494.
(4) Scholtz, J.; Baldwin, R. *Annu. Rev. Biophys. Biomol. Struct.* **1992**, *21*, 95−118.
(5) Munoz, V.; Serrano, L. *Nat. Struct. Biol.* **1994**, *1*, 399−409.
(6) Munoz, V.; Serrano, L. *Curr. Opin. Biotechnol.* **1995**, *6*, 382−386.
(7) Applequist, J. *J. Chem. Phys.* **1963**, *38*, 934−941.
(8) Dill, K. *Biochemistry* **1985**, *24*, 1501−1509.
(9) Dill, K.; Stigter, D. *Adv. Protein Chem.* **1995**, *46*, 59−104.
(10) Privalov, P.; Khechinashivili, N. *J. Mol. Biol.* **1974**, *86*, 665−684.
(11) Finkelstein, A.; Shakhnovich, E. *Biopolymers* **1989**, *28*, 1667−1680.
(12) Sali, A.; Shakhnovich, E.; Karplus, M. *Nature* **1994**, *369*, 248−251.
(13) Wolynes, P.; Onuchic, J.; Thirumalai, D. *Science* **1995**, *267*, 1619−1620.
(14) Klimov, D.; Thirumalai, D. *Folding Des.* **1998**, *3*, 127−139.
(15) Knott, M.; Chan, H. *Chem. Phys.* **2004**, *307*, 187−199.
(16) Moghaddam, M. S.; Shimizu, S.; Chan, H. *J. Am. Chem. Soc.* **2005**, *127*, 303−316.
(17) Chan, H. *Proteins: Struct., Funct., Genet.* **2000**, *40*, 543−571.
(18) Dill, K.; Alonso, D.; Hutchinson, K. *Biochemistry* **1989**, *30*, 5974−5985.
(19) Alonso, D.; Dill, K. *Biochemistry* **1991**, *30*, 5974−5985.
(20) Stigter, D.; Alonso, D.; Dill, K. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 4176−4180.
(21) Chan, H.; Shimizu, S.; Kaya, H. *Methods Enzymol.* **2004**, *380*, 350−379.

10.1021/ja067153s CCC: $37.00

One of the problems is experimental: how can we systematically study cooperativity? Helix-coil and coil-globule collapse processes can be studied in homopolymers, controlled by just a few variables, the chain length and simple interaction parameters. But proteins are complex; there are no simple "knobs" that vary the cooperativity. Each protein has a different fold, a different packing, a different amount of secondary structure of different types, a different amount of hydrogen bonding, and different hydrophobic cores. What systematic experiments can explore cooperativity and stability in protein folding?

We believe a new opportunity is afforded by helix-bundle proteins. First, they provide the proper challenge: helix-bundle folding involves both collapse and helix formation. Second, they can be studied systematically through simple independent variables such as the lengths and numbers of helices, in addition to simple solvation and helical-propensity parameters, for simple sequences. Hecht et al.[22] and DeGrado et al.[23] and others have shown that synthetic helix bundles can be designed simply by putting hydrophobic residues on the inside, polar monomers on the outside, and connecting them with simple turns and loops. Third, there is a growing body of helpful experimental data on helix-bundle stabilities, coming from two directions. First, to understand the speed limits in folding kinetics, there are new detailed denaturation studies of "ultrafast folders",[24−30] which happen to be predominantly small helix-bundle proteins. Second, folding has been studied in nonbiological helix bundles, made from simple-sequence *polypeptoids*, which are N-substituted glycines.[31−34] The numbers and lengths of helices have been varied systematically in these peptoid helix bundles, and their denaturation has been studied using temperature and denaturants.[35] Hence, chain molecule helix bundles are providing a new arena for understanding how local and nonlocal interactions can conspire within a single chain molecule to cause it to fold and collapse cooperatively.

Here, we develop a theory that treats both helix formation and the collapse process within a single framework. We account for the hydrogen bond formation within helices and the hydrophobic interactions within helices and among helices. We devise a dynamic programming (DP) method that computes the density of states of all the conformations and thus computes the partition functions of such molecules. We use this model to explore folding cooperativities both in simple single helices and in helix-bundle proteins.

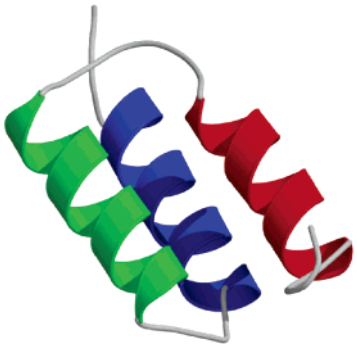

Figure 1. Three-helix bundle protein: the B domain of protein A.

For the single helix-coil process, our approach is more physical than classical helix-coil models such as those of Zimm and Bragg[2] or Lifson and Roig[36] because our approach treats the full polymer physics by actually counting chain conformations in space, including excluded volume approximately, and not just counting one-dimensional distributions of "h" and "c" symbols. We find that this allows us to dispense with the parameter σ previously required in helix-coil theories.

Second, however, we find that an additional interhelical cooperativity term is required to account for the high cooperativities in multiple-helix bundles, perhaps consistent with the findings of Chan and his group that simple two-body-based Gō models are not sufficient to capture folding cooperativity.[16,17,21] Using these three quantities (a hydrogen bond free energy, a hydrophobic interaction free energy, and a cooperative interaction energy), our model gives good agreement with a body of experimental data on helix-coil and helix-bundle stabilities.

2. The Ascending Levels Model

Our chains are configured on a cubic lattice. We consider one-, two-, or three-helix bundles (by one-helix bundle, we mean just a single helix). In this lattice model, there are slices or layers of the lattice that are perpendicular to the axes of the aligned helices in the native structure. We represent a helix bundle as a linear stacking of levels (i.e., there is a first level, a second level, etc.; Figure 2). For example, a three-helix bundle involves three *columns*, within which varying amounts of helix or coil are configured. Within that column, in a coil region, the chain vector starts at one node on the cubic lattice and ends on any neighboring node. In a helix segment, which entails three bonds, the beginning and ending points are nodes on the cubic lattice, but the intermediate beads are located on lattice edges, not on nodes (Figure 2). For this reason, we refer to a helical turn as being *off-lattice*. We believe that this model, despite its simplifications, captures the essentials of helix-bundle physics. A long tradition in polymer statistical mechanics has shown that geometric accuracy of chain detail is less important for capturing the density of states than an ability to capture the proper relative sizes of regions within the phase space. We believe the present model does this.

Figure 3 shows how a protein conformation is constructed as a combination of (1) individual columns, inside of which each three-residue piece of chain is labeled as either c or h, (2)

(22) Kamtekar, S.; Schiffer, J.; Xiong, H.; Babik, J.; Hecht, M. *Science* **1993**, *262*, 1680−1685.
(23) Bryson, J.; Betz, S.; Lu, H.; Suich, D.; Zhou, H.; Oneil, K.; DeGrado, W. *Science* **1995**, *270*, 935−941.
(24) Kubelka, J.; Eaton, W.; Hofrichter, J. *J. Mol. Biol.* **2003**, *329*, 625−630.
(25) Spector, S.; Raleigh, D. *J. Mol. Biol.* **1999**, *293*, 763−768.
(26) Fan, K.; Wang, J.; Wang, W. *Phys. Rev. E: Stat. Phys., Plasmas, Fluids, Relat. Interdiscip. Top.* **2001**, *64*, 4−7.
(27) Mayor, U.; Johnson, C.; Dagett, V.; Fersht, A. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 13518−13522.
(28) Yang, W.; Gruebele, M. *Nature* **2003**, *423*, 193−197.
(29) Dimitriadis, G.; Drysdale, A.; Myers, J.; Arora, P.; Radford, S.; Oas, T.; Smith, A. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 3809−3814.
(30) Zhu, Y.; Alonso, D.; Maki, K.; Huang, C.; Lahr, S.; Daggett, V.; Roder, H.; DeGrado, W.; Gai, F. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 15486−15491.
(31) Kirshenbaum, K.; Zuckermann, R.; Dill, K. *Curr. Opin. Struct. Biol.* **1999**, *9*, 530−535.
(32) Kirshenbaum, K.; Barron, A.; Goldsmith, R.; Armand, P.; Bradley, E.; Truong, K.; Dill, K.; Cohen, F.; Zuckermann, R. *Folding Des.* **1997**, *2*, 369−375.
(33) Kirshenbaum, K.; Barron, A. E.; Goldsmith, R. A.; Armand, P.; Bradley, E. K.; Truong, K. T. V.; Dill, K. A.; Cohen, F. E.; Zuchermann, R. N. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 4303−4308.
(34) Burkoth, T.; Beausoleil, E.; Kaur, S.; Tang, D.; Cohen, F.; Zuckermann, R. *Chem. Biol.* **2002**, *9*, 647−654.
(35) Lee, B.-C.; Zuckermann, R. N.; Dill, K. A. *J. Am. Chem. Soc.* **2005**, *127*, 10999−11009.
(36) Lifson, S.; Roig, A. *J. Chem. Phys.* **1961**, *34*, 1963−1974.

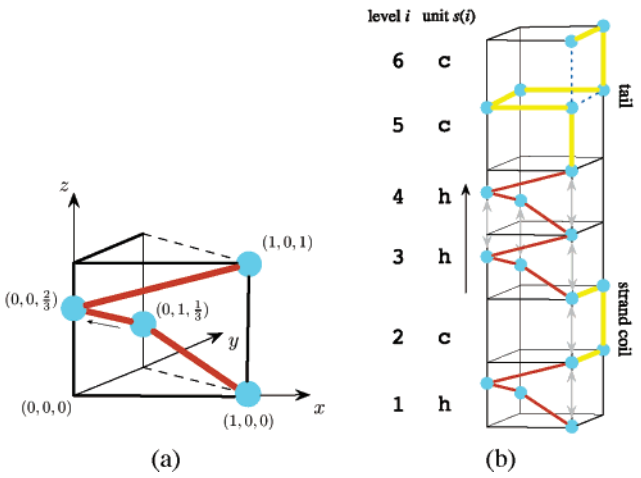

Figure 2. Ascending levels model: (a) an example of a helical unit on strand 1 in a three-dimensional lattice. (b) A six-level (19-residue) single helix. The strand contains a length one *strand coil* and a length two *tail*. Double arrows in gray denote both hydrogen bonds and hydrophobic interactions, while dashed lines in deep blue denote hydrophobic interactions only.

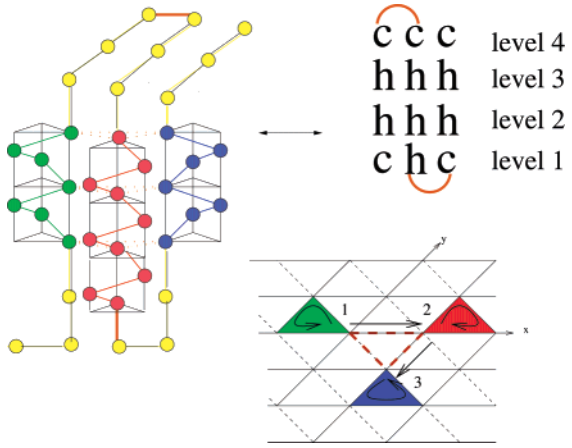

Figure 3. Ascending levels model for a three-helix bundle: the helices in the bundle are represented in terms of levels or layers on the lattice. Here, a four-level three-helix bundle is shown on a cubic lattice together with its associated h/c sequence. Also shown is a slice through the XY plane, indicating how the three strands are situated in a given layer on the lattice. The dotted brown line indicates interhelical interactions.

loops connecting the columns, and (3) *tails* that are connected to a column at one end and free at the other end. Here are the construction rules we use. We call *n* consecutive coil units at the beginning or the end of the helix bundle a *length-n tail* (Figure 2). We call *n* consecutive coil units sandwiched between two helical units on the same strand a *length-n strand coil*.

1. Each helix unit or coil *unit* consists of three monomers. The chain trace of a helix unit follows the *off-lattice* path described above, whereas the chain trace of a coil unit follows the edges of a cubic lattice.

2. Each of the one, two, or three strands of the chain has an identical number of 3-mer units.

3. Two helix units at the same level on different strands make an interhelical (i.e., hydrophobic) contact. In the case of the first and third strands, an interhelical contact is along the diagonal of the lattice.

4. Looking at a three-helix bundle from above, the helices of each strand are stacked above one another.

5. Two helices on a strand separated by *n* coil units are distance *n* apart.

6. Aside from the fully open chain, each strand of a helix bundle must contain at least one helix unit.

7. We allow at most c_m consecutive coil units in a helix bundle.

8. Excluded volume constraint for three-helix bundles: To be physically viable, length-1 strand coils on the middle strand must have three consecutive coil units beside it on an adjacent strand.

2.1. The Interaction Energies in the Model. We consider three types of interactions: hydrogen bonds, hydrophobic interactions, and an interhelical cooperative interaction when one helix interacts with another helix. A hydrogen bond in the model occurs when consecutive helix units have an intrahelical contact of unit distance along the *z* axis, at every third residue. A hydrophobic interaction occurs when two monomers that are not neighbors in the chain sequence are spatial neighbors. Every hydrogen bond is regarded as having the same energy as every other, and every hydrophobic contact is similarly identical to every other hydrophobic contact.

We compute the energy, *E*, of a chain configuration as

$$E = N_{hb}\cdot\epsilon_{hb} + N_{\phi}\cdot\epsilon_{\phi} + N_{c}\cdot\epsilon_{c} \qquad (1)$$

where ϵ_{hb}, ϵ_{ϕ}, and ϵ_{c} are the energies (in units of kilocalories per mole), respectively, for a hydrogen bond, a hydrophobic interaction, and a cooperative interaction. The count of each such type of interaction is N_{hb}, N_{ϕ}, and N_{c}, respectively. Our helix bundle is a homopolymer, and hence a hydrophobic interaction results whenever two nonconsecutive monomers are on adjacent lattice sites or form an interhelical contact. Hydrophobic contacts can arise as interactions between coil units and the rest of the chain. Intrahelical contacts are taken to involve both a hydrogen bond and a hydrophobic interaction. A cooperative interaction arises in a two-helix bundle or a three-helix bundle if there is an interhelical contact between two strands at the same level. In a sense, cooperative interaction entails more than just the individual helical monomers; it also requires that the monomers are in helical arrangements with neighbors. Hence, our cooperativity term could be arguably called a multibody term.

For treating thermal denaturation, we take the interaction energies, ϵ, to be constant, independent of temperature. In addition, we are interested in effects of chemical denaturants. For chemical denaturants at low concentrations, we assume that the hydrogen bond and hydrophobic interaction energies depend linearly on denaturant, and hence we take $\epsilon_{hb} = a - b\cdot[D]$ and $\epsilon_{\phi} = c - d\cdot[D]$ where *a* is the hydrogen bond energy at zero denaturant concentration, *c* is the hydrophobic interaction at zero denaturant, [D] is the denaturant concentration, and *b* and *d* are the slopes of the denaturant dependence for the formation of one hydrogen bond or one hydrophobic contact in solution.

Our dynamic programming method, discussed below, gives the density of states, the count of all the different conformations of the chain. From the density of states, $g(E)$, we get the Boltzmann weighted population for each energy *E*:

$$p(E) = g(E)\cdot\exp(-\beta E)$$

where $\beta = 1/kT$ is the reciprocal of the Boltzmann constant

times temperature. The sum of all such populations is the partition function:

$$Q = \sum_E p(E)$$

From the partition function, we can derive experimentally accessible quantities such as heat capacity, c_P, and fraction native, f_N:

$$c_P = \frac{d}{dT}\left[\frac{1}{Q}\sum_E Ep(E)\right]$$

and

$$f_N = \frac{1}{Q}\sum_E \frac{\nu}{\nu_{max}} p(E)$$

The only hydrogen bonds and interhelical contacts that we consider are those that are native. We compute the degree of nativeness of a given chain conformation as the number of hydrogen bond and two-helix interactions, $\nu = N_{hb} + N_c$, divided by the number of these interactions in the native state, $\nu_{max} = (N_{hb} + N_c)_{max}$. Near-native structures have $\nu/\nu_{max} \approx 1$. In the case where the native state is a single helix, the fraction of native molecules reduces to fractional helicity since ν/ν_{max}, which is the number of hydrogen bonds over the total possible number of hydrogen bonds in the native state, is a measure of helicity.

3. The Dynamic Programming Approach to Partition Functions

Here, we give a brief discussion of our dynamic programming approach[37,38] to computing chain partition functions; full details are given in the Supporting Information. This approach is general and can be applied, in principle, to different types of chain folds and to different kinds of models, on-lattice or off-lattice. The approach works best for chain folds with a linear ordering (i.e., a top and bottom) and with only certain kinds of nonlocal interactions. In this article, we focus on helix-bundle folds, using a cubic lattice model of the chain.

The basic idea of dynamic programming is to divide a big problem into smaller subproblems and to share overlapping subproblems in one stage of the calculation to the next, to reduce the search. We illustrate on a single helix. The density of states function (DSF) of all possible sequences of h and c can be computed simultaneously, eliminating redundant computations. Our application of dynamic programming is similar to a proof by induction on the level, i, of the chain. We start with a base DSF, having $i = 1$, by computing the density of states for a chain of k consecutive c, $k = 0, 1, 2, ...$, forming the bottom tail of the helix. These are precomputed by exact enumeration and stored as a matrix array. Next, we use induction for the DSF for level i from the DSF for level $i - 1$. Figure 4 shows this induction going from $i = 3$ to $i = 4$. Suppose we know the DSF of all sequences up to three levels with an h on the third level. Having an h at level 3 allows us to treat conformations above and below level 3 independently. The level-4 DSF is derived from our level-3 DSF by taking the convolution (defined in the Supporting Information) with the DSF of a precomputed length-2 tail. This gives us the density of states for all conformations indicated in Figure 4.

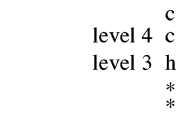

Figure 4. If level 3 is an h, then we convolute (defined in the Supporting Information) the density of states of all sequences having an h in level 3 with the DSF of a precomputed length two tail. We assume chain conformations above and below level 3 do not interact. The symbol * is a wild card, representing either h or c. The sequences shown represent states (4, 2, 0) discussed in the Supporting Information.

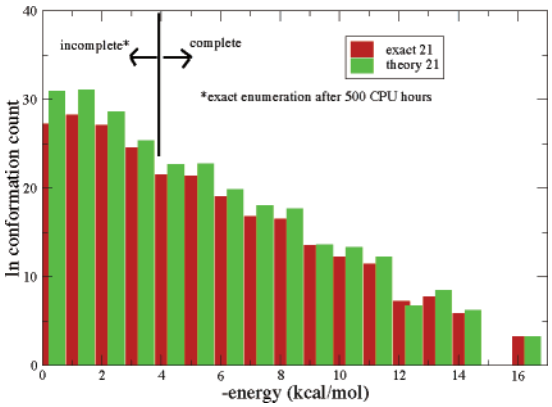

Figure 5. Density of states comparison: dynamic programming model vs exact enumerations for 21-monomer single-helix chains. Exact enumeration is for a self-avoiding walk in a cubic lattice with restrictions 1−3 of section 2. Exact enumerations were incomplete (because they exceeded our computer resources) for energies greater than −4 kcal/mol.

4. Results

4.1. Validating the Model against Exact Enumerations. Even with our simple model, the size of the search space grows exponentially with the chain length. For one-helix and three-helix bundles with n levels, there are approximately 2^n and 8^n possible conformations, respectively. Hence, enumerating a 21-mer single-helix ensemble takes more than 500 cpu hours. The advantage of our dynamic programming method is that computational complexity grows only as $O(nc_m)$ for a single helix, as $O(n^2c_m)$ for a two-helix bundle, and as $O(n^3c_m)$ for a three-helix bundle. As a result, even the computation of a 93-mer, 10-level, three-helix bundle takes only minutes to compute with the dynamic programming approach.

Our dynamic programming method provides only an approximation of the partition function since DP treats local parts of the helix bundle independently. How do we test that the model partition function is accurate? The deepest test is to compute the density of states from the model and to compare it to the *exact* density of states, determined by full enumeration by computer. When performing full enumeration, we remove constraints 2−8 from section 2 since these constraints are artificial in the sense that they were imposed to enable DP. Of course, full enumeration is only viable for chains that are sufficiently short. Below, we show that the density of states is highly accurate, both for a single helix and for a three-helix bundle, when tested against such exact short-chain enumerations.

Figure 5 compares the dynamic programming model result for a single helix compared to exact enumerations. In the Supporting Information, Figure 3 shows the corresponding test

(37) Bellman, R. *Dynamic Programming*; Princeton University Press: Princeton, NJ, 1957.
(38) Cormen, T. H.; Leiserson, C. E.; Rivest, R. L.; Stein, C. *Introduction to Algorithms*, 2nd ed.; MIT Press: Cambridge, MA, 2001.

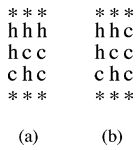

Figure 6. (a) Example of a sequence whose density of states is zero for all energy levels. The h c h strand coil in the middle strand is too constrained by the presence of a helix in either strand 1 or 3. (b) A related physically viable sequence in which the adjacent levels on strand 3 are all c.

for the three-helix bundle. These figures show that the conformation counts are accurate.

The reason for the computational advantage of our method is that the dynamic programming method accurately captures local excluded volume but neglects global excluded volume effects. Our tests here indicate that this approximation works well. The most serious challenge we found was that three-helix bundles with short strand coils in the second strand flanked by helices in the first and third strands are overconstrained and impossible in our lattice model. To eliminate these sequences, we required that strand coils of length one have three consecutive coils at the same level on either strand one or strand three. An example of a nonviable and a viable sequence are given in Figure 6. This is constraint 8 in section 2.

4.2. The Helix-Coil Transition: Comparison with Experiments. As a first test, any model of helix bundles must be able to correctly describe one-helix bundles (i.e., the simple helix-coil transition). The helix-coil transition has been widely studied in the *Baldwin peptide*, Ac−Y(AEAAKA)$_8$F−NH$_2$.[39] Figure 7 compares our model predictions to the experimental heat capacity and fractional helicity of those peptides. We find a good fit if the hydrogen bond energy is $\epsilon_{hb} = 0.786$ kcal/mol and the hydrophobic interaction energy is $\epsilon_\phi = 0.218$ kcal/mol.

We also explored chemical denaturation in the Baldwin helix. Baldwin et al. studied single helices of lengths $n = 14$, 20, 26, 32, and 50 by urea denaturation (Figure 8).[40] (We convert CD data to fractional helicity using

$$[\theta]_H = H_0 \cdot \left(1 - \frac{x}{n}\right) + H_U \cdot [\text{Urea}]$$

and

$$[\theta]_C = C_0 + C_U \cdot [\text{Urea}]$$

where $0 \le x \le 3$ is a constant used to correct for nonhydrogen bonded carbonyls that do not contribute to $[\theta]_H$.[41] We found the best fit using the value $x = 1/5$. Baseline helix and coil values H_0, H_U, C_0, and C_U were obtained from the best fit values for the exchange model in Table 1 of Baldwin et al.[40]) We used the following energy parameters for modeling all the Baldwin peptide helix-coil experiments: $\epsilon_h = 0.786 - 0.011 \cdot [\text{Urea}]$ and $\epsilon_\phi = 0.218 - 0.014 \cdot [\text{Urea}]$.

4.2.1. Thermodynamic Parameters for the Helix-Coil Transition. Baldwin et al. have shown that ln s is linearly proportional to the concentration of urea:

$$\ln s = \ln s_o - \frac{m \cdot [\text{urea}]}{kT} \quad (2)$$

where m is the slope, T is absolute temperature, $k = 1.98$ cal·mol^{-1}·K^{-1} is the Boltzmann constant, and s_o is the propagation parameter at zero denaturant.[40] The propagation parameter in the Zimm−Bragg model is defined as the equilibrium constant for the addition of one more hydrogen bond to the end of an already existing sequence of hydrogen bonds.[7] The propagation parameter can be expressed in terms of thermal quantities as

$$s = \exp\left(\frac{\Delta S^o}{k} - \frac{\Delta H^o}{kT}\right) \quad (3)$$

where ΔS^o and ΔH^o are the standard entropy and enthalpy, respectively, for the formation of an added hydrogen bond. In our model, $\Delta S^o < 0$ when a hydrogen bond is formed following consecutive hydrogen bonds. There are four possible lattice sites available to a coil monomer but only a single lattice site available to a helix. The change in conformational entropy when a hydrogen bond is formed is given by

$$\frac{\Delta S^o}{k} = \ln \frac{1}{4} = -1.39 \quad (4)$$

The hydrogen bond enthalpy, ΔH^o (in units of kilocalories per mole; see section 4.2) is

$$\frac{\Delta H^o}{kT} = \frac{-(0.786 - 0.011 \cdot [\text{urea}])}{kT} \quad (5)$$

Combining eqs 3, 4, and 5, we conclude that $s_o = 1.06$ and $m = 0.011$ kcal/(mol·M·residue) at 0 °C. This is within a reasonable range of the experimental values of $s_o = 1.35$ and $m = 0.023$ kcal/(mol·M·residue) at the same temperature. By multiplying the m value by 50, the number of residues of the Baldwin one-helix bundle, we get theory and experimental m values for the helix of $m = 0.55$ and 1.15 kcal/mol·M, respectively (see the bottom of Table 1).

The second Zimm−Bragg quantity is the nucleation parameter, σ.[7] Because one hydrogen bond and one helical turn in our model involve three monomers, σ is the conformational entropy penalty for freezing two steps in the start of a helical turn. Hence, we have

$$\sigma = \left(\frac{1}{4}\right)^2 = 0.0625 \quad (6)$$

compared with a previously determined value of 0.003.[42]

4.3. Agreement with Experimental Data: Testing the Three-Helix Model. Now we shift our attention from single helices to three-helix bundles. We study the 58-residue B domain of protein A (BdpA) (Figure 1), a three-helix bundle, for which Oas et al. have collected extensive data. We study F13W*, which is a fluorescent mutant of BdpA, versus GuHCl at 37 °C as well as the equilibrium denaturation of F13W*, versus temperature at 2.2 M GuHCl.[29] Because these helices have an amino acid composition different from that of the Baldwin peptide, we require different average energy parameters. Figure 9 compares the experiments with our model using the parameters $\epsilon_{hb} = 0.6$, $\epsilon_\phi = 0.47 - 0.05 \cdot [\text{GuHCl}]$, and a two-helix term, $\epsilon_c = 0.69 \cdot \epsilon_\phi$. These parameter values give a ratio of hydrogen bond energy to hydrophobic interaction energy (at 0 M GuHCl) of

(39) Scholtz, J.; Marqusee, S.; Baldwin, R.; York, E.; Stewart, J.; Santoro, M.; Bolen, D. *Proc. Natl Acad. Sci. U.S.A.* **1991**, *88*, 2854−2858.
(40) Scholtz, J.; 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.; Qian, H.; York, E.; Stewart, J.; Baldwin, R. *Biopolymers* **1991**, *31*, 1463−1470.
(42) Scholtz, J.; Qian, H.; York, E.; Stewart, J.; Baldwin, R. *Biopolymers* **1991**, *88*, 1463−1470.

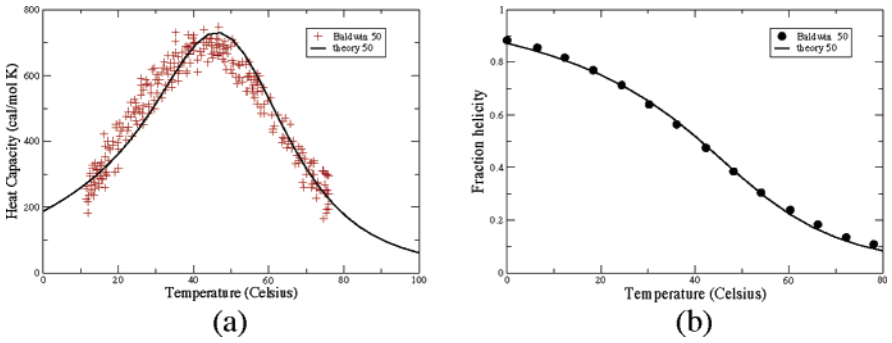

Figure 7. Our model captures heat capacity (a) and melting curve (b) of a 50-monomer helix-coil transition.

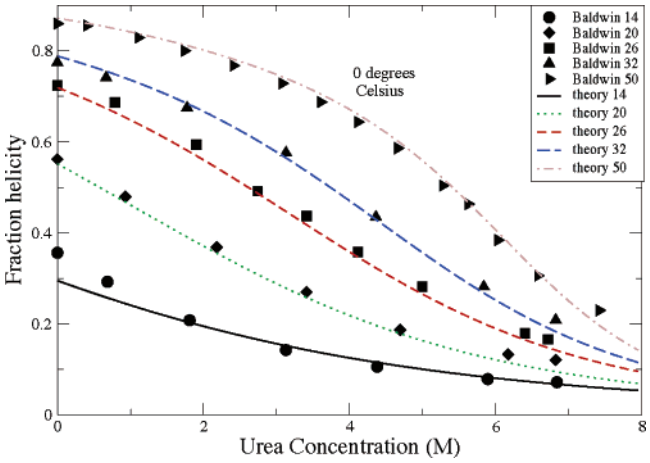

Figure 8. Present model captures well urea denaturation for helical peptides with repeating sequence Ala-Glu-Ala-Ala-Lys-Ala and chain lengths varying from 14 to 50 residues.

Table 1. Summary of Thermodynamic Parameters

helix bundle	temp (°C)	m	*m* expt	ΔG^{H_2O} (s_0)	ΔG^{H_2O} (s_0) expt
F13W*	25	2.30	1.46 ± 0.02	8.17	5.14 ± 0.45
α_3C	25	1.77	2.1	5.61	5.5
peptoid 3H	25	0.64	0.53 ± 0.07	4.04	3.4 ± 0.4
peptoid 2H	25	0.42	0.42 ± 0.03	2.78	2.7 ± 0.2
peptoid 1H	25	0.21	0.28 ± 0.03	1.22	1.5 ± 0.2
Baldwin 1H	0	0.55	1.15	(1.06)	(1.35)

0.6/0.47 = 1.277, which is close to the atomically detailed model results of Irbäck et al.,[43] giving 2.8/2.2 = 1.273. We found that the experimentally observed sharpness of this transition required a cooperative interaction energy that is roughly two-thirds the strength of the hydrophobic contact interaction, ϵ_ϕ. The dotted curve in Figure 9 shows the model prediction in the absence of the cooperative interaction energy, for comparison.

We also tested our model on α_3C, a synthetically designed 71-residue three-helix bundle, related to α_3D.[44] Figure 10 shows that our model captures thermal denaturation of α_3C for different GuHCl concentrations using energy parameters $\epsilon_{hb} = 0.6$, $\epsilon_\phi = 0.51 - 0.06 \cdot$[GuHCl], and $\epsilon_c = 0.19 \cdot \epsilon_\phi$. These values indicate that the protein α_3C has less intrinsic cooperativity than BdpA, even despite the greater chain length of the former. Table 1 compares various model thermal parameters with experiments for the various foldamers we tested.

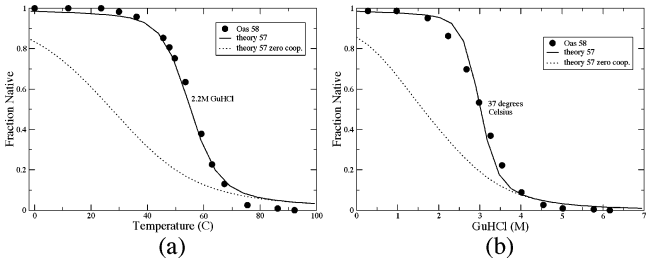

Figure 9. Predictions of the three-helix bundle model vs experiments of Oas et al. for denaturation by (a) temperature and (b) guanidine for F13W*. Temperature denaturation is carried out at 2.2 M GuHCl, and chemical denaturation is carried out at 37 °C. The dotted curve in each plot shows the model prediction, with the cooperative interaction energy set to zero.

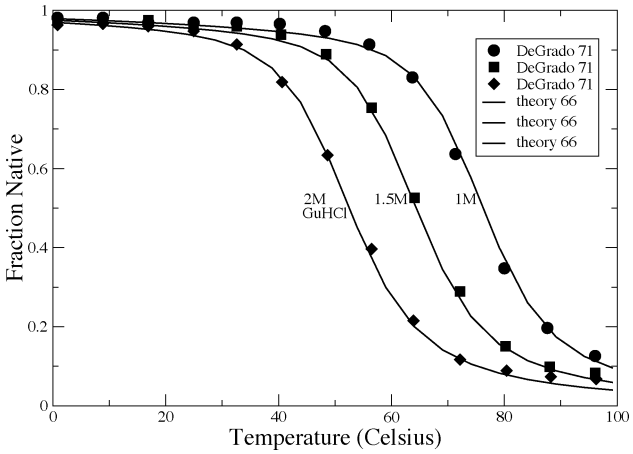

Figure 10. Thermal denaturation of α_3C for different GuHCl concentrations.

Not surprisingly, our two-helix energy parameter appears to be sequence-dependent. One explanation of this is that we are comparing cooperative energies between a natural and synthetic protein. The ϵ_{hb} and ϵ_ϕ values for the two three-helix bundles are similar.

4.4. The Conformational Populations. The value of a statistical mechanical theory, such as the present one, is in predicting not only average quantities, which are experimentally observable, but also the conformational distributions. In particular, we can study the nature of the cooperativity of the transition: if there are two distinct states populated at the midpoint of the transition, we have a two-state transition, such as a first-order phase transition in macroscopic systems. If there is a single broad state populated at the midpoint, we have a higher-order transition. Experiments show that the folding of small globular proteins is two-state.[10,45] We know of no experiments that yet bear on whether the helix-coil transition

(43) Irbäck, A.; Sjunnesson, F.; Wallin, S. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 13614−13618.
(44) Bryson, J.; Desjarlais, J.; Handel, T.; DeGrado, W. *Protein Sci.* **1998**, *7*, 1404−1414.

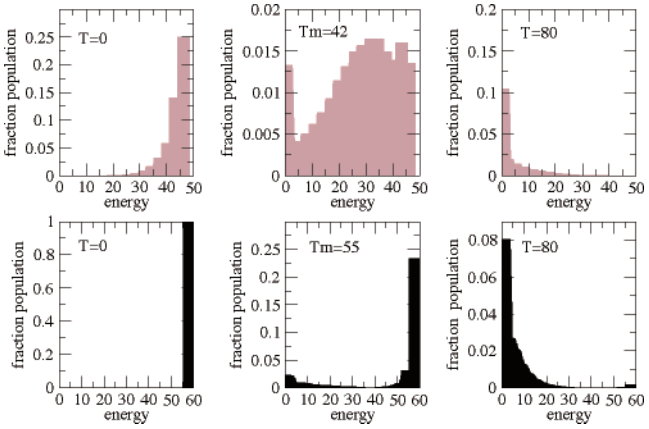

Figure 11. Top: Baldwin peptide transition is two-state, but very close to a continuous transition. At the melting temperature, there is a broad ensemble of non-native structures. Bottom: Oas three-helix bundle protein transition is two-state. At the melting temperature, near-native and near-open structures dominate the population.

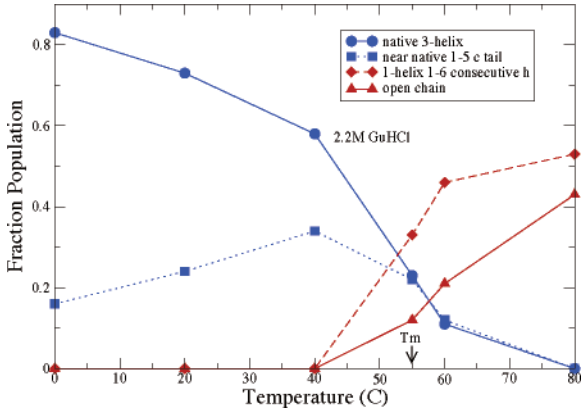

Figure 12. Dominant populations for the Oas 57-mer three-helix F13W*. At the midpoint $T_m = 55$, the two dominant populations are nativelike conformations (blue) or relatively open conformations (red). Unzipping of the helices occurs from the ends of the strands.

is two-state. We explore these issues with the present model.

Our philosophy is to first determine whether our model is consistent with all the known relevant data for the experimentally observable averaged quantities (described in the previous section), and then, if so, to predict the distribution functions using the same parameter values. This then gives what we believe are the current best estimates of the underlying distributions and allows us to interpret the basis for the cooperativity in the model. On this basis, Figure 11 shows our prediction that both the Baldwin 50-mer helix and the Oas three-helix bundle F13W* undergo two-state transitions. However, because both transitions are broad, it follows that the free energy barrier between the two states, in each case, is small. Figure 13b shows that F13W* has a near-zero folding activation energy at the mid-melting temperature. Our results agree with a recent study of λ_{6-85} by Gruebele and co-workers who predicted that $\Delta G\dagger$ for F13W* is only 0.45 kcal/mol at 37 °C; hence, they refer to it as a downhill folder.[28] Oas et al., however, observed exponential kinetics that are usually associated with a folding barrier, indicating some disagreement among experimentalists about the nature of the folding kinetics. Our studies here are limited to the equilibrium properties.

Figure 12 gives a more detailed breakdown of the populations for the Oas three-helix 57-mer, F13W*, as a function of temperature. We find that the protein B helices denature first from the ends, not the middle regions, since the former gives an advantage in conformational entropy. Raising the temperature from low temperatures melts the native structure, first by melting the tails of the two end helices, then the bundle falls apart into strands, each of which is a mix of helix and coil. Finally, the remaining helices melt out.

The model predicts that no two-helix conformation of the Oas three-helix bundle molecule is ever stable; the population of two-helix bundles is less than one percent of the total population at all temperatures (Figure 12). Figure 13a shows that the free energy of two-helix bundles is never simultaneously lower than that of a single helix and three-helix at 10 °C. Moreover, we find that the instability of two-helix bundles is not limited to this protein. Figure 14 shows that two-helix conformations are unstable at all chain lengths, at least for chain sequences for which these parameters are applicable. The three-helix bundle is more stable than the two-helix bundle because of the additional helix−helix interactions in the former. The one-helix bundle is more stable than the two-helix bundle because the former has more chain entropy and a higher density of hydrogen bonds (although fewer hydrophobic interactions). Comparing helix bundles of different length each having six turns on each strand, we find an increase in stability and cooperativity with an increase in the number of strands (Figure 15).

4.5. Peptoid Helix Bundles. Proteins are not the only polymers that can form helix bundles. Recently, Lee et al. have synthesized *polypeptoid* molecules that appear to fold into helix bundle structures. Lee et al. have done systematic studies of the unfolding of those molecules in mixed solvents acetonitrile and water (ACN).[35] To predict FRET efficiencies, we use dynamic programming to keep track of the distance between the two tagged monomers (acting as flourescent probe and receptor) and average over all conformations. We then use a relationship between distance and fluorescence to estimate FRET efficiency (see Supporting Information for details). For peptoid monomer sequences in ACN, we obtain the best fits with the parameters $\epsilon_{hb} = 1.59$ and $\epsilon_\phi = 1.19 - 0.08\cdot$[ACN]. The ratio of ϵ_{hb} to ϵ_ϕ at 0 M ACN is approximately the same (around 1.3) for peptoids as it is for proteins. Surprisingly, different cooperativity parameters of $\epsilon_c = 0.2\cdot\epsilon_\phi$ and $\epsilon_c = -0.3\cdot\epsilon_\phi$ for two-helix and three-helix bundle peptoids, respectively, are necessary to match Lee's experimental data. Figure 16 compares the model predictions to Lee's experimental unfolding data. We find one respect in which the peptoid helix bundles differ significantly from protein helix bundles. Whereas our models of three-helix proteins require a positive cooperativity parameter, our model of three-helix peptoids requires a negative cooperativity parameter. The dotted line in Figure 16 predicts peptoid three-helix bundle denaturation with positive cooperativity, $\epsilon_c = 0.2\cdot\epsilon_\phi$. The implication is that the three-helix bundle peptoids may be able to take advantage of hydrogen bonding and hydrophobic interactions but that these particular peptoid helices are not as well designed for packing or other cooperative interactions as are the two proteins we studied. Consistent with this observation, it has not yet been possible to crystallize peptoid helix bundles and obtain unique structures.

(45) Dill, K.; Bromberg, S.; Yue, K.; Fiebig, K.; Yee, D.; Thomas, P.; Chan, H. *Protein Sci.* **1995**, *4*, 561−602.

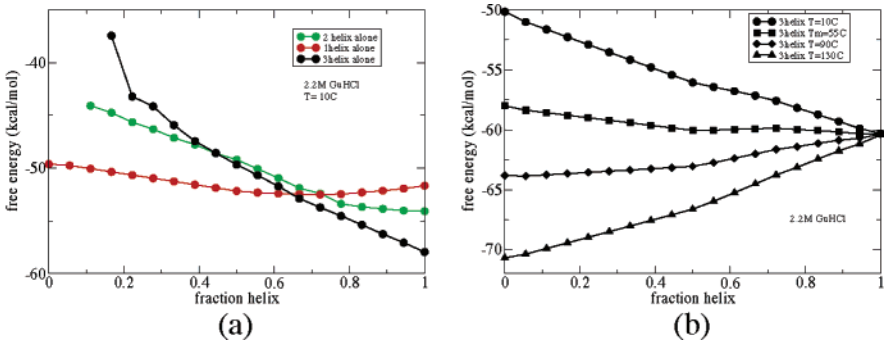

Figure 13. (a) Free energy diagrams of one-, two-, and three-helix bundle components of F13W* at 10 °C. No two-helix state is ever as stable as one- or three-helix conformations. The free energy of a two-helix bundle is never simultaneously lower than that of a single helix and three-helix bundle. (b) The free energy profiles for F13W* at various temperatures.

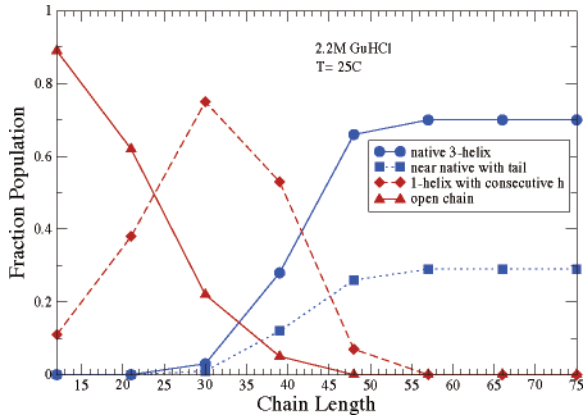

Figure 14. Dominant populations for three-helix bundles at different chain lengths, based on the F13W* energy parameters. Two-helix bundles are always less than 1% of the population.

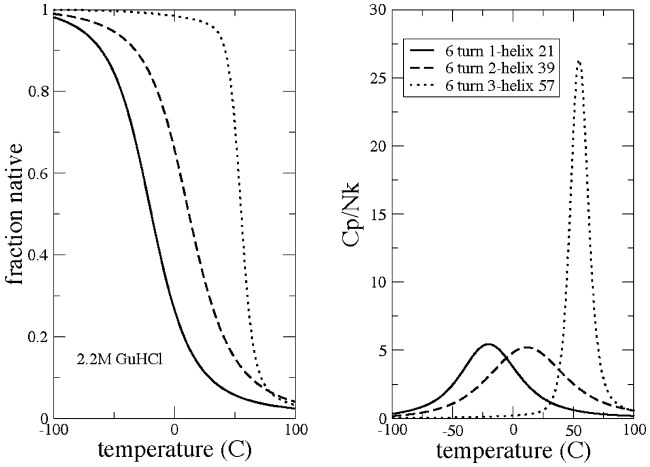

Figure 15. Comparing one-, two-, and three-helix bundles. Fraction native (left) and heat capacity per residue (right). One- and two-helix bundles have essentially the same cooperativity. Three-helix bundles are more cooperative.

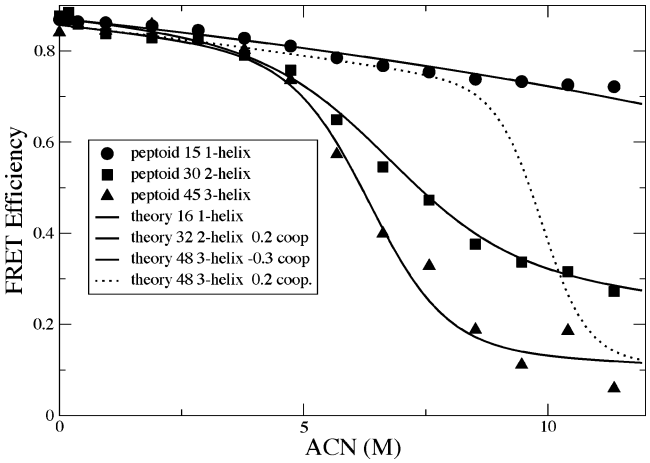

Figure 16. Model predictions for ACN-induced unfolding of peptoid helix bundles. The two-helix bundle denatures with a positive cooperativity parameter whereas the three-helix bundle denatures with a negative cooperativity parameter. The dotted line shows the expected denaturation plot for the three-helix bundle with a positive cooperativity parameter.

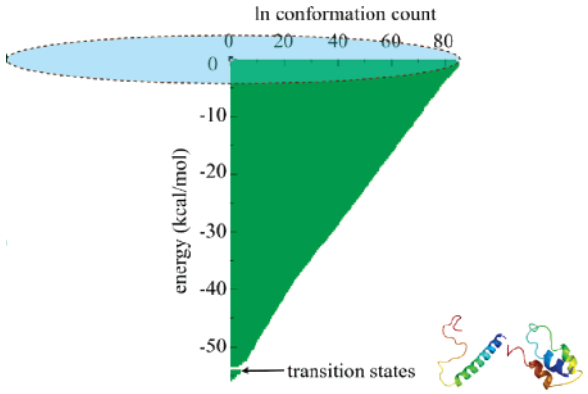

Figure 17. Energy landscape for the three-helix bundle F13W*. The transition states are single helices and partially folded three-helix bundles. Notice that the bottleneck state is very far down the landscape.

Another key difference the model predicts between peptoids and proteins: in proteins, two-helix bundles are not stable, whereas in peptoids they are. Consistent with this model prediction, Lee et al. constructed two-helix bundles, whereas there are few two-helix bundles in proteins as far as we are aware. Interestingly, most α-fibrous proteins are two-stranded coiled coils (for example, tropomyosin, myosin, paramyosin, intermediate filament proteins).[46] The reason for the greater stability of coiled coils compared to two-helix bundles, we believe, is that coiled coils have systematic side chain interactions resulting from heptad sequence repeats with hydrophobic residues at the "a" and "d" positions and charged residues at the "e" and "g" positions. These interactions allow a "knobs-into-holes" packing between residues in different α helical chains and strong interchain ionic interactions.[47] In globular proteins by contrast, such ideal coiled-coil packings are less common.[46]

(46) Cohen, C.; Parry, D. *Proteins: Struct., Funct., Genet.* **1990**, *7*, 1−15.

Table 2. Summary of Calorimetric Cooperativity Data

helix bundle	C	[GuHCl]	T_{max} (°C)	$C_{P,max}$	ΔH_{vH}	ΔH_{cal}	$(\Delta H_{vH}/\Delta H_{cal})$
F13W*	0	0	62	961	29.3	59.1	0.50
F13W*	0.69	0	104	2644	54.7	65.5	0.84
F13W*	0.69	2.2	55	2511	46.3	57	0.81
α_3C	0	0	89	1887	44.4	73.2	0.61
α_3C	0.19	0	100	2619	53.9	75.2	0.72
α_3C	0.19	1	76	2500	49.2	70.4	0.70
α_3C	0.19	1.5	64	2410	46.7	68.0	0.69
α_3C	0.19	2	52	2292	43.9	65.6	0.67
peptoid 3H	−0.3	0	707	1020.4	88.4.0	121.8	0.72
peptoid 3H	0.2	0	773	1152.0	100.2	131.3	0.76
Baldwin 1H	0	0	46	730	24.3	45.7	0.53

4.6. Cooperativity in Protein Folding. Protein folding is regarded as two-state when calorimetric measurements give a value of the ratio of van't Hoff to calorimetric enthalpies to be $\Delta H_{vH}/\Delta H_{cal} \approx 1$. There are some experimental challenges in subtracting baselines, leading to challenges in determining this number accurately. Nevertheless, this condition is approximately satisfied for typical single-domain proteins. Recently, Chan et al. have shown that existing polymer collapse and folding models, including the highly specific Gō model, fail to give $\Delta H_{vH}/\Delta H_{cal} \approx 1$.[21] The most cooperative folding model is currently that of Kaya and Chan, which is based on a favorable coupling between helix formation and the packing of the native core, in addition to an extra stabilizing energy for the ground state.[48] Their model predicts a $\Delta H_{vH}/\Delta H_{cal}$ ratio of 0.91 for a 55-mer chain sequence. They conclude that few models have yet been able to account for the high degrees of cooperativity of the folding of small proteins.

Table 2 shows the values that our model predicts for $\Delta H_{vH}/\Delta H_{cal}$. Of the molecules we studied, the Oas three-helix bundle F13W* at zero denaturant is predicted to be the most cooperative, with $\Delta H_{vH}/\Delta H_{cal} = 0.84$. The synthetic three-helix bundle, α_3C, should be less cooperative, with $\Delta H_{vH}/\Delta H_{cal} = 0.72$ at zero denaturant. Peptoid three-helix bundles, requiring a negative value of the intrinsic cooperativity parameter, are predicted to have $\Delta H_{vH}/\Delta H_{cal} = 0.72$. However, because they are too stable for thermal denaturation in water, this prediction about peptoids is not yet directly testable. For the Baldwin 50-mer helix, our model predicts a $\Delta H_{vH}/\Delta H_{cal}$ ratio of 0.53 using eq 4 in the Supporting Information for the van't Hoff enthalpy. Calculating the van't Hoff enthalpy instead at the population midpoint gives a slightly lower value of 0.49 for $\Delta H_{vH}/\Delta H_{cal}$. Experimental values for the Baldwin peptide range between 0.19 and 0.35 depending on their technique for establishing the baseline.[39]

Our model gives an explanation for the folding cooperativity in helix-bundle proteins. First, we find that the cooperativity does not simply reside within the helix-coil transitions of the individual helices themselves. Individual helices are neither sufficiently stable nor sufficiently cooperative to account for the experimental data on helix-bundle proteins. Second, our predicted cooperativities are also not the consequence of an amorphous nonspecific collapse process, as might be predicted from homopolymer collapse theories, because in our present model (as well as in real helix-bundle proteins, we believe), compactness of the chain cannot occur in the absence of substantial secondary structure.

Rather, as an indicator of the origins of cooperativity in the model, it is most informative to look at the least-populated conformations when the protein is at the midpoint of its denatured-to-native transition (Figure 17). We find that the states that are least populated are those partial folds having individual helices that overshoot their native lengths or partial folds having helices that are bundled but are just short stubs that are not the full native lengths. Such structures are part of the ensemble of maximum free energy that separates the denatured from native states. What is the thermodynamically most challenging step? We find that once two helices have formed near a common turn, so that the helices can pair together, the rest of the folding process is downhill.

Figure 17 also shows the location of this free energy bottleneck on an energy landscape. We believe this is the first quantitative energy landscape for a real protein that accurately captures the thermodynamics. Interestingly, it shows that the bottleneck state is very far down the landscape, very close to the native structure.

5. Conclusions

We introduce a new dynamic programming method for calculating the partition functions of chain molecules in which both local and nonlocal interactions play a role. In particular, we illustrate the method on chain-molecule helical bundles. We consider three types of energies: the hydrogen bonds within helices, the local and nonlocal hydrophobic interactions, and a cooperative interaction when helices come together into contact. The dynamic programming approach is vastly more efficient than full conformational enumeration and yet reproduces the densities of states quite accurately, for the cases we have tested.

Our ascending levels model gives good predictions for the thermal denaturation properties of the Baldwin peptide helices, the B domain of protein A studied by Oas and colleagues, and the synthetic protein α_3C. In addition, the model predicts the underlying conformational distributions throughout the melting processes. It shows that the conformational transitions in the Baldwin peptide and the Oas three-helix bundle are two-state, although with very small free energy barriers, and it shows that the former is "nearly'' higher-order. Interestingly, we find that the designed three-helix peptoid of Lee et al. folds with negative cooperativity. We find that the Oas three-helix bundle unfolds by fraying from the helical ends, then the bundle unfolds, then remaining helices become coiled. The model predicts that two-helix bundles should be unstable in protein backbones and stable in peptoids. Also, interhelical cooperative interactions are necessary to capture helix-bundle cooperativity.

(47) Crick, F. *Acta Crystallogr.* **1953**, *6*, 689−697.
(48) Kaya, H.; Chan, H. *Proteins: Struct., Funct., Genet.* **2003**, *52*, 510−523.

Acknowledgment. We thank the NSF, the NIH for Grant GM34993, Amgen, and UC Discovery for their support. We thank Kingshuk Ghosh for helpful discussions.

Supporting Information Available: Details of our dynamic programming approach, theory versus exact enumeration, van't Hoff and calorimetric enthalpy computations, and peptoid model. This material is available free of charge via the Internet at http://pubs.acs.org.

JA067153S