Density of states for a short overlapping-bead polymer: Clues to a mechanism for helix formation?

James E. Magee,[*] Leo Lue, and Robin A. Curtis

School of Chemical Engineering and Analytical Science, The University of Manchester, P. O. Box 88, Sackville Street,
Manchester M60 1QD, United Kingdom

The densities of states are evaluated for very short chain molecules made up of overlapping monomers, using a model which has previously been shown to produce helical structure. The results of numerical calculations are presented for tetramers and pentamers. We show that these models demonstrate behaviors relevant to the behaviors seen in longer, helix-forming chains, particularly "magic numbers" of the overlap parameter, where the derivatives of the densities of states change discontinuously, and a region of bimodal energy probability distributions, reminiscent of a first-order phase transition in a bulk system.

I. INTRODUCTION

Helices are a common structural motif in biological molecules, from the α-helix in proteins to the culturally iconic double helix observed in double-stranded DNA. In a living cell, the adoption of stable helical structures allows these molecules to place functional groups in specific positions and orientations, and holds the polymer backbone away from the solvent, protecting it from chemical attack. The consensus view of helix formation follows the work of Pauling *et al.* [1]: biological helices are stabilized by orientationally dependent hydrogen bonding, with their chirality arising from the chirality of the polymer molecule.

Those same properties which make helical molecules so useful in living cells also make them useful in the context of nanotechnology. Unfortunately, while our understanding of biological helices is good at explaining why polypeptides and polynucleotides do form helices, it does not provide useful prescriptions for developing alternative helix-forming molecular architectures. To gain the understanding necessary to develop such prescriptions, many workers have considered "reduced models" [2–7] for helix formation, which attempt to capture the underlying physics of the phenomenon in as simple a manner as possible.

Over recent years, simulation studies of such reduced models have yielded surprising results. In particular, several polymer models have been proposed which produce helical structure while interacting via isotropic potentials [2–5]; that is, helix formation without "designed-in" preferred interactions, with spontaneous chiral symmetry breaking. Maritan *et al.* [4,5] have shown that helices are "maximally compact" structures for stringlike objects. This suggests that helix formation arises from geometric symmetry breaking, akin to crystallization. A better understanding of how this symmetry breaking can arise should lead to better prescriptions for helix-forming architectures.

In the study of α-helix formation in polypeptides, the starting point is the observation that helices are quasi-one-dimensional objects, which can be looked at as spin chains.

The standard approach [8–11] is to attribute amino acid residue conformations to spins, either H ("helix," that is, capable of forming a hydrogen bond compatible with a helical structure) or C ("coil," otherwise). A spin chain representation is then made up of these states; in the simplest form [9], residues that are neighbors along the peptide backbone interact according only to their H or C attribution and amino acid type. Modern versions of this approach [8] include many-body "capping interactions," which are nonpairwise, nonlocal interactions between residues; the strength of these interactions, however, still depends only upon the residue type and H or C attribution. Such models have achieved considerable success in helical structure prediction for polypeptides. For more general helix-forming systems, the proper attribution of a backbone segment to H or C type is not clear. However, the success of the spin chain approach to helix formation in polypeptides suggests that a similar approach may be fruitful.

For a linear polymer of spherically symmetric monomers, single monomers are not the equivalent of amino acid residues for helix formation, as they have no internal degrees of freedom. From symmetry arguments, the minimum possible such building block must be a tetramer; helices break chiral symmetry, and a tetramer is the shortest length chain which may exhibit chirality. Similarly, the behavior of a pentamer should contain information on how neighboring chiral centers interact, and so forth for longer chains.

In this paper, we seek complete enumeration of the partition function for tetramers and pentamers, using a simple polymer model which has previously been shown to produce helices [3]. This enumeration is performed using a methodology similar to that followed by Taylor [12] for short tangent square-well chains. The intention is to identify the building blocks necessary for helix formation in longer chains, and the origins of the behaviors which allow helix formation in longer chains. The methodology and results of this enumeration are intended as a staging post for the construction of generic spin chain models of helix formation

The remainder of the paper is structured as follows. In Sec. II, the polymer model that is to be studied is described. In Sec. III, the method by which the partition functions for the model are calculated is given. The results calculated from
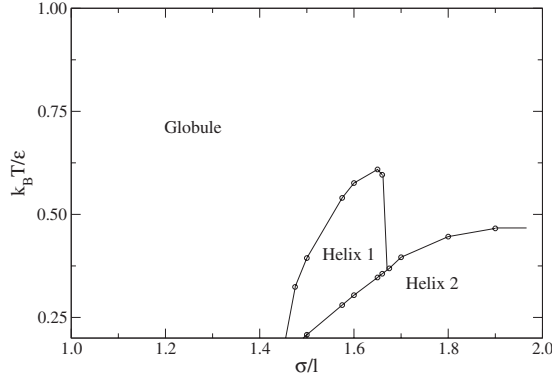
_____
*j.magee@manchester.ac.uk

FIG. 1. Schematic phase diagram from simulation for a helix-forming 20mer, as described in the main text. Reproduced from Ref. [3].

these partition functions are described in Sec. IV. Finally, in Sec. V, these results and their implications are discussed.

II. MODEL

The polymer model consists of a linear chain, bond length l, of N hard spherical monomers with diameter σ. The degree of overlap between monomers is determined by the reduced parameter σ/l. For $\sigma/l=1$, this is the familiar tangent sphere polymer model. We consider chains with $\sigma/l \geq 1$, that is, with overlapping monomers. Interactions between non-bonded monomers (separation r) are given by an isotropic square-well potential

$$u(r) = \begin{cases} \infty, & r \leq \sigma, \\ -\epsilon, & \sigma < r \leq \lambda\sigma, \\ 0, & \lambda\sigma < r, \end{cases} \quad (1)$$

where λ is the well width (taken as 1.5 in this work), and the well depth ϵ sets the energy (and hence temperature) scale. We follow the protein literature, by denoting interactions between particles where $\sigma < r < \lambda\sigma$ as *contacts*, and interactions where $r < \sigma$ as *overlaps*. Interactions between monomers separated by two bonds along the chain are referred to as 1-3 interactions; interactions for monomers separated by three bonds are referred to as 1-4 interactions, and so forth.

In previous simulation work, we have used a version of this model where individual bond lengths were allowed to vary by $\pm 10\%$. It has been suggested that such bond length variation can enhance the ergodicity of a simulation compared to rigid bonds [13]; further, this allows the configurational and momentum parts of the partition function to be factorized. With such bond length fluctuation, the system has been shown to form helices for 20mers (polymers of length $N=20$). The observed phase diagram is shown in Fig. 1; the system is observed to form two distinct helical phases, "helix 1" (stable at higher temperatures, and with a smaller radius) and "helix 2" (stable at lower temperatures, and with a larger radius). The following work does not include such bond flexibility, as the extra degree of freedom per bond would make the problem very much less tractable.
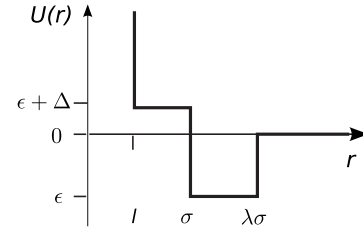

FIG. 2. A core-softened potential. If the shoulder height $\Delta \gg kT$, the effective core diameter will be σ rather than l; hence a chain of monomers interacting via such a potential with bond length l would act like the overlapping square-well monomer model presented here.

Physical relevance

With any such "reduced model," however interesting the behaviors, the question of physical relevance must be answered. The idea of overlapping monomers is consistent with the van der Waals radii of atoms in "realistic" potentials such as CHARMM [14], where atomic radii are often larger than the bond length to neighboring atoms. On a larger scale of approximation, if amino acid residues are approximated by interacting spheres, the radii of gyration for amino acids can be larger than their center of mass spacing along the peptide chain.

Since protein molecules form intrachain hydrogen bonds, an interesting parallel can be made to "core-softened potentials" (see Fig. 2), which have been used to study the anomalous behavior of water [15–17]. These isotropic potentials have a shoulder (diameter σ) around a repulsive core (diameter l), representing close-packed but non-hydrogen-bonded pairs, and an outer well (diameter $\lambda\sigma$) which represents hydrogen-bonding interactions. In a chain of such monomers with bond length l, if the difference between the potential energy of the shoulder and the potential energy in the minimum is sufficiently larger than $k_B T$, the *effective* repulsive core diameter will be the shoulder diameter; at low temperatures, a chain of such monomers would behave like the overlapping square-well monomer model presented here, forming a helical structure.

III. METHODS

We consider four- and five-length polymers of the type described above, as shown in Fig. 3. The position of mono-
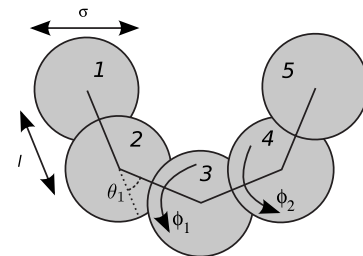

FIG. 3. A cartoon of the model. Five monomers are shown, of diameter σ, bond length l, with bond angle θ_1, and the two dihedral angles ϕ_1 and ϕ_2 indicated.

mer i is denoted by $\mathbf{R}_i$. Bond vectors are defined as $\mathbf{r}_i = \mathbf{R}_{i+1} - \mathbf{R}_i$. Separation between monomers i and j is denoted r_{ij}. The bond angle around monomer i is defined as the angle between bond $\mathbf{r}_{i-1}$ and $\mathbf{r}_i$, that is, $\cos \theta_i = \mathbf{r}_{i-1} \cdot \mathbf{r}_i$. The dihedral (torsional) angle ϕ_i is defined as the angle between the planes formed by the vector pairs $(\mathbf{r}_{i-1}, \mathbf{r}_i)$ and $(\mathbf{r}_i, \mathbf{r}_{i+1})$, relative to the *cis* conformation (i.e., $\phi = 0$ is *cis*, $\phi = \pi$ is *trans*). We take positive ϕ as a right-handed rotation. This is arbitrary, however, as the underlying model is achiral. We do not consider translations and rotations of the entire molecule; as such, we fix the position of the first monomer, as well as the

plane made by the vectors $\mathbf{r}_1$ and $\mathbf{r}_2$. The configurational integral of such an n-mer is given by $\mathcal{Z}_n$, defined as

$$\mathcal{Z}_n = \left(\prod_{i=1}^{n-3} \int_{-\pi}^{\pi} d\phi_i \right) \left(\prod_{j=1}^{n-2} \int_0^{\pi} l^2 \sin \theta_i d\theta_i \right) \exp[-\beta E(\{\mathbf{R}_i\})], \tag{2}$$

where $E(\{\mathbf{R}_i\})$ is the total configurational energy for the system [the sum of the pairwise interactions in Eq. (1)] and $\beta = 1/k_B T$, the inverse temperature.

A. Tetramer

We initially consider a tetramer. The configurational integral is given by

$$\mathcal{Z}_4 = l^4 \int_{-\pi}^{\pi} d\phi_1 \int_0^{\pi} \sin \theta_1 d\theta_1 \int_0^{\pi} \sin \theta_2 d\theta_2 \exp\{-\beta[u(r_{13}) + u(r_{24}) + u(r_{14})]\}. \tag{3}$$

Separations are given by

$$r_{i-1,i+1}^2 = 2l^2(1 + \cos \theta_i) = l^2 x_i \tag{4}$$

and

$$
\begin{aligned}
r_{i-1,i+2}^2(\theta_i, \theta_{i+1}, \phi_i) &= l^2[(1 + \cos \theta_i + \cos \theta_{i+1})^2 + \sin^2 \theta_i + \sin^2 \theta_{i+1} - 2 \sin \theta_i \sin \theta_{i+1} \cos \phi_i] \\
&= l^2[(x_i + x_{i+1} - 2)^2/4 + x_i(4 - x_i)/4 + x_{i+1}(4 - x_{i+1})/4 - \cos \phi_i \sqrt{x_i(4 - x_i)} \times \sqrt{x_{i+1}(4 - x_{i+1})}/2] \\
&= l^2 y_i(x_i, x_{i+1}, \phi_i).
\end{aligned}
\tag{5}
$$

Since we are working with variables of squared separation, for notational convenience we also define $a = (\sigma/l)^2$. Physical bounds for x_i are $a \leq x_i \leq 4$, since a separation of less than σ represents an overlap. It is natural to switch variables in the configurational integral to separations x_i, giving

$$\mathcal{Z}_4 = (1/4) \int_{-\pi}^{\pi} d\phi_1 \int_0^4 dx_1 \int_0^4 dx_2 \exp(-\beta\{u(l\sqrt{x_1}) + u(l\sqrt{x_2}) + u[l\sqrt{y_1(x_1,x_2,\phi_1)}]\}). \tag{6}$$

Since we are working in a square-well system with discretized energies, it is now convenient to switch to a density-of-states representation:

$$\mathcal{Z}_4 = (1/4) \sum_{k=0}^{3} \omega_4(k) \exp(\beta \epsilon k), \tag{7}$$

where $\omega_n(k)$ is the density of states for the n-mer with k contacts. For the tetramer, we can write the appropriate integrals as

$$\omega_4(0) = \int d\phi_1 \int_{\min(4,\lambda^2 a)}^4 dx_1 \int_{\min(4,\lambda^2 a)}^4 dx_2 \Theta(y_1(x_1,x_2,\phi_1) - \lambda^2 a), \tag{8}$$

$$
\begin{aligned}
\omega_4(1) = {} & 2 \int_{-\pi}^{\pi} d\phi_1 \int_{\min(4,\lambda^2 a)}^4 dx_1 \int_a^{\min(4,\lambda^2 a)} dx_2 \Theta(y_1(x_1,x_2,\phi_1) - \lambda^2 a) + \int_{-\pi}^{\pi} d\phi_1 \int_{\min(4,\lambda^2 a)}^4 dx_1 \int_{\min(4,\lambda^2 a)}^4 dx_2 \\
& \times [\Theta(y_1(x_1,x_2,\phi_1) - a) - \Theta(y_1(x_1,x_2,\phi_1) - \lambda^2 a)],
\end{aligned}
\tag{9}
$$

$$
\begin{aligned}
\omega_4(2) = {} & 2 \int_{-\pi}^{\pi} d\phi_1 \int_{\min(4,\lambda^2 a)}^4 dx_1 \int_a^{\min(4,\lambda^2 a)} dx_2 [\Theta(y_1(x_1,x_2,\phi_1) - a) - \Theta(y_1(x_1,x_2,\phi_1) - \lambda^2 a)] \\
& + \int_{-\pi}^{\pi} d\phi_1 \int_a^{\min(4,\lambda^2 a)} dx_1 \int_a^{\min(4,\lambda^2 a)} dx_2 \Theta(y_1(x_1,x_2,\phi_1) - \lambda^2 a),
\end{aligned}
\tag{10}
$$

$$\omega_4(3) = \int_{-\pi}^{\pi} d\phi_1 \int_a^{\min(4,\lambda^2 a)} dx_1 \int_a^{\min(4,\lambda^2 a)} dx_2 [\Theta(y_1(x_1,x_2,\phi_1) - a) - \Theta(y(x_1,x_2,\phi_1) - \lambda^2 a)], \tag{11}$$

where $\Theta(x)$ is the Heaviside step function. The min terms exist to prevent unphysical limits of integration when $\lambda^2 a \geq 4$ (in which case 1-3 contacts are always on). We note that $\omega_4(k)$ is a sum of integrals of the general form

$$f_4 = 2 \int_0^{\pi} d\phi_1 \int_{x_{1l}}^{x_{1h}} dx_1 \int_{x_{2l}}^{x_{2h}} dx_2 \Theta(y_1(x_1,x_2\phi_1) - h), \tag{12}$$

where we have used the symmetry of the system to simplify the ϕ_1 integral. The integrand is nonzero for that region of (x_1,x_2,ϕ_1) space for which $y_1 > h$. We can solve Eq. (5) to find the bound of this space with respect to x_2 (or, by symmetry, x_1) for given (ϕ,h), which we call x_c:

$$x_c(x,\phi,h) = 2\{x[t(4-x) + (h-1)] + \text{sgn}[\cos(\phi_1)]\sqrt{t(4-x)x} \times \sqrt{[2x(2t+h-1) - tx^2 - (h-1)^2]}\}/\{x[x + t(4-x)]\}, \tag{13}$$

where we use $t = \cos^2\phi$, and $\text{sgn}(x)$ returns the sign of x. Similarly, we also solve for the value of ϕ at which $y_i = h$ for given x_i and x_{i+1}, which we denote ϕ_c:

$$\phi_c(x_i, x_{i+1}, h) = \begin{cases} \pi, & \Phi < -1, \\ \arccos(\Phi), & |\Phi| < 1, \\ 0, & \Phi > 1, \end{cases} \tag{14}$$

where the ratio Φ is defined as

$$\Phi = \frac{2 + x_i x_{i+1} - 2h}{\sqrt{x_i(4 - x_i)x_{i+1}(4 - x_{i+1})}}. \tag{15}$$

We consider the shape of this boundary:

$$\left.\frac{\partial \phi_c}{\partial x_i}\right|_{x_{i+1}} = \frac{2\sqrt{x_{i+1}(4 - x_{i+1})}}{\sin \phi_c \sqrt{x_i(4 - x_i)}}\{2(1 - h) - x_i[x_{i+1} + (1 - h)]\}. \tag{16}$$

The only part of this equation which can be negative is the term enclosed in braces. Within the range $a \leq x_i \leq 4$, $a \leq h \leq \lambda^2 a$, $1 \leq a < 4$, and $\lambda^2 > 1$, it can be shown that $(\partial \phi_c / \partial x_i)|_{x_{i+1}}$ is nonpositive. By symmetry, $(\partial x_c / \partial x)|_\phi$ is therefore also nonpositive. As such, we can write Eq. (12) as

$$f_4 = 2 \int_0^{\pi} d\phi_1 \int_{x_{1l}}^{x_{1h}} dx_1 \int_{\min\{x_{2h}, \max[x_{2l}, x_c(x_1,\phi,h)]\}}^{x_{2h}} dx_2. \tag{17}$$

The max term picks the larger of the original lower limit (x_{2l}) and the value of x_2 below which the Heaviside function integrand in Eq. (12) becomes zero. The min prevents the unphysical result of the lower limit becoming larger than the upper limit.

We note that, by symmetry, if $x_1 = x_c(x_2, \phi, h)$, then $x_2 = x_c(x_1, \phi, h)$. Hence, we can immediately see that the solution to $x_{2h} = x_c(x_1, \phi_1, h)$ with respect to x_1 is $x_1 = x_c(x_{2h}, \phi, h)$. For $x_1 < x_c(x_{2h}, \phi, h)$, the upper and lower limits on the innermost integral are equal, and hence the contribution to the integral is zero. Thus,

$$f_4 = 2 \int_0^{\pi} d\phi_1 \int_{\min\{x_{1h}, \max[x_{1l}, x_c(x_{2h},\phi,h)]\}}^{x_{1h}} dx_1$$
$$\times \int_{\max[x_{2l}, x_c(x_1,\phi,h)]}^{x_{2h}} dx_2. \tag{18}$$

Since $x_c(x, \phi, h)$ is a monotonically decreasing function of x across the range of interest, we can now propagate the max term in the middle integral out to the dihedral integral:

$$f_4 = 2 \int_{\phi_c(x_{1l},x_{2h},h)}^{\pi} d\phi_1 \int_{x_{1l}}^{x_{1h}} dx_1 \int_{\max[x_{2l}, x_c(x_1,\phi,h)]}^{x_{2h}} dx_2$$
$$+ 2 \int_0^{\phi_c(x_{1l},x_{2h},h)} d\phi_1 \int_{\min[x_{1h}, x_c(x_{2h},\phi,h)]}^{x_{1h}} dx_1$$
$$\times \int_{\max[x_{2l}, x_c(x_1,\phi,h)]}^{x_{2h}} dx_2. \tag{19}$$

The remaining max and min terms can then be propagated out in a similar fashion, remembering that $\phi_c(x, x', h)$ is a monotonically decreasing function of x and x' across the range of interest:

$$f_4 = 2 \int_{\phi_c(x_{1l},x_{2l})}^{\pi} d\phi_1 \int_{x_{1l}}^{x_{1h}} dx_1 \int_{x_{2l}}^{x_{2h}} dx_2$$
$$+ 2 \int_{\max[\phi_c(x_{1l},x_{2h}),\phi_c(x_{1h},x_{2l})]}^{\phi_c(x_{1l},x_{2l})} d\phi_1 \int_{x_c(x_{2l})}^{x_{1h}} dx_1 \int_{x_{2l}}^{x_{2h}} dx_2$$
$$+ 2 \int_{\max[\phi_c(x_{1l},x_{2h}),\phi_c(x_{1h},x_{2l})]}^{\phi_c(x_{1l},x_{2l})} d\phi_1 \int_{x_{1l}}^{x_c(x_{2l})} dx_1 \int_{x_c}^{x_{2h}} dx_2$$
$$+ 2 \int_{\phi_c(x_{1l},x_{2h})}^{\max[\phi_c(x_{1l},x_{2h}),\phi_c(x_{1h},x_{2l})]} d\phi_1 \int_{x_{1l}}^{x_{1h}} dx_1 \int_{x_c}^{x_{2h}} dx_2$$
$$+ 2 \int_{\min[\phi_c(x_{1l},x_{2h}),\phi_c(x_{1h},x_{2l})]}^{\phi_c(x_{1l},x_{2h})} d\phi_1 \int_{x_c(x_{2l})}^{x_{1h}} dx_1 \int_{x_{2l}}^{x_{2h}} dx_2$$
$$+ 2 \int_{\min[\phi_c(x_{1l},x_{2h}),\phi_c(x_{1h},x_{2l})]}^{\phi_c(x_{1l},x_{2h})} d\phi_1 \int_{x_c(x_{2h})}^{x_c(x_{2l})} dx_1 \int_{x_c}^{x_{2h}} dx_2$$
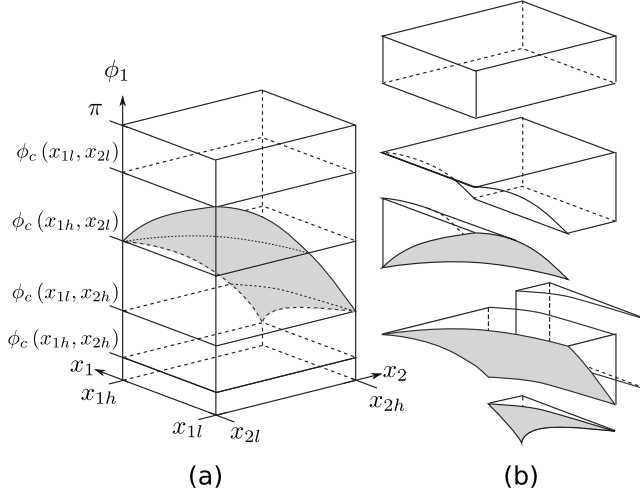
FIG. 4. Graphical interpretation of the integrals in Eq. (20) when $\phi_c(x_{1h}, x_{2l}) > \phi_c(x_{1l}, x_{2h})$. (a) shows the entire space in x_1, x_2, ϕ_1; the total integral is the volume above the gray surface. (b) breaks the volume corresponding to the total integral into the parts listed in Eq. (20).

$$+ 2 \int_{\phi_c(x_{1h}, x_{2h})}^{\min[\phi_c(x_{1l}, x_{2h}), \phi_c(x_{1h}, x_{2l})]} d\phi_1 \int_{x_c(x_{2h})}^{x_{1h}} dx_1 \int_{x_c}^{x_{2h}} dx_2. \tag{20}$$

For notational convenience, superfluous arguments to the functions ϕ_c and x_c have been omitted; that is, h and variables of integration. There are min and max terms in the dihedral integral since, without knowing more about the original limits x_{il} and x_{ih}, it is not possible to tell whether $\phi_c(x_{1h}, x_{2l}, h) < \phi_c(x_{1l}, x_{2h}, l)$. Figure 4 shows a graphical representation of the integral presented in Eq. (20).

The first integral is trivial. The remaining integrals include a term $\int x_c(x_{1l}, \phi_1, h) d\phi_1$; this is analytically tractable, resulting in terms involving elliptic integrals, but is simpler to treat numerically. The final two terms include integrals of the form $\int\int x_c(x_1, \phi_1, h) dx_1 d\phi$, which are not analytically tractable, and are thus treated numerically.

Calculation of the f_4 integrals via Eq. (20) allows calculation of the density of states via Eqs. (8)–(11), from which the equation of state and energy probability distributions can be determined. Structural information, in the form of the dihedral angle probability distributions, is also easily accessible. The dihedral density of states $g_4(\phi_1, k)$ is given by integrals like Eqs. (8)–(11) without the integral over the dihedral angle. For the tetramer, these give tractable though lengthy analytic forms. The probability of observing a given dihedral angle is then given by

$$P(\phi_1; T) = \sum_{k=0}^{4} g_4(\phi_1, k) \exp(\beta\epsilon k) / \mathcal{Z}_4. \tag{21}$$

B. Pentamer

An equivalent procedure may be carried out for a pentamer. We first introduce the 1-5 separation $z_i(x_i, x_{i+1}, x_{i+2}, \phi_i, \phi_{i+1}) = r_{i-1, i+3}^2 / l^2$:

$$z_i = x_i + x_{i+2} - \sqrt{x_{i+1}(4 - x_{i+1})}$$
$$\times [\cos\phi_i x_{i+2} \sqrt{x_i(4 - x_i)} \cos\phi_{i+1} x_i \sqrt{x_{i+2}(4 - x_{i+2})}]/4$$
$$+ (x_{i+1} - 2)[x_i x_{i+2} - \cos\phi_i \cos\phi_{i+1}$$
$$\times \sqrt{x_i(4 - x_i) x_{i+2}(4 - x_{i+2})}]/4 + \sin\phi_i \sin\phi_{i+1}$$
$$\times \sqrt{x_i(4 - x_i) x_{i+2}(4 - x_{i+2})}/2, \tag{22}$$

where the arguments to z_i have been omitted. The partition function for the pentamer is given by

$$\mathcal{Z}_5 = (1/8) \sum_{k=0}^{6} \omega_5(k) \exp(\beta\epsilon k). \tag{23}$$

Equivalent expressions to Eqs. (8)–(11) are simple to construct, using the equivalent form to Eq. (12):

$$f_5 = 2 \int_0^\pi d\phi_1 \int_{-\pi}^\pi d\phi_2 \int_{x_{1l}}^{x_{1h}} dx_1 \int_{x_{2l}}^{x_{2h}} dx_2 \int_{x_{3l}}^{x_{3h}} dx_3$$
$$\times \Theta(y_1 - h_1)\Theta(y_2 - h_2)\Theta(z_1 - h_3), \tag{24}$$

where we have suppressed the arguments of y_1, y_2, and z_1 for notational ease. This integral is constructed (without loss of generality) such that ϕ_1 is always right handed. Explicit bounds of integration due to 1-4 interactions can be treated in the same manner as for the tetramer case. Bounds for the x_1 and x_2 integrals as a function of ϕ_1, and for the x_2 and x_3 integrals as a function of ϕ_2, are determined exactly as in Eq. (20). This leads to single ranges of integration for x_1 and x_3, and two sets of ranges of integration for x_2. The proper range of integration over x_2 is then the overlap of these two ranges. Explicitly treating the bounds of integration due to 1-5 interactions is not trivial, and as such the resulting integral is treated numerically. Dihedral probability distributions $P(\phi_1, \phi_2, T)$ can be calculated from dihedral densities of states $g_5(\phi_1, \phi_2, k)$ in an analogous manner to the tetramer.

IV. RESULTS

Using the results presented in Sec. III, we have evaluated the full partition functions for tetramers and pentamers. Results for the tetramer have been calculated with the MATHEMATICA symbolic algebra package, using Gauss-Kronrod numerical integration. Results for the pentamer have been calculated using ten-point Gauss-Legendre quadrature [18]. Both methods of integration have been checked by comparison against the tangent chain results presented by Taylor [12]. The pentamer results have been verified against short Monte Carlo simulations (data not shown).

A. Tetramer

To validate the method, we compare our calculated densities of states for tetramer tangent square well chains ($\sigma/l = 1$) to those presented by Taylor [12]. These results are shown in Table I. It can be seen that the results are equivalent to four significant figures aside from an unimportant multiplicative factor. The method of Taylor does not use explicit limits of integration, instead numerically integrating the

TABLE I. Comparison of the densities of states for a square-well tetramer chain with $\sigma/l=1.0$ and $\lambda\sigma=1.5$ calculated in this work [$\omega_4(k)/4l^4$, with the factor $1/4$ (as described in Sec. III)] and by Taylor [12] [$\omega_4^{(\text{Taylor})}(k)$, or $g_4^{(k)}$ in the original terminology]. Suppression of the unimportant multiplicative factor of 8π in the work of Taylor leads to the difference in the values; it can be seen that with this factor included the values differ only in the fourth and fifth significant figures.

k	$\omega_4(k)/4l^4$	$\omega_4^{(\text{Taylor})}(k)$	$\omega_4(k)/[32\pi\omega_4^{(\text{Taylor})}(k)]$
0	4.78131	0.19029	0.999750
1	5.59121	0.22247	0.999987
2	2.42528	0.09650	0.999986
3	0.62013	0.02467	1.000170

Heaviside functions in Eq. (12); strictly, the method presented here should be more accurate, though these results suggest the difference is not significant.

The calculated densities of states as a function of σ/l are shown in Fig. 5. The densities of states for the $k=0$ and 1 states are zero for $\sigma/l \gtrsim 4/3$. For overlaps greater than this "magic number," 1-3 interactions become always on—that is, $r_{i-1,i+1} \leqslant \lambda\sigma$ for all values of θ_i with $\lambda=3/2$. This also gives rise to a kink (discontinuity in the derivative) of $\omega_4(2,\sigma/l)$. This is because at $\sigma/l=4/3$, the first term in Eq. (10) (which refers to the density of states for tetramers with a single 1-3 contact and a 1-4 contact) becomes zero, as the limits on the x_1 integral become equal.

As $\sigma/l \rightarrow 2$, the polymer becomes increasingly rigid, and the available conformational space vanishes. The calculated densities of states show the correct behavior at this limit.

Properties calculated from these densities of states are shown in Figs. 6 (energy) and 7 (heat capacity). At $\sigma/l=4/3$, the slopes of the energy and heat capacity contours show discontinuities in their derivatives. As such, the derivatives $(\partial U/\partial\sigma)_T$ and $(\partial C_V/\partial\sigma)_T$ have singularities at $\sigma/l=4/3$; however, these are not physically meaningful response functions. In simulated systems [3], bond lengths are not rigid, and bond length fluctuations will have the effect of smoothing out the discontinuity.

Although the tetramer does not show any other discontinuities, it does show a line of maxima in heat capacity with respect to temperature. We follow Taylor [12] and Zhou *et al.* [13] in ascribing these maxima to collapse of the tetramer into compact conformations. The strength of these maxima
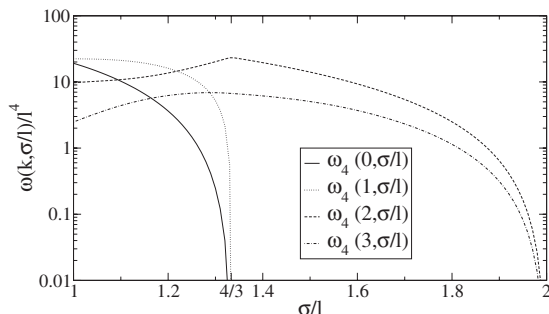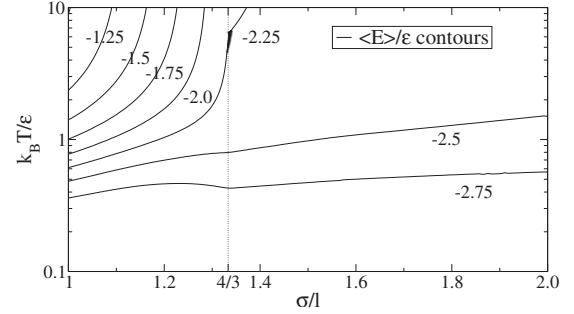

FIG. 6. Ensemble average energies $\langle E\rangle/\epsilon$ for tetramers plotted against σ/l and temperature T. Solid lines show energy contours at the labeled value. Note the discontinuities in the slope of the energy contours at $\sigma/l=4/3$ (shown by the dotted line).

can be seen to decrease with increase in σ/l. Further, the line of maxima shows reentrance with respect to σ/l, with the temperature at which heat capacity is maximal itself having a maximum with respect to σ/l at a point below $\sigma/l=4/3$.

Representative results for the torsional behavior of the tetramer are shown in Fig. 8, where we show the probability $P(\phi_1;T)$ for four values of σ/l. At low values of the overlap ($\sigma/l \lesssim 1.48$), we see maxima in $P(\phi_1;T)$ for nonzero ϕ_1 at all temperatures, with the maxima becoming stronger and moving closer to zero (*cis* conformation) as temperature decreases. For intermediate values of overlap [$1.48 \lesssim \sigma/l < \sqrt{(3+\sqrt{5})/2}$], weak maxima in $P(\phi_1;T)$ are seen for nonzero ϕ_1 at high temperatures; however, the most probable conformation becomes $\phi_1=0$ (*cis* conformation) at low temperature. For large values of overlap [$\sigma/l > \sqrt{(3+\sqrt{5})/2}$], $P(\phi_1;T)$ has only a single maximum at $\phi_1=0$ for all temperatures. The points separating the two regimes $\{\max[P(\phi_1;T)]=0$ and $\neq 0\}$ can be calculated analytically, as the points at which $[\partial P(\phi_1;T)/\partial\phi_1]|_{\phi_1=0}=0$. The calculated line in overlap-temperature space is shown in Fig. 9(a). The upper limit of this line is the value of σ/l at which $\phi_c(x_i=x_{i+1}=h=a)=0$. For values of overlap equal to or larger than this, it is not possible for the polymer to exhibit 1-4 overlaps, and there is no steric hindrance to $\phi_1=0$ states, which are the points of closest 1-4 approach. The lower limit of this region occurs at the point where the maximum of $P(\phi_1;T=0)$ [see
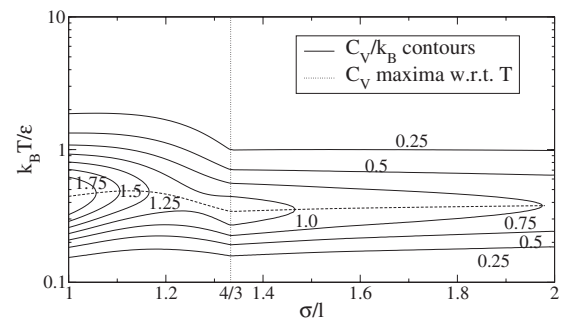

FIG. 5. Densities of states for tetramers plotted against σ/l.

FIG. 7. Configurational heat capacity C_V for tetramers plotted against σ/l and temperature. Solid lines show contours at the labeled value. The dashed line shows the line of maxima in C_V with respect to temperature T. Note the discontinuities in the slope of the heat capacity contours at $\sigma/l=4/3$ (shown by the dotted line).
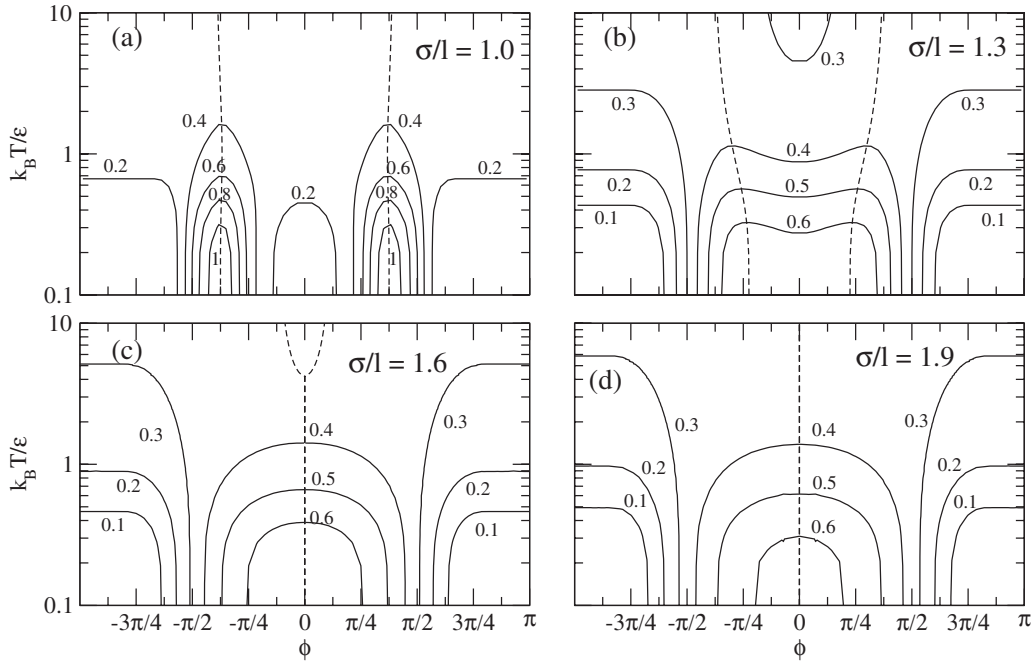
FIG. 8. Contour plots of $P(\phi_1;T)$ for $\sigma/l=$ (a) 1.0, (b) 1.3, (c) 1.6, and (d) 1.9. Solid lines show contours at the labeled value. Dashed lines show maxima in $P(\phi_1;T)$.

Fig. 9(b)] becomes zero. The value of σ/l at this limit does not admit a simple interpretation or expression.

B. Pentamer

A comparison between the densities of states calculated here for pentamer tangent square-well chains with those presented by Taylor is provided in Table II. Results are equivalent to four significant figures. The calculated densities of states as a function of σ/l are shown in Fig. 10. Once again, we see the highest-energy densities of states going to zero at $\sigma/l=4/3$ as 1-3 interactions become always on, combined with a kink in the density of states for the highest remaining energy. All densities of states tend to zero as $\sigma/l \to 2$, where the available conformational space becomes zero. There are
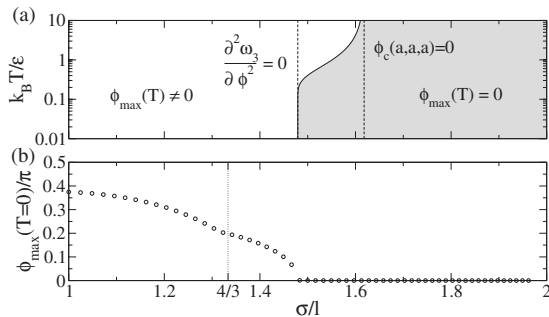

FIG. 9. (a) Behavior of $P(\phi=0;T)$ with respect to σ/l; to the left of the solid line, $P(\phi=0)$ is a minimum, while to the right, $P(\phi=0)$ is a maximum. Directly on the line, $P(\phi=0)$ is a point of inflection. Dashed lines indicate the upper and lower bounds in σ/l of the line. (b) The behavior of the maximum of $P(\phi_1;T=0)$ (that is, in the ground state) with respect to σ/l. There is a kink in the line at $\sigma/l=4/3$, indicated by the dotted line.

two further behaviors, not seen in the tetramer. The most obvious is that the density of the lowest-energy state $\omega_5(6)$ becomes zero at $\sigma/l=\sqrt{7/2}$. For values of overlap larger than this, the pentamer has become so stiff that it cannot bend back on itself far enough to make 1-5 contacts.

A further interesting behavior is observed at intermediate values of σ/l where the ground state $\omega_5(6)$ becomes of the same order of magnitude as $\omega_5(5)$. Indeed, for $1.53 \lesssim \sigma/l \lesssim 1.56$, $\omega_5(6) > \omega_5(5)$. This gives rise to a concavity in the entropy $S(k)=k_B \ln \omega(k)$ of the system with respect to energy at $E=-5\epsilon$, which can be studied using the discrete analog to the second derivative, $S''(k)=[S(k+1)-2S(k)+S(k-1)]$; the function is concave if $S''(k)$ is negative. The concavity results in a bimodal probability distribution function $P(E;T,\sigma/l)$ (illustrated in Fig. 11). In analogy to the study of phase tran-

TABLE II. Comparison of the densities of states for a square-well tetramer chain with $\sigma/l=1.0$ and $\lambda\sigma=1.5$ calculated in this work $[\omega_5(Ek)/4l^4$, with the factor $1/4$ as described in Sec. III] and by Taylor $[\omega_5^{(Taylor)}(k)$, or $g_5^{(k)}$ in the original terminology]. Suppression of an unimportant multiplicative factor of $16\pi^2$ in the work of Taylor leads to the difference in the values; it can be seen that with this factor included the differences in the values are negligible.

k	$\omega_5(k)/4l^4$	$\omega_4^{(Taylor)}(k)$	$\omega_4(k)/[64\pi^2\omega_4^{(Taylor)}(k)]$
0	12.963393	0.08206	1.000386
1	21.368822	0.13531	1.000071
2	14.300928	0.09057	0.999908
3	7.300283	0.04626	0.999342
4	2.284838	0.01447	0.999924
5	0.0633290	0.004012	0.999590
6	0.035383	0.0002222	1.008395
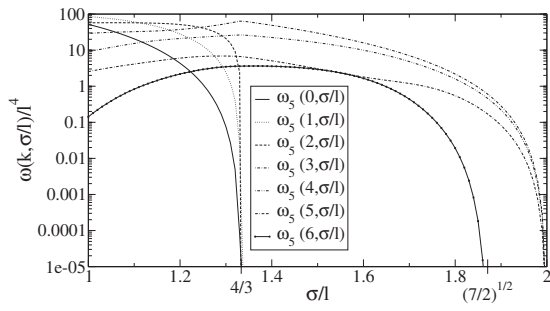
FIG. 10. Densities of states for pentamers plotted against σ/l.

FIG. 12. Ensemble average energies $\langle E \rangle/\epsilon$ for pentamers plotted against σ/l and temperature T. Solid lines show energy contours at the labeled value. The dashed line shows the state coexistence line. Note the discontinuities in the slope of the energy contours at $\sigma/l = 4/3$ and $\sqrt{7/2}$ (shown by dotted lines).

sitions, we find the line of temperatures at which the two peaks of these bimodal probability distributions have equal weight—a line of "state coexistence." This line is plotted alongside the data in Figs. 12 (energy) and 13 (heat capacity), and runs from $\sigma/l \approx 1.14$ to $\sigma/l \approx 1.72$. These end points are at nonzero temperature, and occur where the curvature of the free energy at $E = -5\epsilon$ becomes zero. The end points are not associated with heat capacity divergences.

The thermodynamic data show the expected discontinuities in the slope of energy and heat capacity contour at the magic numbers $\sigma/l = 4/3$ and $\sqrt{7/2}$. The lower magic number corresponds to the loss of high-energy states, as for the tetramer. The larger magic number, corresponding to the loss of the $k=6$ state, gives a discontinuity in the energy at zero temperature (from $E = -6\epsilon$ to $E = -5\epsilon$). As for the tetramer, bond length fluctuations in real systems would act to smooth out these discontinuities.

The pentamer also shows a line of heat capacity maxima, which lies at lower temperature than the state coexistence line. Both these lines show a discontinuity in slope at $\sigma/l = 4/3$. Both lines are doubly reentrant, showing one maximum below $\sigma/l = 4/3$, and another above $\sigma/l = 4/3$. The line of heat capacity maxima connects with the discontinuity in energy at $\sigma/l = \sqrt{7/2}$.

The dihedral behavior of the pentamer at zero temperature (ground state) is shown in Fig. 14, in four representative

plots of the ground state dihedral densities of states $g_5(\phi_1, \phi_2, 6)$ and $g_5(\phi_1, \phi_2, 5)$. These are equivalent to unnormalized dihedral probability distributions for the system at $T=0$. We see that, for the three lowest values of σ/l, the probabilities are peaked at points on the $\phi_1 = \phi_2$ diagonal; the dihedrals prefer to take the same sign. This continues to the magic number $\sigma/l = \sqrt{(3+\sqrt{5})/2}$, where, as for the tetramer, 1-4 overlaps can no longer occur, and steric effects no longer prevent *cis* conformations. As the degree of overlap tends to this number, and the amount of steric interference decreases, the maxima move closer to $\phi_1 = \phi_2 = 0$. For $\sigma/l \gtrsim \sqrt{(3+\sqrt{5})/2}$, the probability distributions become unimodal at $\phi_1 = \phi_2 = 0$. This should be compared with the behavior for the tetramer (see Fig. 9), where $\phi_{\max}(T=0)$ is zero for $\sigma/l \gtrsim 1.48$; this effect is due to the additional steric interference from 1-5 overlaps.

V. DISCUSSION AND CONCLUSIONS

In the previous section, it has been shown that the tetramer and pentamer show a rich and surprising range of behaviors. Specifically, these are magic numbers of the overlap
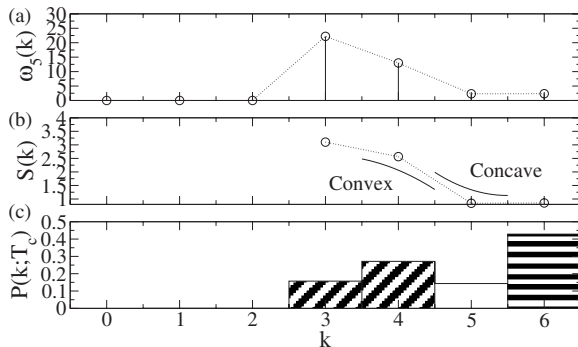

FIG. 11. Concavity in the entropy and bimodal energy probability distribution for $\sigma/l = 1.55$. (a) The density of states $\omega_5(k)$. (b) The entropy as a function of k; note the concavity at $k=5$. (c) The probability distribution function $P(k,T)$ at the state coexistence temperature. The function is bimodal, and the total weights of the two "states" [$k < 5$ (diagonal shading) and $k > 5$ (horizontal shading)] are equal. Dashed lines serve as a guide to the eye.
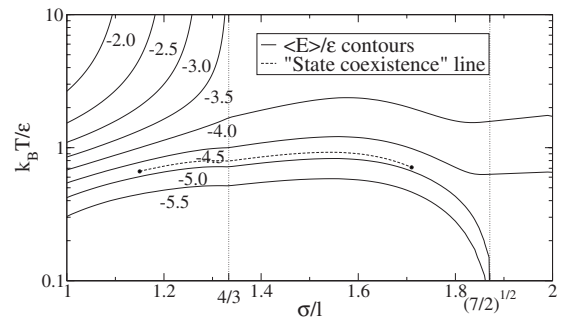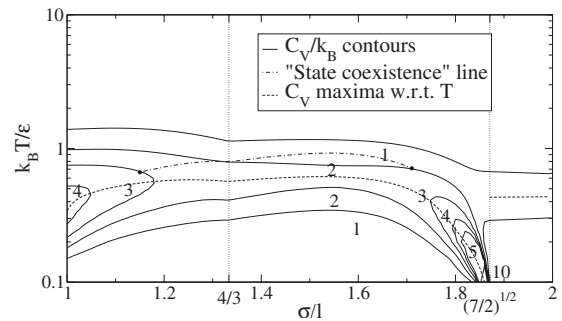

FIG. 13. Configurational heat capacity C_V for pentamers plotted against σ/l and temperature. Solid lines show contours at the labeled value. The dashed line shows the line of maxima in C_V with respect to temperature T. The dot-dashed line shows the state coexistence line—the position of the end points of this line near a contour line is purely coincidental. Note the discontinuities in the slope of the heat capacity contours at $\sigma/l = 4/3$ and $\sqrt{7/2}$ (shown by dotted lines).
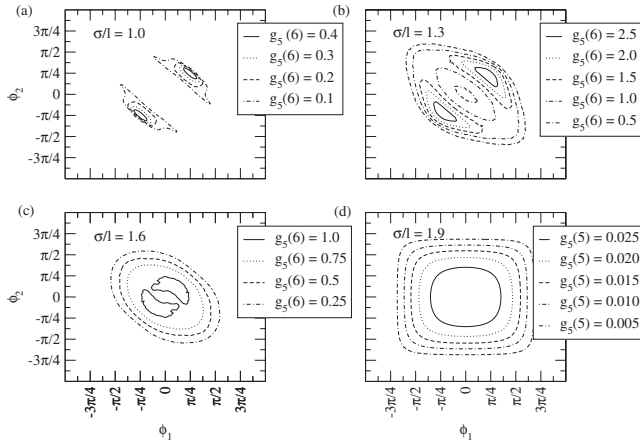
FIG. 14. Ground-state dihedral densities of states $g_5(\phi_1,\phi_2,6)$ for $\sigma/l=$(a) 1.0, (b) 1.3, and (c) 1.6, and (d) $g_5(\phi_1,\phi_2,5)$ for σ/l $=1.9$; the ground state is -5ϵ for $\sigma/l=1.9$. Lines show contours as denoted in figure legends.

σ/l where the derivatives of the densities of states change discontinuously, maxima in specific heat with respect to temperature, and a region of bimodal energy probability distributions, reminiscent of a first-order transition in bulk systems. In general, the behavior of long polymer chains cannot be directly inferred from the behavior of very short chains such as those studied in this work. If, however, interactions between monomers widely spaced along a chain can be neglected, the behavior of very short chains can be used as a basis for a spin chain model. Such interactions may be neglected when chains become very stiff (for, e.g., large values of σ/l, or after helix formation). In this case, the behavior of the very short chains may be considered the "building block" for the behavior of longer chains.

The magic numbers that are observed correspond to discontinuous changes in the derivatives of the densities of states. At $\sigma/l=4/3$, 1-3 contacts become always on and high-energy densities of states become zero. At σ/l $=\sqrt{(3+\sqrt{5})/2}$, the chain becomes so stiff that it cannot bend back upon itself far enough for 1-4 overlaps to occur. Similarly, at $\sigma/l=\sqrt{7/2}$, the chain becomes so stiff that 1-5 contacts can no longer occur, and the ground state for the pentamer is lost. These discontinuities in the densities of states are associated with discontinuities in the energy and compressibility with respect to the parameter σ. The magic numbers are similar in principle to the cutoff λ values noted by Taylor [12] for tangent chains—this work has not examined the effects of changing the well width parameter λ, but it is obvious that the values of these magic numbers will depend upon that parameter, and that cutoff values of λ will also exist for this model. As has been noted above, the discontinuities across lines of constant σ/l in this system will be smoothed in simulations with variable bond length; however, the effects should still be visible. We particularly note the sudden loss of stability of the helix 1 phase at $\sigma/l\approx1.675$ in previous simulation work [3] (see Fig. 1). Given the 10% bond length fluctuation allowed in those simulations, this loss of stability may coincide with the magic number at $\sigma/l=\sqrt{(3+\sqrt{5})/2}\approx1.618$, suggesting that the more tightly

wound helix 1 phase is stabilized by steric interference of 1-4 contacts. This supposition is supported by the observed loss of double-peaked dihedral angle probability distributions for overlaps above this magic number, suggesting that the more loosely wound helix 2 phase is stabilized by steric interference between monomers spaced further along the chain.

The low-temperature maxima in the specific heat for these short polymers appear to be a continuation of the specific heat maxima observed for short tangent chains [12,13]. We follow these previous works in interpreting these maxima as signatures of collapse to close-packed, low-energy conformations. This interpretation appears confirmed by the presence of bimodal energy probability distributions for the pentamer, with a line of state coexistence which roughly parallels the line of maxima.

For the tetramer, the line of specific heat maxima shows reentrance below $\sigma/l=4/3$, having a maximum with respect to temperature. For the pentamer, both the lines of specific heat maxima and of state coexistence are doubly reentrant, showing maxima below and above $\sigma/l=4/3$. The reentrance of the state coexistence line can be easily explained by reference to the densities of states shown in Fig. 10. Consider the system for $\sigma/l\geq4/3$. For overlaps just above this point, the ground state density of states (the entropy of the low-energy state) is increasing while all other densities of states are decreasing with increasing overlap. Hence, the low-energy state becomes more stable, and coexistence moves to higher temperature. The ground state density of states soon begins to decrease, but as long as it is decreasing more slowly than the higher-energy densities of states, its stability continues to increase. However, on closer approach to σ/l $=\sqrt{7/2}$, the ground state density of states decreases faster than the higher-energy densities of states, and stability decreases. The same argument holds for the line when σ/l $<4/3$. If we interpret the maximum in heat capacity as a result of structural competition between the ground state and higher-energy states (following Stanley *et al.* [19]), we can make the same argument for the reentrance in the lines of maxima for both the tetramer and pentamer. Physically, increasing the overlap of the chain makes configurations with lower energy (more contacts) more likely at first (as monomers are drawn into each other's square wells), but then begins to cut into these low-energy states as the chain becomes too stiff to bend back upon itself and make contacts. We attribute the reentrance of the stability of the helix 1 phase in previous work to this same competition between effects.

Although it seems reasonable to attribute the behavior of the phase boundary between the helix 1 and globule phases to effects seen in the pentamer, it should be noted that the state coexistence seen in the pentamer represents collapse of the pentamer, rather than helix formation. Although the dihedral probability distributions shown in Fig. 14 do show double peaks at nonzero dihedral angles, this is not a sufficient criterion for helicity. The cross-correlation coefficient of these distributions is not significantly above zero; the total statistical weight associated with dihedrals away from the peaks is still large enough to outweigh the correlated peaks. However, the clear double-peaked structure does suggest that the physics necessary for helix formation *is* contained in

these simple, small systems, particularly in the steric interference due to 1-4 overlaps.

While these results appear to clarify certain behaviors observed in simulations, they do raise further questions. Under the interpretation we have offered here, the nature of the helix-2 phase is unclear; this phase is observed to be stable up to $\sigma/l = 1.9$ in simulation [3], where the chain is too stiff for 1-5 overlaps to be the root of the observed chirality. Further, the question of how the helix transition connects (or does not connect) to the crystallizationlike transition observed in simulations for the tangent chain system remains unresolved. Follow-up work, developing a spin chain model for helix formation using the results presented here, is under way; it is hoped that this approach will shed light upon these questions.

ACKNOWLEDGMENTS

[1] L. Pauling and R. Corey, Proc. Natl. Acad. Sci. U.S.A. **37**, 235 (1951).

[2] Y. Snir and R. D. Kamien, Science **307**, 1067 (2005).

[3] J. E. Magee, V. R. Vasquez, and L. Lue, Phys. Rev. Lett. **96**, 207802 (2006).

[4] A. Maritan, C. Micheletti, A. Trovato, and J. R. Banavar, Nature (London) **406**, 287 (2000).

[5] D. Marrenduzzo, A. Flammini, A. Trovato, J. R. Banavar, and A. Maritan, J. Polym. Sci., Part B: Polym. Phys. **43**, 650 (2005).

[6] J. P. Kemp and Z. Y. Chen, Phys. Rev. Lett. **81**, 3880 (1998).

[7] V. Varshney, T. E. Dirama, T. Z. Sen, and G. A. Carri, Macromolecules **37**, 8794 (2004).

[8] A. J. Doig, Biophys. Chem. **101**, 281 (2002).

[9] B. Zimm and J. Bragg, J. Chem. Phys. **28**, 1246 (1958).

[10] S. Lifson, J. Chem. Phys. **34**, 1963 (1961).

[11] J. H. Gibbs and E. A. diMarzio, J. Chem. Phys. **28**, 1247 (1958).

[12] M. P. Taylor, J. Chem. Phys. **118**, 883 (2003).

[13] Y. Q. Zhou, M. Karplus, J. M. Wichert, and C. K. Hall, J. Chem. Phys. **107**, 10691 (1997).

[14] A. D. MacKerell *et al.*, J. Phys. Chem. B **102**, 3586 (1998).

[15] M. R. Sadr-Lahijany, A. Scala, S. V. Buldyrev, and H. E. Stanley, Phys. Rev. Lett. **81**, 4895 (1998).

[16] E. A. Jagla, J. Chem. Phys. **111**, 8980 (1999).

[17] G. Franzese, G. Malescio, A. Skibinsky, S. V. Buldyrev, and H. E. Stanley, Nature (London) **409**, 692 (2001).

[18] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C*, 2nd ed. (Cambridge University Press, Cambridge, U.K., 2002).

[19] H. E. Stanley, S. V. Buldyrev, M. Canpolat, M. Meyer, O. Mishima, M. R. Sadr-Lahijany, A. Scala, and F. W. Starr, Physica A **257**, 213 (1998).