# Protein Folding with Implicit Crowders: A Study of Conformational States Using the Wang−Landau Method

Travis Hoppe* and Jian-Min Yuan

Department of Physics, Drexel University, Philadelphia, Pennsylvania 19104, United States

ABSTRACT: In this paper we introduce the idea of the implicit crowding method to study the statistical mechanical behaviors of folding of β-sheet peptides. Using a simple bead-lattice model, we are able to consider, separately, the conformational entropy involving the bond angles along the backbone and the orientational entropy associated with the dihedral angles. We use a Ising-like model to partially account for the dihedral angle entropy and, implicitly, the hydrogen-bond formations. We also compare our results to recent experiments and find good quantitative agreement on the predicted folded fraction. On the basis of the predictions from the scaled particle theory, we investigate changes in the melting temperature of the protein, suggesting crowding enhanced stability for a variant of trpzip hairpin and a slight instability for the larger β-sheet designed proteins.
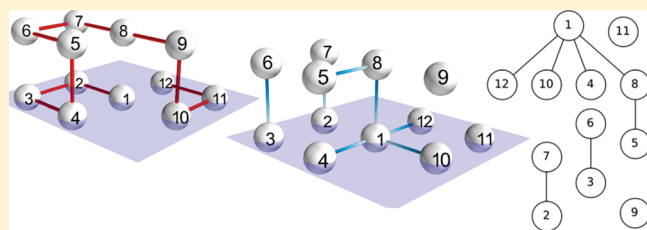
I. INTRODUCTION AND MOTIVATION

The idealization of dilute conditions in conventional in vitro biophysical experiments has long been recognized to ignore the important aspect of crowding. In typical cellular environments, the experimental, computational, and theoretical results have shown crowding agents to affect the stability of proteins and the rates of protein folding. The measured and predicted effects of crowding, however, are varied and seem to be dependent on both the protein and the crowders themselves. For a summary of the recent developments in the field since 2004; see the excellent review by Zhou, Rivas, and Minton.[1] Crowding effects are still actively being explored, with most efforts focusing on the entropic effects to elucidate the response common to all crowding agents. While energetic interactions may exist between the protein and the crowding agents, a simplified yet effective treatment of a crowder is that of a steric, inert particle, affecting the entropy of the protein's conformational state according to its compactness and shape.[2,3]

In one limit, where the crowding particles are much larger in volume and mass than the protein, good results have been obtained by approximating the crowders through localization. Typically, this is modeled as confinement between parallel plates, or in spherical or cylindrical cavities.[4−6] In these approximations, the conformations that are extended beyond the confinement wall are excluded. In the other limit, where the crowders approach the size of a typical residue, the effects of excluded volume dominate. Here, not only are the extended conformal states of the protein highly perturbed but also intermediate states are proportionally less favored to those states that are compact.[7]

In this article, we investigate the effects of crowders on the folding properties of β-sheet proteins using a lattice model. While lattice-based approaches are numerous, those that connect their results to physical experiments are less so. We are motivated by the experiments done by Gai and co-workers[8−10] and examine

our model in light of their observations. We recognize, however, that simple lattice-bead models capture only a portion of the conformational entropy; typically only the gross features of the backbone are correctly modeled. To accommodate for the oriential entropy of the dihedral or $\phi-\psi$ angles, we introduce an Ising-like model. Each bead along the chain is associated with a binary internal state, with an interaction potential requiring the energetic contributions of two beads to have the same state. Since the correct configuration of the dihedral angles is essential for proper hydrogen-bond formations, this extra degree of freedom gives a physically motivated Hamiltonian that implicitly includes the hydrogen-bond contacts. We model the positional entropy of the backbone by projecting it onto a face-centered cubic (fcc) lattice. We use a Go-like Hamiltonian to model the native connections and to ensure the existence of a unique ground state. With the conformational entropy defined we propose a model that attempts to capture the salient aspects of macro-molecular crowding using a detailed density of states (DOS) calculation.

Once the DOS has been determined, we use the results of the scaled particle theory (SPT) to approximate the effects of crowding on several β-sheet proteins. The use of the SPT to study the conformational states of the protein folding process is, of course, not new and has been studied previously; see refs 11−14 and the references cited therein. We model the protein as a right circular cylinder since the native state of β-sheets are naturally disk-like. The crowders, Ficoll 70, are modeled as spherocylinders accounting for their observed elongation.[15,16] Our treatment is unique among the previous studies in that we use the Wang−Landau algorithm to determine density of states for the positional and orientational entropies, separately. This gives us an accurate measurement of the crowding effects across

ACS Publications © XXXX American Chemical Society

A

dx.doi.org/10.1021/jp107809r | J. Phys. Chem. B XXXX, XXX, 000–000

the conformations of the density of states and the ability to compute thermodynamic quantities to high accuracy.

The organization of the article is as follows. In section II we introduce the Hamiltonian and the effective free energies associated with crowding and the dihedral angles. In this section, we explicitly define the cost, both enthalpically and entropically for each conformation. We demonstrate how the density of states can be factored into two terms, greatly speeding up the calculation using the Wang—Landau algorithm. We then describe the experimental results in section III and fit our model to the experimental observations. We use the SPT to determine the effect of crowders on the thermodynamic quantities and discuss the implications. Finally, we use the results to make predictions for future experiments.

II. METHODS

Our protein is coarse grained to a chain of beads and projected as a self-avoiding walk onto a fcc lattice, with a "bead" representing an amino acid residue. The fcc lattice was chosen over the traditional cubic lattice to provide more degrees of freedom. Previous works have found the fcc lattice to be a more natural fit to the secondary structures of α-helices and β-sheets.[17,18] The choice of lattice is not arbitrary, as higher coordination numbers and different symmetries may better represent the underlying structure. For a summary on the effect of lattice choice see Pierri et al.[19]

Two beads are considered nearest-neighbors if they are on adjacent sites on the lattice. With the underlying lattice being defined by a set of primitive vectors $\mathbf{e}$, we define two lattice points $\mathbf{x}_i, \mathbf{x}_j$ to be nearest neighbors if there exists a vector $\mathbf{v} \in \mathbf{e}$ such that $\mathbf{x}_i = \mathbf{x}_j + \mathbf{v}$. For convenience, we define the twelve lattice steps in Cartesian coordinate space (x, y, z) that form the base set of a face-centered cubic lattice

$$\mathbf{e} = l \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & -1 & -1 & -1 & -1 \\ 1 & -1 & 0 & 0 & 1 & 1 & -1 & -1 & 0 & 0 & 1 & -1 \\ 0 & 0 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 0 & 0 \end{bmatrix}^T$$

These twelve vectors define the nearest-neighbors for a given lattice point. Here $l = 3.8$ Å is the length scale of the lattice, which is the average spacing between two C_α atoms. Let the set of all backbone conformations be denoted by $\mathscr{C}$ with the vector $\mathbf{c} \in \mathscr{C}$ representing an individual conformation on the lattice.

Our model Hamiltonian is a modification of the Go model,[20] where the only energetic contributions are either from the attraction of the residues that are predefined native contacts or from the repulsion of the nonnative ones. The Go model, primarily a model of minimal frustration, typically ignores the potential from non-native contacts. Models with the repulsive terms added[21,22] create a frustrated energy landscape since more structural information is encoded in the Hamiltonian. In addition, the high coordination number of the fcc lattice does not always admit a unique native state for some structures without the nonnative term. We let $\mathbf{G}$ be a symmetric matrix of native contacts, $\mathbf{G}_{ij} = 1$ if positions i and j are native contacts of the protein, otherwise $\mathbf{G}_{ij} = 0$.

In addition to the Go-like native contacts we further require that the beads have the correct orientations. To achieve this, each bead has a binary internal state, representing the correct (or incorrect) range of values of its dihedral angles. This is similar to the ideas presented in the Muñoz—Eaton (ME) model[23] where each amino acid is allowed two internal states, folded or unfolded. While the ME model has been solved exactly in a restricted form[24]

and incorporated into more extensive models,[25] we exploit the fact that the ME model generates a density of internal states that is easily decoupled with the positional microstates, thus leading to more precise estimate of the thermodynamic variables. The permutations of these internal states generate an ensemble of microstates; let the set of all such state sequences be denoted $\mathscr{S}$, with the vector $\mathbf{s} \in \mathscr{S}$ representing a particular sequence of the internal states. It will be useful to refer to the total number of folded beads for a state $\sigma = \sum_{i=1}^{L} \mathbf{s}_i$, with the unfolded/folded states defined, indexed by $s_i = 0/s_i = 1$, and amino acid residue count L.

Our Hamiltonian depends on the number of native and non-native contacts of all the beads on the lattice and on their internal state

$$\mathscr{H}(\mathbf{c},\mathbf{s}) = -\sum_{i=1}^{L}\sum_{j=i+2}^{L} \omega_{ij}[J_+\mathbf{s}_i\mathbf{s}_j\mathbf{G}_{ij} - J_-(1-\mathbf{G}_{ij})] \quad (1)$$

where $\omega_{ij} = 1$ if residues i and j are nearest-neighbors on the lattice and J_+, J_- represent the strength of the Go model's native and non-native contacts, respectively. A more intuitive form can be written by counting the number of contributions from the native k_+ and non-native k_- contacts:

$$\begin{aligned} \mathscr{H}(\mathbf{c},\mathbf{s}) &= -J_+k_+ + J_-k_- \\ k_+ &= \sum_{i=1}^{L}\sum_{j=i+2}^{L} \omega_{ij}\mathbf{s}_i\mathbf{s}_j\mathbf{G}_{ij} \\ k_- &= \sum_{i=1}^{L}\sum_{j=i+2}^{L} \omega_{ij}(1-\mathbf{G}_{ij}) \end{aligned} \quad (2)$$

There are two entropic effects incorporated into the free energy, the crowding and the dihedral angle restrictions. The free energy of a state $\mathbf{c}, \mathbf{s}$ is

$$\mathscr{F}(\mathbf{c},\mathbf{s}) = \mathscr{H}(\mathbf{c},\mathbf{s}) - \beta\Delta\psi(\sigma) - \beta\Delta\mu(\mathbf{c}) \quad (3)$$

Here $\beta = 1/k_B T$, $\beta\Delta\psi(\sigma)$ is the free energy term associated with the entropy of the dihedral angle orientation, and $\beta\Delta\mu(\mathbf{c})$ is an entropic cost of inserting the protein into a solution of crowders (both terms to be defined in later sections). When the crowders are implicitly modeled as hard-particles, there is only entropic cost for insertion; the term is truly a free energy contribution. If, however, the crowder specifically interacts with the protein, this contribution must be included in the Hamiltonian. The model admits three fitting parameters (J_+, J_-, h), with h setting the energy scale of the dihedral angle term $\beta\Delta\psi(\sigma)$. Additionally, the crowding term, $\beta\Delta\mu(\mathbf{c})$, depends on the concentration and the geometry of the crowders.

The positional conformation $\mathbf{c}$ determines the number of non-native contacts k_-; thus we take $\Omega(\mathbf{c},\sigma,k_+)$ as the density of states. The partition function can be factored by summing up to the maximum number of native contacts k_+^*:

$$\mathscr{Z} = \sum_{\mathbf{c}\in\mathscr{C}}\sum_{\mathbf{s}\in\mathscr{S}} e^{-\beta\mathscr{F}(\mathbf{c},\mathbf{s})}$$

$$= \sum_{\mathbf{c}\in\mathscr{C}}\sum_{\sigma=0}^{L}\sum_{k_+=0}^{k_+^*}[\Omega(\mathbf{c},\sigma,k_+)e^{\beta\Delta\psi(\sigma)+\beta\Delta\mu(\mathbf{c})}\sum_{\mathbf{s}\in\mathscr{S},\Sigma s_i=\sigma} e^{-\beta\mathscr{H}(\mathbf{c},\mathbf{s})}]$$

$$(4)$$

A. Conformational Entropy of Dihedral Angles.
Associating an entropic cost with the correct dihedral angles is an idea that

B

dx.doi.org/10.1021/jp107809r |*J. Phys. Chem. B* XXXX, XXX, 000–000

goes back to the original Zimm—Bragg (ZB)[26] and Lifson—Roig (LR) models.[27] These models were first used for helix—coil transitions and later extended to include sheets.[28−30] Letting each bead have an internal state, native or non-native, allows us to capture some of the detail present in more complex models yet still retain the simplicity inherent in lattice models. It is the lack of spatial degrees of freedom that separate the ZR, LR type models from the one presented here. In our model each conformation defines a new Ising-like subproblem, where we consider the entropy associated with the ensemble of "spins" of only the nearest neighbor contacts. Our model is actually the generalized variant of the spin systems, commonly referred to as the Potts model. There are two major distinctions between the Ising and Potts models; the spin directions are not necessarily restricted to two states and the strength of spins in contact are determined by an interaction matrix. We still retain the two-state model, folded/unfolded, but our interaction matrix has only a single nonzero term, contributing only when both spins are in the folded state. This is evident by the $s_i s_j$ term in the Hamiltonian since $s_i s_j \in \{0, 1\}$. However, each folded residue comes at a price; the entropic cost of restricted dihedral angles for a single bead is

$$\beta \Delta \psi(\sigma) = h \sum_{i=1}^{L} s_i = h\sigma \qquad (5)$$

B. Reduction of State Space - Conformational Decoupling.
Since the Hamiltonian, and thus the free-energy, is a function of both positional conformations c, and the orientational conformations σ, k_+, it would seem that a calculation of the full density of states $\Omega(c,\sigma,k_+)$, is necessary. We will show, however, that given a positional conformation, one can decouple the internal states by grouping similar conformations into isomorphic macrostates. Consider the matrix

$$\mathbf{X}_{ij}(c) = \mathbf{G}_{ij}\omega_{ij}(c) \qquad (6)$$

of the positive energetic contributions to the Hamiltonian when one ignores the internal states. We can map this symmetric matrix to a simple, but possibly disjoint graph, $\mathbf{X}_{ij} \rightarrow g(c)$, by observing that $\mathbf{X}$ is an adjacency matrix. Let $\mathrm{Aut}(g(c))$ define the automorphism group to which c belongs. Each $\mathrm{Aut}(g(c))$ is a permutation group, whose members are related by mapping the vertices of the graph onto itself through permutation such that the resulting graph is isomorphic to the original. While all graphs in the same automorphism group are isomorphic to each other, we should note that not every graph is physically realizable in our lattice model due to the restriction that no two beads can occupy the same lattice site. The key to decomposing the density of states, however, is the grouping of the graphs, and hence the conformations. Each automorphism group defines a finite-graph for a spin-state system. The density of states for this finite-graph system is calculated by considering, over all possible "spins" of the system, the restricted counts of σ and k_+. The internal states are identical for all members of a particular automorphism group, thus we can decouple the density of states as

$$\Omega(c,\sigma,k_+) = \Omega_1(c)\,\Omega_2(\sigma,k_+;X(c)) \qquad (7)$$

As an illustrative example of the decoupling method, consider a 12-bead homopolymer defined over a cubic lattice where every connection is favorable ($\mathbf{G}_{ij} = 1$ if $|i - j| > 1$) in the particular conformation shown in Figure 1a. All of the native connections,
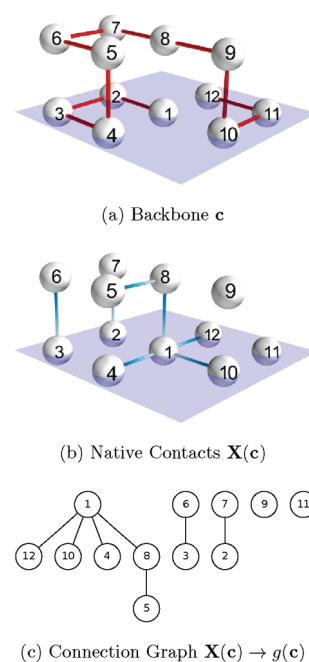

(a) Backbone **c**

(b) Native Contacts **X(c)**

(c) Connection Graph **X(c)** → g(c)

Figure 1. Sample homopolymer (with all connections favorable) on a cubic lattice with (a) the backbone and (b) the energetically favorable connections. The problem of finding the density of states for the internal conformations of each bead for this conformation is then reduced to solving the density of states of the Potts model over the graph shown in (c). Note that the labels in (c) are shown only to guide the eye, all valid permutations of the indices belong to Aut($g(c)$) and hence define the same Potts subproblem. The **G** model in the Hamiltonian would make some of the connections in **X(c)** unfavorable, further simplifying the problem. Additionally, our model is defined over a fcc lattice, while the example shown above is a cubic lattice for illustrative purposes.

which in this case are simply all nearest neighbors, are shown in Figure 1b. Abstracting the representation to a graph in Figure 1c shows the finite system that solves the density of states over the Potts model. Usually, when a Potts/Ising-type system is solved, the underlying graph has a high degree of symmetry (cubic lattices or Cayley trees are common examples). For a typical graph produced by our model, this symmetry is broken, forcing us to numerically compute $\Omega_2(\sigma,k_+;X(c))$. However, the number of edges in the graph determine the maximum number of favorable connections. Since this number is small, the convergence of the Wang—Landau algorithm on this portion of the DOS is rapid.

In general, grouping a particular c to its automorphism group requires solving the graph isomorphism problem multiple times. While specialized algorithms exist,[31] the computational solution is unique in its complexity class and lacks a simple invariant that definitively determines isomorphism.[32] The problem is greatly simplified if the energetic matrix **G** permits only a few, highly degenerate sets of graphs. The Hamiltonian defined for β-sheets happens to be one of these favorable partitions. Since the β-sheet structure is essentially planar, moving perpendicular to the strand direction identifies a column of connections. When abstracted to a graph, the only structure possible is that of a linear chain (unary tree), whose length is limited by the number of strands. The set of conformations $\mathcal{C}$ for β-sheets create graph structures that have a high degeneracy and consequently low cardinality in the set of unique automorphism groups. Additionally, checking for graph isomorphism of linear chains is trivial since these graphs can be

uniquely determined by a single number, the chain length. These two facts combined significantly speed the computation of the decoupled density of states $\Omega_1(\mathbf{c})\,\Omega_2(\sigma,k_+;\mathbf{X}(\mathbf{c}))$.

C. Implicit Crowding Effects. We study the effects of crowders to our protein system using the scaled particle theory. In one of the original formulations of SPT,[33] one calculates the work done to expand a spherical cavity of radius R in a hard sphere fluid of radius r_c. The centers of the fluid particles are excluded from the cavity region, which implies that meaningful values of R must be greater than $-r_c$. This work $W(R)$ is the configurational part of the chemical potential for a single solute particle $\Delta\mu(\mathbf{c}(R))$. The relationship between the probability of any particular configuration and the work required to create it is $P(R) = \exp(-\beta\Delta W(R))$, where $P(R)$ is the probability that there are no centers of fluid particles of radius r_c in the spherical region $(R + r_c)$. If we approximate both our protein and crowders as hard-spheres, we can calculate the work required to create the proper cavity. One first calculates the probability of finding a cavity whose size can accommodate only a single spherical solute particle. This yields

$$W(-r_c \leq R \leq 0) = -kT \ln(1 - (4/3)\pi\eta(R + r_c)^3) \quad (8)$$

where η is the number density of the solvent. One expands $W(R)$ around $R = 0$ to the second order (noting that the leading term for large R must be the pressure–volume term $^4/_3\pi p R^3$ with p as the pressure of the fluid) and obtains

$$\beta\Delta\mu(\mathbf{c}(R)) = \beta\big(\Delta\mu(\mathbf{c}(0)) + \Delta\mu'(\mathbf{c}(0))R$$
$$+ \frac{1}{2}\,\Delta\mu''(\mathbf{c}(0))R^2 + \frac{4}{3}\,\pi p R^3\big) \quad (9)$$

These terms can be computed by the continuity of $W(R)$ and its first two derivatives at $R = 0$. The pressure can be found by substituting in the exact solution of the Percus–Yevick equation, yielding a density approximation as a function of the packing fraction ϕ. This density route is not unique among thermodynamic pathways. Expressions have been worked out for both compressibility and viral routes.[34] Each of these pathways amount to a smoothing in the structural information of the fluid as one "turns-on" the density field. The compressibility and viral routes tend to yield better approximations to the solvation free energy, giving

$$(\beta\Delta\mu)_{\text{spherical}} = \frac{\phi(-2 + 7\phi - 11\phi^2)}{2(1-\phi)^3} - \ln(1 - \phi)$$
$$+ \frac{18\phi^3}{(1-\phi)^3}\left(\frac{R}{2r_c}\right) - \frac{18\phi^2(1+\phi)}{(1-\phi)^3}\left(\frac{R}{2r_c}\right)^2$$
$$+ \frac{8\phi(1+\phi+\phi^2)}{(1-\phi)^3}\left(\frac{R}{2r_c}\right)^3 \quad (10)$$

The above treatment by SPT assumes, however, that the cavity created is spherical, a condition that is not rigorously satisfied for the crowders nor the proteins examined in this study. The native states of the proteins studied in this work can be well approximated by a right circular cylinder, as the β-sheet structures are disk-like. Additionally our crowder, Ficoll 70, is known to have an elongated shape[15] and recent predictions model them as spherocylinders with diameters of 28 Å and an end-to-end length of 184 Å.[16]

The extension of SPT to work with aspherical mixtures can be expressed through the activity coefficient.[35,36]

The activity coefficient γ_i is the measure of the deviation of the ith species at the actual composition of the solution from the chemical potential of an ideal solution as given by the equation

$$RT \ln \gamma_i \equiv \mu_i - \mu_i^0 \quad (11)$$

where μ_i^0 is the chemical potential of a reference state. For hard-convex particles the nonideality of a particular species of interest can be found by computing an expression as a function of the volume V_i, surface area S_i and the Kihara support function H_i[37] of that species, given by

$$\ln \gamma_i = -\ln(1 - \langle V\rangle) + \frac{H_i\langle S\rangle + S_i\langle H\rangle + V_i\langle 1\rangle}{1 - \langle V\rangle}$$
$$+ \frac{H_i^2\langle S\rangle^2 + 2V_i\langle H\rangle\langle S\rangle}{2(1-\langle V\rangle)^2} + \frac{V_i\langle H^2\rangle\langle S\rangle^2}{3(1-\langle V\rangle)^3} \quad (12)$$

where $\langle X\rangle \equiv \Sigma\rho_i X_i$ and $\langle 1\rangle \equiv \Sigma\rho_i$ are the averages over the different species, with ρ_i as the number density of that species. For a right circular cylinder and spherocylinder, respectively, the Kihara support functions are $H_{\text{sphereocylin}} = r\pi/4 + L/2$ and $H_{\text{cylinder}} = r + L/4$ with r as the radius and L as the length of the cylindrical section.

In the process of calculating the density of states, we can sample the conformations to determine the parameters for the activity coefficient. For each conformation of the peptide we calculate a best fit circular cylinder by pointing the axis of the cylinder along the largest principal axis of the bead positions then scale the radius and length so all beads fit inside. Using eq 11 and 12, we can determine the free energy due to crowders in our system.

D. The Wang–Landau Density of States Method. Wang–Landau (WL) sampling[38] is a generic algorithm to calculate the relative density of states (DOS) for a given system. The algorithm starts off with the initial a priori ansatz that all conformational states are equally likely $\Omega(\zeta) = 1$, where ζ is a conformation of the system. Traditionally, the calculation of the density of states was computed as a function of the energy of the system. Like others,[39,40] we use the Wang–Landau method to determine the density of states for a *conformation* of the system rather than the energy explicitly. Each conformation is still directly related to a numerical energy. By calculating the DOS for conformations, we can delay the calculation of the energy. This has the advantage that multiple simulations are not required for each set of the system parameters (J_+, J_-, h, ϕ, r_c).

In the WL method, the density of states is iteratively refined, crudely at first to ensure a large sampling, and then with greater precision as Ω converges. Similar to a typical Monte Carlo simulation, the algorithm has an acceptance rate. However, unlike traditional Metropolis–Hastings simulations, the process can not be modeled as a Markov chain since the transition matrix itself is iteratively refined. The WL acceptance rate is

$$P(\zeta_A \rightarrow \zeta_B) = \min\left(1, \frac{\Omega(\zeta_A)}{\Omega(\zeta_B)}\frac{n_{B\rightarrow A}/n_B}{n_{A\rightarrow B}/n_A}\right) \quad (13)$$

where n_A is the number of outgoing moves from states A and $n_{A\rightarrow B}$ is the number of moves from A to B (similarly defined for n_B and $n_{B\rightarrow A}$). If the moves are reversible, then $n_{B\rightarrow A}/n_{A\rightarrow B} = 1$. These factors are necessary for detailed balance if the move set

chosen has a variable number of moves from each state. Once any particular state had been selected, the density of states is modified by $\Omega(\zeta) \rightarrow f\Omega(\zeta)$, where f is a constant that is slowly reduced to unity during the simulation. All of the computations carried out in this paper took $f_0 = e^1 \approx 2.71828$, $f_{\text{final}} = e^{-9}$, and $f_{i+1} = (f_i)^{1/2}$. Each time a conformation was selected, the DOS is updated along with a histogram of visits for that conformation $H(\zeta)$. The factor f was reduced when $H(\zeta)$ was no less than 90% of the average number of visits for all conformations $\langle H(\zeta) \rangle$. Once the factor f had been reduced, we reset the histogram of visits for all conformations, $H(\zeta) = 0$, and began the process again.

Our move set consists of pull moves, which were first defined over a cubic lattice[41] and later for triangular lattice models.[42] Pull moves are an ergodic, reversible move set that modify the positional conformations by moving the chain along a path defined by a pair of beads adjacent in chain sequence $(i, i + 1)$. Since the number of moves are finite and easily computable, we can quickly determine the factors necessary for detailed balance (i.e., n_A, $n_{A \rightarrow B}$, etc).

When converged, the WL method gives a flat histogram. That is, the averaged fraction of time spent at each macrostate approaches the same constant. Here we define a macrostate as the set of all conformations with the same energy level. Not every microstate is visited during the simulation, nor would it be possible due to the exponential growth in the DOS as a function of chain length. We assume that the visits to each state are ergodic; subsequent visits to a macrostate will visit each conformation an equal number of times over long averages. This idea is reasonable when we consider detailed balance is obeyed for the Monte Carlo simulation and is employed by Wust and Landau.[40] We exploit this observation to determine a probability distribution for a second observable as a function of the first. For instance, we can step through the conformations to compute the probability distribution of the activity coefficient for the protein in a particular conformation **c**. Doing so prevents the need for a multiplicative increase in the density of states (thus speeding up the convergence of the WL algorithm), yet it still provides us with a reasonable estimate of an extended DOS $\Omega(\mathbf{c}, r_g)$.

III. RESULTS

A. Model Calibration. Unlike α-helices, the β-sheet motif has been difficult to study experimentally due to its propensity to aggregate. Recently there has been a spate of designed peptides that exhibit the β-sheet motif, albeit with extremely broad thermal transitions.[43] We consider and describe below, three experimentally designed β-sheets peptides used in this study. These designed proteins were specifically chosen to test the model against their readily available experimental measurements.

The first peptide (sequence: RFSEV$^{\text{D}}$[PG]KKFITS$^{\text{D}}$[PG]-KTYTEV$^{\text{D}}$[PG]KKILQ, nicknamed $^{\text{D}}$P$^{\text{D}}$P$^{\text{D}}$P) is a 28-residue chain with a natural four-strand β-sheet structure. This designed peptide was studied experimentally by Xu et al.[8] as an extension of the peptide $^{\text{D}}$P$^{\text{D}}$P-II first proposed by Gellman and co-workers.[44] A schematic model of the native state for the peptide is shown in Figure 2a. The second peptide (sequence: RFIEV$^{\text{D}}$-[PG]KKFITS$^{\text{D}}$[PG]KTYTE, nicknamed $^{\text{D}}$P$^{\text{D}}$P) is a 20-residue chain with a natural three-strand β-sheet secondary structure.[45] The final peptide (sequence: GEWTWAD[AT]KTWTWTE, nicknamed trpzip4-m1) is a 16-residue variant of the tryptophan zippers studied by Cochran et al.[43] and later by Du et al.[10] Compared to the designed peptides, the tryptophan zippers have
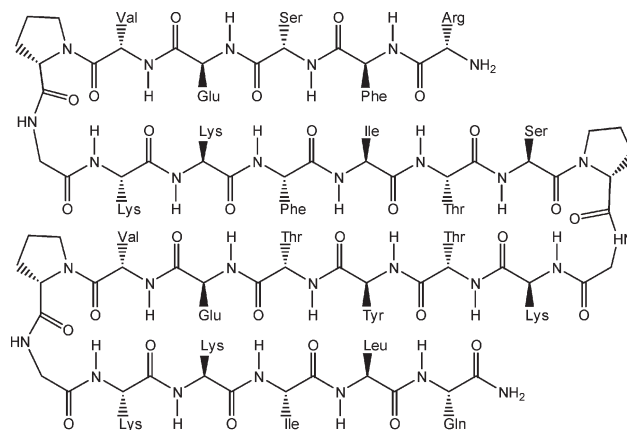
(a) Chemical Representation

Figure 2. Native state of the peptide $^{\text{D}}$P$^{\text{D}}$P$^{\text{D}}$P represented schematically. The Pro-Gly amino acid residues at each end are combined to a single bead to allow for the proper turn structure on the fcc lattice.

significantly higher stabilities (with a difference of approximately $\Delta G_{\text{unf}} \approx 1.0$ kcal mol^{-1} at 298 K).

An artifact of the fcc lattice forces the β-hairpin to be made over an odd number of lattice sites; thus we replace the Pro-Gly residues in the first two designed proteins and the Ala-Thr residues in the third peptide trpzip4-m1, with a combined residue (denoted with a square bracket, e.g., [PG]). The experimental measurements on these peptides have been carried out using temperature jump experiments, for details see refs 8 and 10.

We use the WL method to calculate the conformations of the positional $\Omega_1(\mathbf{c})$ and orientational $\Omega_2(\sigma, k_+; \mathbf{X}(\mathbf{c}))$ density of states. The model is calibrated by fitting J_+, J_-, and h to the data of three experiments.[8−10] We calculate the fraction of β-sheet contacts, observable from the experiments, by taking the expectation of $\langle \alpha\Theta(\alpha) \rangle$. Here $\langle \bullet \rangle$ is the standard Boltzmann average, Θ is the Heaviside step function, and $\alpha(\mathbf{c})$ measures how close the protein is to its native state. Since the experimental data measure the fraction of β-sheet contacts (inferred from a circular dichroism measurement), we calibrated our model with physically similar observable,

$$\alpha(\mathbf{c}) = (k_+ - k_-) / k_+^* \tag{14}$$

This fraction of β-sheet contacts was used to calibrate the three free parameters. We note that without the non-native term J_- in the Hamiltonian in eq 2 the fraction of sheet contacts would be, as usual, $\langle k_+/k_+^* \rangle$. In Figure 3 we show the fits of the two proteins trpzip4-m1 and $^{\text{D}}$P$^{\text{D}}$P with the fitting parameters given in Table 1. The fits are quite good, encouraging us to make predictions about the system behavior as a function of crowding packing fraction. It is worth noting that the three- and four-stranded designed β-sheets had the best fits with $J_- = 0$, implying that additional stabilization provided by the term was needed only to model trpzip4-m1. This may not be surprising when the larger melting points and the broad thermal transitions of the designed proteins are considered versus those of the smaller β-hairpin peptide (listed in Table 2).

B. Effects of Crowders. In this section we use the model defined above to study the effects of crowders on peptide/protein structures and stability. For one of the peptides we have the experimental results on crowding effects to compare with. In
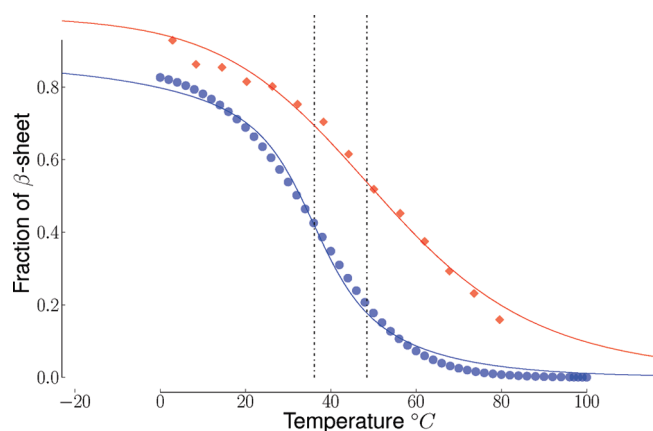
Figure 3. Experimental data of fraction folded versus temperature for trpzip4-m1 (blue circles) and the three stranded β-sheet ^{D}P^{D}P (red diamonds). Model fits are shown with dashed lines of the same color. The thin (black) vertical lines are shown to mark the critical temperatures at 36.1 and 48.5 °C for three-stranded and trpzip4-m1 peptides, respectively. The fit for the four-stranded β-sheet ^{D}P^{D}P^{D}P is similar to the three strand and is not shown for clarity.

Table 1. Fit Parameters of the Free Energy (eq 3) to the Model for Each Peptide[a]

	h	J_+	J_-
trpzip4-m1	9.396	6526.8	4691.6
three-stranded	5.103	2228.8	0.0
four-stranded	4.319	1751.3	0.0

[a] Here J_+, J_- are the strengths of the native and nonnative bonds and h sets the energy scale of the dihedral angle term.

the paper by Mukherjee et al.[9] a significant change (approximately 12 °C) in the melting point of trpzip4-m1 was observed under crowded conditions. The crowder chosen for this experiment was Ficoll 70 (F70) at a concentration of 200 mg/mL. F70 is a compact, highly cross-linked branched copolymer of sucrose and epichlorohydrin with an average molecular weight of 70 000.[46] At 200 and 300 mg/mL the packing fractions are approximately $\phi = 0.13$ and $\phi = 0.20$, respectively.[47,48] We study the effects of crowders by considering the specific heat, $C_V(T) = \beta^2(\langle E^2 \rangle - \langle E \rangle^2)$ and note that in all cases, we observe only a single maxima. We identify this maxima as the melting temperature T_c (alternatively, $(\partial C_V/\partial T)|_{T_c} = 0$).

Heat capacity as a function of temperature for trpzip4-m1 is shown in Figure 4 while the melting points for all peptides are listed in Table 2. As expected, trpzip4-m1 displays crowding enhanced stability with the change of critical temperature $\Delta T_c = [1.03, 1.65]$ °C at packing fractions $\phi = [0.13, 0.20]$, respectively. However, the three- and four-stranded β-sheets exhibit a slight decrease in their critical temperatures with ϕ, indicating an entropically based *instability* caused by the crowders.

The native state for the three- and four-stranded β-sheets are highly aspherical. When the entropic effects of crowders are considered, the system prefers compact conformations that minimize the excluded volume effect. As a consequence of this, the native state ceases to be the minimum free energy conformation at large enough ϕ. Crowding-induced conformational change of the native state has been observed experimentally in a recent work by Dhar et al.[48] who studied phosphoglycerate kinase (PGK) with the same crowders as our simulations (Ficoll 70). In this study, the

Table 2. List of the Experimental and Model Melting Points (°C) for Each Peptide[a]

	experimental	$\phi = 0$	$\phi = 0.13$	$\phi = 0.20$
trpzip4-m1	32.1 ± 0.9	36.12	37.15	37.76
three-stranded	52.6 ± 0.4	48.49	48.23	48.10
four-stranded	50.5 ± 0.8	49.18	49.01	48.94

[a] The experimental melting points are taken from refs 8–10 in a dilute solution without crowders. The calculated melting points from the model are given at the listed values of the packing fraction. Not shown is the experimental value of trpzip4-m1 in the Ficoll 70 solution of 200 mg/mL with $T_C = 44.0 \pm 0.2$ °C.
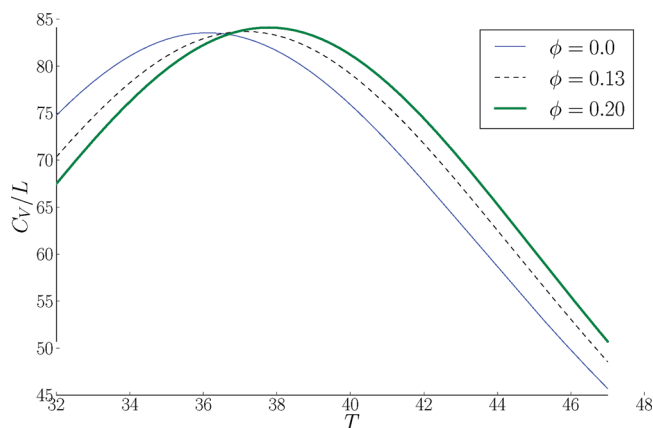

Figure 4. Specific heat per residue count for the protein trpzip4-m1 in the presence of crowders.

conformational states were changed dramatically with crowders; an optimal nonzero packing fraction of crowders was found to increase the protein's activity. In our simulation, there was no shift to a new distinct native state at higher crowding concentrations. Rather, we observed a gradual shift toward more compact conformations at the expense of breaking energetically favored bonds, a general collapse of the β-sheet. As an example of the native conformation, which is enthalpically favored, versus an entropically favored one (Figure 5). Previous studies that showed crowding enhanced stability often dealt with globular wild-type proteins, whose natural environment required them to operate in crowded conditions. In contrast to the PGK study, the two larger peptides in our study were not wild-type, rather they were designed and studied because of the fact that they folded into β-like conformations at realistic temperatures without aggregation. This suggests further experimentation on the designed peptides to determine if the destabilization of the native state against those of the unfolded and intermediate states under crowded conditions can be observed experimentally.

To assess the effect on the conformational states, we examine the Boltzmann averaged excess chemical potential from the native state as a function of temperature and ϕ

$$\beta \langle \Delta \mu_{\text{ex}}(T) \rangle = \langle \ln \gamma_i - \ln \gamma_N \rangle \qquad (15)$$

Figure 6 of this free energy term for trpzip4-m1 illuminates several interesting structural features from an ensemble perspective. At large temperatures we see that this excess chemical potential approaches a constant, proportional to the change in the unfolded states due to the crowders. Conversely, at very low temperatures crowders have no effect on the only viable con-
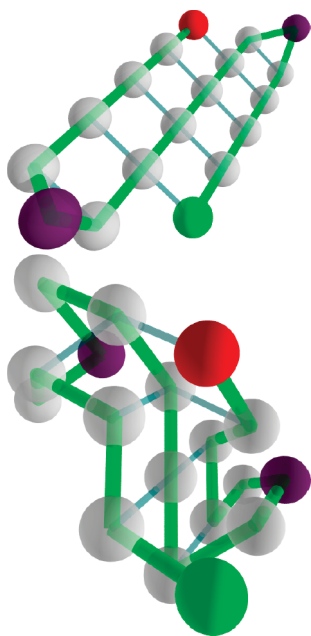
Figure 5. Example of native state (top) and intermediate state (bottom) of the three-stranded peptide. The top state has ten bonds (shown as thin blue lines) while the bottom has eight bonds, making the native state favored energetically. However, the ratio of activity coefficients $\ln \gamma_1 / \ln \gamma_2$ is 1.13 at 200 mg/mL ($\phi = 0.13$) and favors the eight-bond structure due to the entropic crowding effects. The C and N terminus marked with red and green beads, respectively, and the combined Pro-Gly amino acid residues are purple beads.
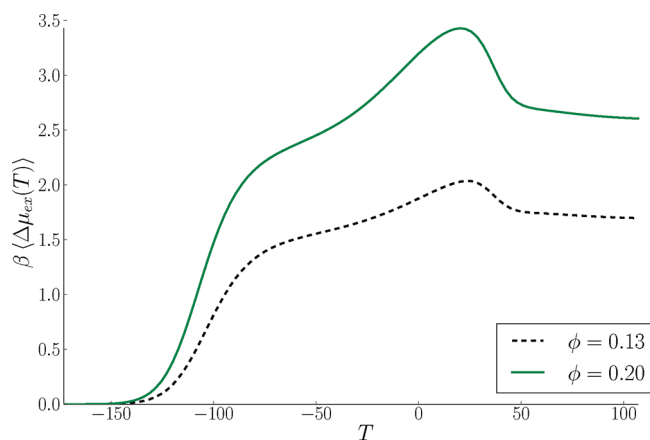

Figure 6. Excess chemical potential (defined in eq 15) for the protein trpzip4-m1 in the presence of crowders.

formation, the native state. Near the folding transition temperature, the effect is large and nonlinear.

IV. CONCLUSION

In this paper we developed a coarse-graining model for proteins by combining the Ising-like state information of the dihedral angles of the modeled β-sheet structures with a fcc lattice model. The density of states has a favorable decoupling that enabled us to use the Wang–Landau method to determine the partition function accurately, giving excellent quantitative agreement with previous in vitro experiments. Using our model and

the predictions of SPT, we found crowding-induced stability, in qualitative agreement with experiment for the smaller peptide, trpzip4-m1. The effect predicted by this model showed a modest change of about $\approx$1 °C, in contrast to the large change observed in the experiments of Mukherjee.[9] We note, however, that these coarse-grained models are approximations and selectively ignore various interactions. The study presented here is an entropic one. If the crowders have enthalpic interactions with the peptide, then these predicted effects will be incomplete. We attribute this underestimation to effects that cannot be explained by excluded volume effects alone.

We found that that the model predicted instability for the designed three- and four-stranded β-sheet peptides. This is consistent with the observation that their native state does not minimize excluded volume effects (it is disk-like rather then globular). This observation alone, however, is not sufficient to predict of crowding based instability. Even if the native state is nonideal, one has to consider the entire ensemble of states as a whole. This was possible using the Wang–Landau method, which allowed us to accurately determine the density of states under the constraints of our model.

The extension of the Go-like contact map to a finite graph presented here is not limited to the β-sheet motif. This entropic model of conformational states can be extended to α-helices or a mixture of secondary structures, as long as the contact graph structure can be decomposed into simple degenerate forms.

■ AUTHOR INFORMATION

Corresponding Author
*E-mail: hoppe@drexel.edu.

■ ACKNOWLEDGMENT

■ REFERENCES

(1) Zhou, H.; Rivas, G.; Minton, A. P. *Annu. Rev. Biophys.* **2008**, *37*, 375.

(2) Hoppe, T.; Yuan, J. *Phys. Rev. E* **2009**, *80*, 011404.

(3) Homouz, D.; Perham, M.; Samiotakis, A.; Cheung, M. S.; Wittung-Stafshede, P. *Proc. Natl. Acad. Sci.* **2008**, *105*, 11754.

(4) Mittal, J.; Best, R. B. *Proc. Natl. Acad. Sci.* **2008**, *105*, 20233.

(5) Wang, W.; Xu, W.; Levy, Y.; Trizac, E.; Wolynes, P. G. *Proc. Natl. Acad. Sci.* **2009**, *106*, 5517.

(6) Zhou, H.; Dill, K. A. *Biochemistry* **2001**, *40*, 11289.

(7) Minton, A. P. *Biophys. J.* **2000**, *78*, 101.

(8) Xu, Y.; Bunagan, M. R.; Tang, J.; Gai, F. *Biochemistry* **2008**, *47*, 2064.

(9) Mukherjee, S.; Waegele, M. M.; Chowdhury, P.; Guo, L.; Gai, F. *J. Mol. Biol.* **2009**, *393*, 227.

(10) Du, D.; Tucker, M. J.; Gai, F. *Biochemistry* **2006**, *45*, 2668.

(11) Mittal, J.; Best, R. B. *Biophys. J.* **2010**, *98*, 315.

(12) Ping, G.; Yang, G.; Yuan, J. *Polymer* **2006**, *47*, 2564.

(13) Batra, J.; Xu, K.; Qin, S.; Zhou, H. *Biophys. J.* **2009**, *97*, 906.

(14) Minton, A. P. *Biophys. J.* **2005**, *88*, 971.

(15) Asgeirsson, D.; Venturoli, D.; Fries, E.; Rippe, B.; Rippe, C. *Acta Physiol.* **2007**, *191*, 237.

(16) Fodeke, A. A.; Minton, A. P. *J. Phys. Chem. B* **2010**, *114*, 10876.

(17) Peto, M.; Sen, T. Z.; Jernigan, R. L.; Kloczkowski, A. *J. Chem. Phys.* **2007**, *127*, 044101.

G

dx.doi.org/10.1021/jp107809r |*J. Phys. Chem. B* XXXX, XXX, 000–000

(18) Pokarowski, P.; Kolinski, A.; Skolnick, J. *Biophys. J.* **2003**, *84*, 1518.

(19) Pierri, C. L.; Grassi, A. D.; Turi, A. *Proteins* **2008**, *73*, 351.

(20) Abe, H.; Go, N. *Biopolymers* **1981**, *20*, 1013.

(21) Rathore, T. A. K. IV, N.; de Pablo, J. J. *Biophys. J.* **2008**, *94*, 4473.

(22) Hoang, T. X.; Cieplak, M. *J. Chem. Phys.* **2000**, *112*, 6851.

(23) Muñoz, V.; Eaton, W. A. *Proc. Natl. Acad. Sci.* **1999**, *96*, 11311.

(24) Bruscolini, P.; Pelizzola, A. *Phys. Rev. Lett.* **2002**, *88*, 258101.

(25) Chung, H. S.; Tokmako, A. *Proteins* **2008**, *72*, 488.

(26) Zimm, B. H.; Bragg, J. K. *J. Chem. Phys.* **1959**, *31*, 526.

(27) Lifson, S.; Roig, A. *J. Chem. Phys.* **1961**, *34*, 1963.

(28) Mattice, W. L.; Scheraga, H. A. *Biopolymers* **1984**, *23*, 1701.

(29) Hong, L. *J. Chem. Phys.* **2008**, *129*, 225101.

(30) Schreck, J. S.; Yuan, J. *Phys. Rev. E* **2010**, *81*, 061919.

(31) McKay, B. D. *Congr. Numer.* *30*, 4787 (1981).

(32) Kbler, J.; Schning, U.; , Torn, J. *The graph isomorphism problem: its structural complexity*; Birkhauser Verlag: Basel, Switzerland, 1993.

(33) Lebowitz, J. L.; Helfand, E.; Praestgaard, E. *J. Chem. Phys.* **1965**, *43*, 774.

(34) Chen, Y.; Weeks, J. D. *J. Chem. Phys.* **2003**, *118*, 7944.

(35) Boublk, T. *Mol. Phys.* **1974**, *27*, 1415.

(36) Minton, A. P. *Method Enzymol.* **1998**, *295*, 127.

(37) Kihara, T. *Rev. Mod. Phys.* **1953**, *25*, 831.

(38) Wang, F.; Landau, D. P. *Phys. Rev. E* **2001**, *64*, 056101.

(39) Patel, B. A.; Debenedetti, P. G.; Stillinger, F. H.; Rossky, P. J. *J. Chem. Phys.* **2008**, *128*, 175102.

(40) Wust, T.; Landau, D. P. *Phys. Rev. Lett.* **2009**, *102*, 178101.

(41) Lesh, N.; Mitzenmacher, M.; , Whitesides, S. In *Proceedings of the seventh annual inter- national conference on Research in computational molecular biology*; ACM: Berlin, Germany, 2003; pp 188−195.

(42) Bockenhauer, H.; Ullah, A. D.; Kapsokalivas, L.; , Steinhofel, K. In *Algorithms in Bioinformatics*; Springer: Berlin, 2008; Vol. 5251, pp 369−381.

(43) Cochran, A. G.; Skelton, N. J.; Starovasnik, M. A. *Proc. Natl. Acad. Sci.* **2001**, *98*, 5578.

(44) Syud, F. A.; Stanger, H. E.; Mortell, H. S.; Espinosa, J. F.; Fisk, J. D.; Fry, C. G.; Gellman, S. H. *J. Mol. Biol.* **2003**, *326*, 553.

(45) Hudson, F. M.; Andersen, N. H. *Biopolymers* **2006**, *83*, 424.

(46) Venturoli, D.; Rippe, B. *Am. J. Physiol. Renal. Physiol.* **2005**, *288*, F605.

(47) Lavrenko, P. N.; Mikriukova, O. I.; Okatova, O. V. *Anal. Biochem.* **1987**, *166*, 287.

(48) Dhar, A.; Samiotakis, A.; Ebbinghaus, S.; Nienhaus, L.; Homouz, D.; Gruebele, M.; Cheung, M. S *Proc. Natl. Acad. Sci* **2010**, *107*, 17586.

H

dx.doi.org/10.1021/jp107809r |*J. Phys. Chem. B* XXXX, XXX, 000–000