Structural propensities and entropy effects in peptide helix-coil transitions

Ilan E. Chemmama,[1] Adam Colt Pelea,[2] Yuba R. Bhandari,[1] Prem P. Chapagain,[1,*] and Bernard S. Gerstman[1,†]
[1]*Department of Physics, Florida International University, University Park, Miami, Florida 33199, USA*
[2]*Department of Mathematics and Statistics, Florida International University, University Park, Miami, Florida 33199, USA*
(Received 5 April 2012; revised manuscript received 13 July 2012; published 17 September 2012)

The helix-coil transition in peptides is a critical structural transition leading to functioning proteins. Peptide chains have a large number of possible configurations that must be accounted for in statistical mechanical investigations. Using hydrogen bond and local helix propensity interaction terms, we develop a method for obtaining and incorporating the degeneracy factor that allows the exact calculation of the partition function for a peptide as a function of chain length. The partition function is used in calculations for engineered peptide chains of various lengths that allow comparison with a variety of different types of experimentally measured quantities, such as fraction of helicity as a function of both temperature and chain length, heat capacity, and denaturation studies. When experimental sensitivity in helicity measurements is properly accounted for in the calculations, the calculated curves fit well with the experimental curves. We determine values of interaction energies for comparison with known biochemical interactions, as well as quantify the difference in the number of configurations available to an amino acid in a random coil configuration compared to a helical configuration.

DOI: 10.1103/PhysRevE.86.031915 PACS number(s): 87.15.−v, 82.20.−w, 82.35.−x

I. INTRODUCTION

The structural transition from random coil to α-helix is intensely studied as part of the protein folding problem. An early model developed by Schellman [1] assumed that the coil-to-helix transition starts at one end of the peptide and the helical structure propagates down the chain. In the groundbreaking work of Zimm and Bragg [2], the possibility of interspersing sections of helices and coils was included. The probability for a segment of the chain to undergo the coil-to-helix transition is characterized through the use of two statistical parameters: the nucleation probability to form the first helical turn to start a segment, and the helix propagation probability down the chain that represents a cooperative effect between nearest-neighbor amino acids. The Zimm-Bragg model was refined by Lifson and Roig [3] by extending the interactions to next-nearest neighbors, which only increased the rank of the transfer matrix from two to three.

Other investigations have expanded on the work of the previous models (see, for example, Refs. [4–6] and references therein). Models for the 3,10-helix–coil transition [7,8] as well as the β-sheet–coil transition [9] or the α-helix–β-sheet–coil transitions [10,11] have been developed. More recently, models have been presented that incorporate tertiary contact into the formalism of the transition [12]. Models that were developed to describe helix-bundle proteins and the cooperative effect in repeat proteins [13,14] have also been used to describe the single-helix transition.

The theoretical work presented in this paper builds upon other models to investigate the transition from the random coil to the α-helix configuration in terms of the propensity of amino acids to form an α-helix structure, and the hydrogen bond and dipole interactions that occur in α-helices. The helix propensity energy contribution is frequently used in the investigations of peptide structural transitions [15–20], and quantifying its role relative to other interactions is important. We develop a method for obtaining and incorporating the degeneracy factor for a peptide's helix-coil states that allows the exact calculation of the partition function for a peptide with any number of nucleation events. The partition function is used in calculations that allow comparison with experimentally measured quantities, such as fraction of helicity as a function of both temperature and chain length, and heat capacity. This allows the determination of values of propensity and interaction energies for comparison with known biochemical interactions, as well as the quantification of the difference in the number of configurations available to an amino acid in a random coil configuration compared to a helical configuration. Our calculated results are compared to experimental data for engineered single α-helices [21–24]. This comparison shows that the helicity calculation must be modified to take into account experimental sensitivity. When properly accounted for in the calculations, the calculated curves fit well with the experimental curves.

II. PEPTIDE MODEL: CHAIN ENERGY AND CONFIGURATIONAL ENTROPY

The model that we use for investigating the coil-helix transition allows each amino acid to be in one of two possible configurations, either random coil (C) or α-helix (H). Thus, a peptide configuration can be represented by a string of H's and C's. As in cubic lattice models, 4 amino acids are necessary to complete a helical turn instead of the 3.6 amino acids per turn in actual proteins. The energy of a peptide sequence is given relative to the configuration in which all amino acids are in the C configuration, which is defined to have zero energy.

The energy of a peptide configuration is determined by three different terms in the Hamiltonian,

$$H = H_L + H_M + H_{\mathrm{HBD}}. \tag{1}$$

The first term H_L is an energy that depends on the state of an individual amino acid. This term reflects the fact that some amino acids are known to have an especially strong propensity to assume an α-helical configuration. This can be due to

*chapagap@fiu.edu
†gerstman@fiu.edu

hydrophobic side chains burying themselves against the side of a helix, as well as other structural factors [25–27]. We label this as a local effect because it involves only a single amino acid. If an amino acid is in the C configuration, then H_L assigns an energy of zero. If the amino acid is in the H configuration, then H_L assigns a favorable, lower energy of $-E_L$, with E_L a positive quantity. The propensity for an amino acid to assume an α-helical configuration can also be affected by a neighboring residue [26]. The term H_M is a cooperative interaction between nearest-neighbor amino acids along the chain. It encourages the propagation of helical structure by lowering the energy of nearest-neighbor residues by $-E_M$ if they are both in H states. This cooperative term H_M is labeled a medium-range interaction because it involves the interaction of nearest-neighbor residues. The term H_{HBD} represents the combined effect of the hydrogen-bond interaction and dipole interaction that occurs between amino acid j and amino acid $j+4$ if amino acids j, $j+1$, $j+2$, $j+3$, and $j+4$ are all in the H configuration. This energy $-E_{\mathrm{HBD}}$, with E_{HBD} a positive quantity, provides additional stability to segments of the polypeptide chain that have assumed measurable α-helix structure.

Energy considerations favor the H state for each amino acid, but entropy considerations favor C. The transition from random coil to α-helix induces a loss of entropy due to the fact that there are many ways, Ω_C, for an amino acid to be in a random coil C state, but much fewer ways, Ω_H, to be in H. A similar entropic factor is used in a model [28] to explain protein cold denaturation. The relative probabilities depend on the ratio Ω_C/Ω_H. As with Ref. [14], we set $\Omega_H = 1$ so that we need only explicitly include Ω_C. We note that in the equations below, wherever Ω_C appears, it represents the ratio Ω_C/Ω_H. The differences in energy and entropy between the C and H configurations for a single amino acid are incorporated in the relative probability given by the properly weighted Boltzmann factor,

$$P_{HC} = \frac{P_H}{P_C} = \frac{e^{-E_H/RT}}{(\Omega_C e^{-E_C/RT})} = \frac{e^{E_L/RT}}{\Omega_C}, \qquad (2)$$

where $E_C - E_H = 0 - (-E_L) = E_L$. Greater $|E_L|$ favors H; greater Ω_C favors C. Each amino acid in the peptide that switches from coil to helix introduces another P_{HC} factor given in Eq. (2) involving both E_L and Ω_C. Equation (2) is equivalent to the standard expression for the relative probability of two configurations, $P_{HC} = P_H/P_C = e^{-\Delta G/RT}$ where $\Delta G = \Delta E - T\Delta S = G_H - G_C = (E_H - E_C) - T(S_H - S_C)$ and $S_{H/C} = R \ln \Omega_{H/C}$.

The numerical value of Ω_C is expected to depend on the number of amino acids N in the chain, i.e., $\Omega_C = \Omega_C(N)$. The dependence is due to steric hindrance. Longer chains are more likely to self-intersect in the random coil configuration [29]. Self-intersection reduces the number of states that each amino acid can assume, thus lowering the average value of Ω_C for each amino acid in longer chains. As described in a later section, when we apply our model to experimental data, we find $\Omega_C(N)$ to be a decreasing function of N, consistent with self-intersecting steric hindrance.

III. HELIX-COIL PARTITION FUNCTION

To compare our model with experimentally observable quantities such as heat capacity, or the fraction of amino acids in the helical state as a function of both temperature and chain length, it is necessary to calculate the partition function. Since the energy of an amino acid in the coil state is defined as $E_C \equiv 0$, the Boltzmann factor contribution from an amino acid in the coil state is just Ω_C. To determine contributions from an amino acid in the helical state, we define the parameters $l \equiv e^{E_L/RT}$, $m \equiv e^{E_M/RT}$, and $h \equiv e^{E_{\mathrm{HBD}}/RT}$ that represent the Boltzmann factors for amino acids in the helix state from the local interaction, the medium interaction, and the hydrogen-bond-dipole interaction, respectively. The partition function Z in terms of energy levels and their associated degeneracy is given by

$$Z(N,T) = \sum_{i=0}^{N} \sum_{j=0}^{i-1} \sum_{k=0}^{j-3} g(N,i,j,k) \Omega_C^{N-i} l^i m^j h^k, \qquad (3)$$

where N is the number of amino acids in the polypeptide chain. For any specific configuration, the i summation index represents the number of amino acids that are in the helix state (local interaction) leaving $N-i$ amino acids in the coil state, j represents the number of helical pairs (medium interaction), and k represents the number of hydrogen bonds (hydrogen-bond-dipole interaction). The values of i,j,k determine the energy of a configuration, and the degeneracy factor $g(N,i,j,k)$ is the number of distinct multiparticle configurations of a chain of length N that contributes the same Boltzmann factor to Z. For shorter peptides, all configurations can be easily enumerated, as shown in the Appendix. For longer peptides, a numerical method for calculating degeneracy factors is necessary, and is derived as follows.

For the degeneracy factor g, we now show how to construct and count the number of possible binary strings (consisting of C's and H's) for a chain of length N with exactly i, j, and k instances of the substrings H, HH, and HHHHH, respectively. The term "block" will refer to a substring consisting entirely of H's that is maximal. For example, the binary string CHHCCHHHH has only the two blocks HH and HHHH. There can be no more than $N-i+1$ blocks since there must be at least one C between any two blocks. The actual number of blocks, A (of length $\geqslant 1$), represents the number of helical nucleations in the entire string and depends on the specific ordering of C's and H's, with $A_{\max} = N - i + 1$. The number of blocks can be determined as follows. For each block $r = 1, \ldots, A$, the parameter i_r is the number of instances of the substring H and satisfies the condition $i = \sum_{r=1}^{A} i_r$. Similarly, we define j_r to be the number of instances of the substring HH in the rth block and satisfies $j = \sum_{r=1}^{A} j_r$. Since blocks consist entirely of H's, we have $i_r = j_r + 1$ for every r, so that we find $i = \sum_{r=1}^{A} (j_r + 1) = j + A$. Therefore, the number of blocks, A, can be expressed as $A = i - j$.

Continuing to longer blocks, let a represent the number of instances of the triplet substring HHH in a chain and let b represent the number of instances of the quadruplet HHHH. Applying similar reasoning, we find that the number of blocks of length of at least two is $(j-a)$, the number of blocks of length of at least three is $(a-b)$, and the number of blocks of length of at least four is $(b-k)$. If there can be exactly a instances of HHH and b instances of HHHH, then we must choose, from the $N-i+1$ possible positions for

blocks, exactly $(i-j)$ of them to actually have at least one H. That is, the first step in our construction is to distribute $(i-j)$ H's among the spaces between the C's where we avoid creating a block of length two. The number of ways to do this is given by the binomial coefficient $\binom{N-i+1}{i-j}$. In a similar fashion, the number of ways to have exactly $(j-a)$ blocks with length of at least two H's out of $(i-j)$ blocks is $\binom{i-j}{j-a}$, the number of ways to have exactly $(a-b)$ blocks of length of at least three H's out of $(j-a)$ blocks is $\binom{j-a}{a-b}$, and the number of ways to have exactly $(b-k)$ blocks with length of at least four H's out of $(a-b)$ blocks is $\binom{a-b}{b-k}$. Since it does not matter how many blocks are of length of at least five or greater, we are free to distribute the final k H's among the existing $(b-k)$ blocks of length four. Thus, the number of ways to distribute the remaining k H's is given by $\binom{b-1}{k}$. Since our task did not require that we have precisely a and b instances of the substrings HHH and HHHH, respectively, we sum over their possible values $(k \leqslant b \leqslant a \leqslant j)$ to obtain the desired degeneracy factor:

$$g(N,i,j,k) = \sum_{a=k}^{j}\sum_{b=k}^{a}\binom{1+N-i}{i-j}\binom{i-j}{j-a}\binom{j-a}{a-b} \times \binom{a-b}{b-k}\binom{b-1}{k}, \tag{4}$$

where the following rules dictate the mathematical properties of the coefficients: $\forall x \in \mathbb{N}$, $\binom{x}{0} = 1$; $\forall y \in \mathbb{N}^*, \binom{0}{y} = 0$; and $(\forall x,y \in \mathbb{N}^*) \wedge (y > x), \binom{x}{y} = 0$. As an example, in the Appendix we present Table II for $N = 5$ that lists all of the configurations grouped by energy, the corresponding number of entropy factors Ω_C, and the degeneracy g of that energy level. A microdegeneracy factor g' that is explained later is also listed.

IV. COMPARISON WITH EXPERIMENTS

A. Helicity

For comparison with experimental measurements, physical observables can be theoretically calculated by taking various derivatives of the partition function. One of the most natural physical observables for the helix-to-coil transition is the helicity θ representing the fraction of amino acids that are in the helical state, defined as

$$\theta_N = \frac{\langle i_N \rangle}{N}, \tag{5}$$

where $\langle i_N \rangle$ is the average number of amino acids in the helical state in a peptide chain of length N. The helicity can be calculated from the partition function using

$$\theta_N = \frac{l}{NZ}\frac{\partial Z}{\partial l}. \tag{6}$$

We used the partition function Z of Eq. (3) in Eq. (6) to calculate the helicity defined in Eq. (5) as a function of the temperature and chain length to compare to the experimental values from Ref. [22]. The values of the energy parameters E_L, E_M, and E_{HBD} of Eq. (1) were treated as variables in a fitting routine, but were constrained to be within a biochemically reasonable range. We found that the theoretical curves do not fit the experimental data well. This disagreement can be resolved if we modify the theoretical treatment to take into account the sensitivity of the experimental arrangement. An isolated, single amino acid in the helical conformation would, alone, not produce a strong enough circular dichroism (CD) signal to be detected [30]. Instead, it takes a few contiguous amino acids to all be in the helical (H) configuration in order to produce a sufficient signal. To take this into account, we modify our calculations by assuming that a segment in a peptide must have at least three contiguous residues in the helical configuration in order to contribute to θ.

The calculations are modified in the following manner. Equation (5) is changed to denote that only helical segments with at least three amino acids are included in the calculation for the helicity,

$$\theta_{N,3} = \frac{\langle i_{N,3} \rangle}{N} = \frac{\langle i_N - i_{N,1} - i_{N,2} \rangle}{N}, \tag{7}$$

where, for a chain of length N, $i_{N,3}$ is the number of amino acids in helical segments of length three or greater, $i_{N,1}$ is the number of helical amino acids that are in a block of length one (with a random coil amino acid on either side, i.e., . . .CHC. . .), and $i_{N,2}$ is the number of amino acids in helical blocks of length two. In order to calculate $i_{N,1}$, it is necessary to determine the number of isolated singlet H blocks in a chain of length N, and to calculate $i_{N,2}$, it is necessary to determine the number of doublet HH blocks. To calculate the number of singlet blocks, we follow the approach used in the derivation of the degeneracy factor g in Eq. (4). We know that there are $i-j$ blocks of at least one H, and $j-a$ blocks of at least two H's. Therefore, there are $(i-j)-(j-a) = i-2j+a$ blocks of precisely one singlet H. Likewise, since there are $a-b$ blocks of at least three H's, there are $j-2a+b$ blocks of precisely two H's. The number of H's that do not contribute to θ because they are located in singlets or doublet blocks is therefore $i_{N,1}+i_{N,2} = (i_N - 2j + a) + 2(j - 2a + b) = i_N - 3a + 2b$. Inserting this into Eq. (7) gives

$$\theta_{N,3} = \frac{\langle i_{N,3} \rangle}{N} = \frac{\langle 3a - 2b \rangle}{N}. \tag{8}$$

It can be shown that if it is desired to omit only H's in blocks of length one (. . .CHC. . .), then the term $(3a-2b)$ must be changed to $(2j-a)$ when calculating the helicity. If it is desired to remove from the helicity calculation H's that are in blocks of length one, length two, and length three, then $(3a-2b)$ must be changed to $(4b-3k)$.

The helicity $\theta_{N,3}$ can be calculated from Eq. (8) using the following expression:

$$\theta_{N,3} = \frac{1}{NZ}\left[\sum_{i=0}^{N}\sum_{j=0}^{i-1}\sum_{a=0}^{i-2}\sum_{b=0}^{i-3}\sum_{k=0}^{i-4}(3a-2b) \times g'(N,i,j,a,b,k)\Omega_C^{N-i}l^i m^j h^k\right]. \tag{9}$$

The expression for the modified $\theta_{N,3}$ in Eq. (9) contains summations over the indices a and b, along with a modified degeneracy factor g'. Summations over a and b must be included in Eq. (9) because the modified g' is a microdegeneracy

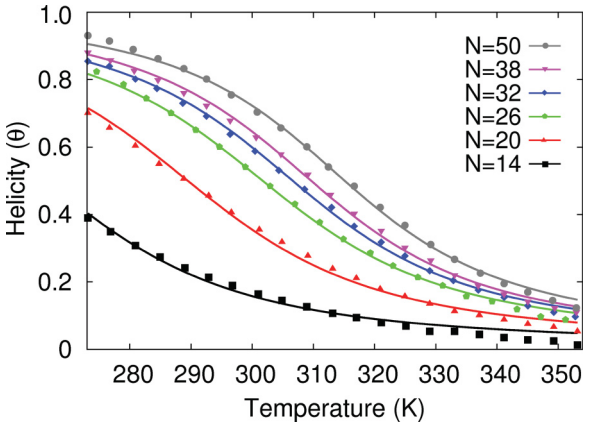

FIG. 1. (Color online) Comparison between calculated and experimental values [23] of helicity θ as a function of temperature for different chain lengths for a homopolymer chain. The ordering of the curves from top to bottom is the same ordering of N used in the legend.

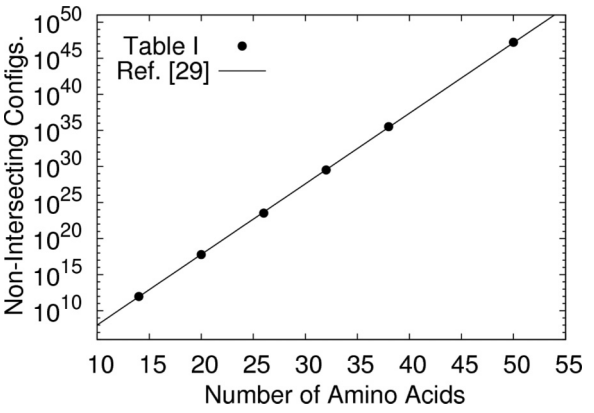

FIG. 2. The points (•) represent the number of non-self-intersecting chain configurations $\Omega_T(N)$ as a function of chain length N determined by using the $\Omega_C(N)$ of Table I of this paper compared to the $\Omega_T(N)$ curve (—) determined from Ref. [29].

factor that does not include the summations over a and b:

$$g'(N,i,j,a,b,k) = \binom{1+N-i}{i-j}\binom{i-j}{j-a}\binom{j-a}{a-b} \times \binom{a-b}{b-k}\binom{b-1}{k}. \quad (10)$$

The microdegeneracy factor g' does not represent all configurations with the same energy. It is a restricted degeneracy factor that gives the number of configurations with the same energy and with patterns containing blocks of helical segments of the same length. An example of the difference between g and g' is given in the Appendix.

Figure 1 shows that the calculated curves using the modified expression for the partition function match the experimental helicity data well. All calculated curves were generated with the same set of values for the energy terms: $E_L = 0.200$, $E_M = 0.810$, and $E_{\mathrm{HBD}} = 0.501$ kcal/mol. The values for Ω_C as a function of N are discussed below.

B. Self-intersecting configurations and the entropy factor $\Omega_C(N)$

Along with E_L, E_M, and E_{HBD}, we also determine Ω_C as a function of N by fitting to the experimental data for helicity. The values of $\Omega_C(N)$ are presented in Table I.

Table I shows that Ω_C decreases as a function of increasing chain length, which is consistent with the fact that longer chains have a greater fraction of configurations that are self-intersecting and not physically allowed. Therefore, some of the Ω_C configurations that can be assumed by individual amino acids in short chains are not possible in longer chains. The values of $\Omega_C(N)$ given in Table I are calculated by using our model of helix-coil transitions to fit the experimental data from helicity studies. These values are consistent with a previous, independent study of the chain-length dependence of the fraction of a chain's configurations that are non-self-intersecting. Reference [29] used a computer lattice model of protein dynamics to investigate protein entropy. Figure 5 of Ref. [29] is a graph of the percentage $f(N)$ of chain configurations that are non-self-intersecting as a function of the number of amino acids N. It is seen in that figure that $f(N)$ is a decreasing function of N, i.e., $f(N) = ae^{-bN}$. The total number of non-self-intersecting chain configurations $\Omega_T(N)$ was expressed as $\Omega_T(N) = f(N)[\Omega_o]^N$, where Ω_o is a constant that represents the number of configurations that a single isolated amino can assume without concern for self-intersection. Figure 2 in this paper compares the number of total non-self-intersecting chain configurations $\Omega_T(N)$ determined by using the $\Omega_C(N)$ of Table I of this paper to the number of non-self-intersecting configurations $\Omega_T(N)$ determined from the work of Ref. [29]. $\Omega_T(N)$ was calculated from the results of this paper using the expression $\Omega_T(N) = [\Omega_C(N) + 1]^{N-3}$, with the $\Omega_C(N)$ for $N = 14$, 20, 26, 32, 38, 50 from Table I. The exponent $N - 3$ is used here because we assume that a chain is not flexible enough, and cannot possibly self-intersect, until it has at least four amino acids. The $\Omega_T(N)$ points calculated from Table I agree well with the $\Omega_T(N)$ curve calculated from Fig. 5 of Ref. [29], even though they are determined independently. As with Ref. [29], the total number of chain configurations increases with N, but at a decreasing rate of growth. The decrease in the rate of growth of $\Omega_T(N)$ as a function of N is now explained by the fact that the entropy factor $\Omega_C(N)$ for each amino acid is a decreasing function of N.

TABLE I. Entropy factor Ω_C as a function of chain length N determined by using Eq. (9) to fit the experimental data for helicity.

N	14	20	26	32	38	50
Ω_C	11.26	10.12	9.55	9.42	9.35	9.11

C. Heat capacity

We use our theoretical treatment to calculate the heat capacity using the partition function Z of Eq. (3) in the following expression [14,31]:

$$C_v = \frac{d\langle E\rangle}{dT} = \frac{2RT}{Z}\frac{dZ}{dT} + RT^2\frac{d}{dT}\left(\frac{1}{Z}\frac{dZ}{dT}\right). \quad (11)$$

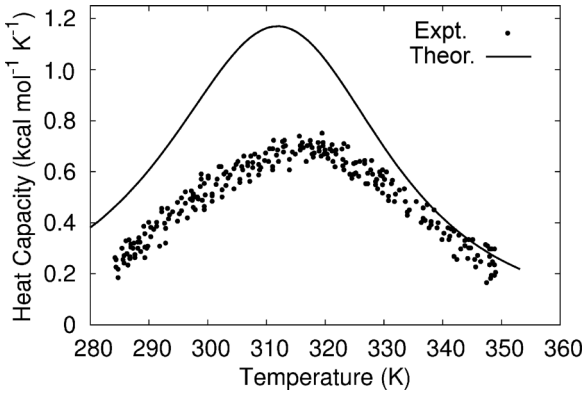

FIG. 3. Theoretically calculated C_v compared to the experimental C_p from Ref. [22] for a chain of length $N = 50$.

In Fig. 3, we compare the theoretically calculated C_v to the experimental C_p from Ref. [22] for a chain of length $N = 50$. The theoretical curve uses the same values for E_L, E_M, E_{HBD}, and Ω_C that were determined from the fits to the helicity data in Fig. 1. The shape of the calculated heat capacity curve matches the experimental data. The calculated values are higher, but the peak position for the theoretical curve occurs within three degrees of the experimental data (312 vs 315 K).

D. Denaturation

We used our model to fit experimental data on the effect of the denaturant urea on the helicity of chains of different lengths at $T = 273$ K from Ref. [24]. As discussed in Refs. [14,32,33], the denaturant is expected to lower the ability of a peptide chain to form α-helices. Following Refs. [14,32], the effect of the denaturant can be included in our model by decreasing the effective strength of E_{HBD} using the expression $E_{\mathrm{HBD}}(C) = E^o_{\mathrm{HBD}} - m_H C$, where C is the concentration of urea, and m_H is a constant whose value is determined by fitting the data. Following Ref. [33], which found that urea decreases the intrinsic propensity of individual amino acids to assume an α-helix configuration, we used an analogous expression for E_L: $E_L(C) = E^o_L - m_L C$. For E^o_L and E^o_{HBD}, we used the same values that we determined for E_L and E_{HBD} from fitting the experimental helicity data in Fig. 1. All curves in Fig. 4 used the same values of $m_L = 0.013$ kcal/(mol M) and $m_H = 0.027$ kcal/(mol M). Our value of m_H is very similar to the analogous quantity $m = 0.029$ kcal/(mol M) given in Ref. [14]. Because of differences in the models, Ref. [14] has no parameter equivalent to m_L for comparison.

As seen in Fig. 4, our model fits the experimental data reasonably well, but not perfectly. Interestingly, the experimental helicity data at a urea concentration of zero is not the same as the experimental helicity data points at 273 K in Fig. 1, which were measured for the same peptides [23]. For $N \geqslant 20$, the experimental helicity data at zero urea concentration is lower than the helicity data at 273 K. This systematic discrepancy may explain why our calculated curves for $N \geqslant 20$ in Fig. 4 are higher than the experimental data at zero urea concentration.

E. Van't Hoff index of cooperativity

An important aspect of protein behavior is cooperativity. In our Hamiltonian of Eq. (1), H_M and H_{HBD} represent

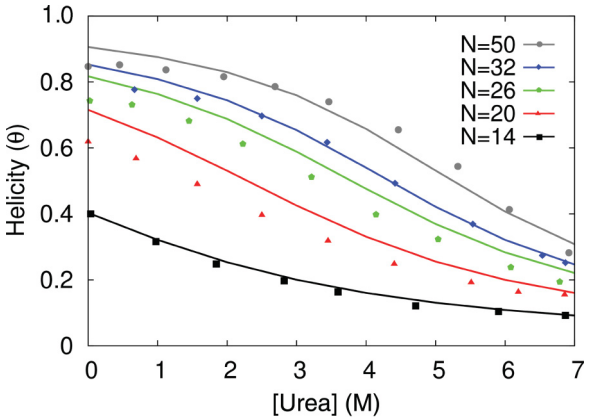

FIG. 4. (Color online) Comparison between calculated and experimental values [24] of helicity θ as a function of urea concentration. The ordering of the curves from top to bottom is the same ordering of N used in the legend.

cooperative, multiparticle interactions, whereas H_L represents a localized, individual amino acid, noncooperative effect. The extent to which these different interactions create cooperative behavior can be quantified through the ratio δ_c of the van't Hoff enthalpy to the calorimetric enthalpy [14,34]. Values of δ_c close to 1.0 signify a highly cooperative system, whereas values close to zero represent a system with little cooperativity. Using our calculational model, we can get an approximate numerical value for δ_c by calculating the ratio of a peptide chain's van't Hoff energy to its calorimetric energy,

$$\delta_c = \frac{\Delta E_{\mathrm{VH}}}{\Delta E_{\mathrm{cal}}}. \tag{12}$$

The van't Hoff energy ΔE_{VH} can be calculated with the expression [34]

$$\Delta E_{\mathrm{VH}} = 4RT^2_{\theta=1/2} \left.\frac{d\theta}{dT}\right|_{\theta=1/2}, \tag{13}$$

where θ is the helicity. The calorimetric energy, defined as the energy difference between the native state and the denatured state, can be calculated with the expression [14,34]

$$\Delta E_{\mathrm{cal}} = \langle E \rangle_{\mathrm{Nat}} - \langle E \rangle_{\mathrm{DN}}. \tag{14}$$

Evaluating these energies using our model for $N = 50$, we get calculated values of $\Delta E_{\mathrm{VH}} = -12.59$ kcal/mol and $\Delta E_{\mathrm{cal}} = -57.04$ kcal/mol. This gives a ratio $\delta_c = 0.22$, which falls nicely in the range of 0.16–0.34 that was determined experimentally in Ref. [22].

V. CONCLUSION

Using the hydrogen bond and helix forming propensity, we developed an exact analytical expression for the partition function for a peptide chain in which amino acids can undergo a helix-coil transition. We compared the calculated results with experimental data. The theoretical treatment agrees well with the experimental measurements when the experimental sensitivity is taken into account. We have shown that the experimental data can be used to determine a numerical value for Ω_C/Ω_H and quantified how the non-self-intersection physical constraint lowers Ω_C as the length of the chain increases.

APPENDIX

As an example of the various contributions to the partition function of Eq. (3), in Table II we give all states with their energies, entropy factors, and degeneracy factors for a short peptide of length $N = 5$.

TABLE II. Example for $N = 5$.

Configuration	Energy	Entropy Factor	g'	g
CCCCC	0	Ω_C^5	1	1
CCCCH, CCCHC, CCHCC, CHCCC, HCCCC	E_L	Ω_C^4	5	5
HCCCH, HCCHC, HCHCC, CHCCH, CHCHC, CCHCH	$2E_L$	Ω_C^3	6	6
CCCHH, CCHHC, CHHCC, HHCCC	$2E_L + E_M$	Ω_C^3	4	4
HCHCH	$3E_L$	Ω_C^2	1	1
HCCHH, HCHHC, CHCHH, CHHCH, HHCHC, HHCCH	$3E_L + E_M$	Ω_C^2	6	6
CCHHH, CHHHC, HHHCC	$3E_L + 2E_M$	Ω_C^2	3	3
HCHHH, HHHCH	$4E_L + 2E_M$	Ω_C^1	2	3
HHCHH	$4E_L + 2E_M$	Ω_C^1	1	
CHHHH, HHHHC	$4E_L + 3E_M$	Ω_C^1	2	2
HHHHH	$5E_L + 4E_M + E_{\mathrm{HBD}}$	1	1	1

[1] J. A. Schellman, C. R. Trav. Lab Carlsberg Ser. Chim. **29**, 230 (1955).
[2] B. H. Zimm and J. K. Bragg, J. Chem. Phys. **31**, 526 (1959).
[3] S. Lifson and A. Roig, J. Chem. Phys. **96**, 1963 (1961).
[4] N.-V. Buchete and J. E. Straub, J. Chem. Phys. B **105**, 6684 (2001).
[5] A. J. Doig, Biophys. Chem. **101-102**, 281 (2002).
[6] D. J. Jacobs and G. G. Woods, Biopolymers **75**, 1 (2004).
[7] C. A. Rohl and A. J. Doig, Protein Sci. **5**, 1687 (1996).
[8] J. K. Sun and A. J. Doig, Protein Sci. **7**, 2374 (1998).
[9] W. L. Mattice and H. A. Scheraga, Biopolymers **23**, 1701 (1984).
[10] L. Hong, J. Chem. Phys. **129**, 225101 (2008).
[11] J. S. Schreck and J. M. Yuan, Phys. Rev. E **81**, 061919 (2011).
[12] A. C. Hausrath, J. Chem. Phys. **125**, 084909 (2006).
[13] A. Lucas *et al*., J. Am. Chem. Soc. **129**, 4272 (2007).
[14] K. Ghosh and K. A. Dill, J. Am. Chem. Soc. **131**, 2306 (2009).
[15] P. Pokarowski, A. Kolinski, and J. Skolnick, Biophys. J. **84**, 1518 (2003).
[16] R. Dima and D. Thirumalai, Biophys. J. **83**, 1268 (2002).
[17] P. D. Thomas and K. A. Dill, Protein Sci. **2**, 2050 (1993).
[18] S. Govindarajan and R. A. Goldstein, Proteins **22**, 413 (1995).
[19] J. Skolnick and A. Kolinski, Science **250**, 1121 (1990).
[20] P. P. Chapagain, Y. Liu, and B. S. Gerstman, J. Chem. Phys. **129**, 175103 (2008).
[21] S. Marqusee and R. L. Baldwin, Proc. Natl. Acad. Sci. USA **84**, 8898 (1987).
[22] J. M. Scholtz *et al*., Proc. Natl. Acad. Sci. USA **88**, 2854 (1991).
[23] J. M. Scholtz *et al*., Biopolymers **31**, 1463 (1991).
[24] J. M. Scholtz *et al*., Proc. Natl. Acad. Sci. USA **92**, 185 (1995).
[25] M. Blaber, X. J. Zhang, and B. W. Matthews, Science **260**, 1637 (1993).
[26] C. N. Pace and J. M. Scholtz, Biophys. J. **75**, 422 (1998).
[27] J. Wang and J. A. Feng, Protein Eng. **16**, 799 (2003).
[28] A. V. Badasyan *et al*., Phys. Rev. E **83**, 051903 (2011).
[29] P. P. Chapagain, J. L. Parra, B. S. Gerstman, and Y. Liu, J. Chem. Phys. **127**, 075103 (2007).
[30] R. J. Kennedy *et al*., J. Am. Chem. Soc. **124**, 934 (2002).
[31] K. A. Dill and S. Bromberg, *Molecular Driving Forces* (Garland Science, Oxford, 2002).
[32] D. O. V. Alonso and K. A. Dill, Biochem. **30**, 5974 (1991).
[33] W. Li, M. Qin, Z. Tie, and W. Wang, Phys. Rev. E **84**, 041933 (2011).
[34] H. Kaya and H. S. Chan, Proteins: Struct. Funct. Genet. **40**, 637 (2000).