Collapse transitions in a flexible homopolymer chain: Application of the Wang-Landau algorithm

D. T. Seaton, T. Wüst, and D. P. Landau
Center for Simulational Physics, University of Georgia, Athens, Georgia 30602, USA
(Received 1 October 2009; published 5 January 2010)

The thermodynamic behavior of a continuous homopolymer is described using the Wang-Landau algorithm for chain lengths up to $N=561$. The coil-globule and liquid-solid transitions are analyzed in detail with traditional thermodynamic and structural quantities. The behavior of the coil-globule transition is well within theoretical and computational predictions for all chain lengths, while the behavior of the liquid-solid transition is much more susceptible to finite-size effects. Certain "magic number" lengths ($N=13,55,147,309,561$), whose minimal energy states offer unique icosahedral geometries, are discussed along with chains residing between these special cases. The low temperature behavior near the liquid-solid transition is rich in structural transformations for certain chain lengths, showing many similarities to the behavior of classical clusters with similar interaction potentials. General comments are made on this size dependent behavior and how it affects transition behavior in this model.

DOI: 10.1103/PhysRevE.81.011802 PACS number(s): 36.20.−r, 64.60.Cn, 02.70.Uu, 05.70.−a

I. INTRODUCTION

The simulation of flexible homopolymers has been a topic of vigorous research over the past decade, with studies varying from investigations of single chains to the interaction of chains with surfaces or pores. In studies of single chains, the models have varied as much as the physical situations, with some being on-lattice [1,2], some in the continuum [3–9], as well as bond-fluctuation lattice models [10–12]. The level of interest is due in part to the number of interesting technological and biological applications, and in part from the availability of extended ensemble simulation methods. These methods allow efficient sampling of all thermodynamic behavior where more traditional methods are unable to probe the rough energy landscapes associated with these systems [13]. Regardless of the model and method used, certain characteristic features remain consistent: As the temperature of a distended, single chain is decreased, a collapse transition from an unwound coil to a globular state (coil-globule) occurs, followed by another transition from a liquid-globule to a "solid," compact globule (liquid-solid). Recent studies have attempted to illuminate the behavior of these two transitions using Wang-Landau sampling. One set of studies using the bond-fluctuation model [10,11], and another employing off-lattice simulations [6,7], have analyzed these transitions for a number of different chain lengths. Noticeable discrepancies resulted in thermodynamic properties and it was unclear whether these were due to differences between the models or in the details of the application of the method. As we began our own work [14,15], we found that our results were not in good agreement with those seen in the previously mentioned off-lattice case, particularly for the qualitative behavior of the liquid-solid transition.

In addition, lattice models using only nearest-neighbor interactions have further highlighted the complexities of this transition, showing that behavior does not develop smoothly with chain length [1,2]. One might think simulating with long range interactions in the continuum would cause this behavior to decrease or disappear, however, subtle complexities near the liquid-solid boundary remain. Recent studies using multicanonical simulations [3,4] compared classical Lennard-Jones clusters and their own continuous homopolymer equivalent, highlighting the behavior of chains near the liquid-solid transition. A number of unique, chain length dependent, transformations can occur in this region, making the finite-size analysis of the liquid-solid transition much more complicated than previously thought. These finite-size effects have an excellent correlation with the behavior of classical clusters, e.g., the "magic number" set ($N=13,55,147,309,561,\ldots,$), whose minimum energy states all have icosahedral geometries. This magic number behavior, along with other unique geometries associated with particular chain lengths, has also been shown to occur in simulations of homopolymers [3,4,9]. In cluster simulations, the global minimum energy structures play an important role in transition behavior near the liquid-solid transition, and with the above correlations with homopolymers, these types of effects must be considered.

In this article, we present results of Wang-Landau simulations of flexible homopolymers for chain sizes up to $N=561$; including detailed discussion of the simulation and analysis methods. First, we focus on the general behavior of our largest chain lengths, analyzing the coil-globule and liquid-solid transition as has been done in previous studies. We then focus on a particular set of chain lengths between the $N=13$ and 55 magic numbers. This set exemplifies the finite-size effects associated with classical Lennard-Jones clusters. Estimates of the infinite-chain transition temperatures are provided when appropriate, along with discussion of finite-size effects seen for many chain lengths. We also make qualitative and quantitative comparisons with a few of the most recent studies of homopolymer behavior.

II. MODEL

A chain of N identical monomers is defined in continuous space, with each monomer having bonded and nonbonded interactions. Nonbonded monomers interact via a truncated-shifted Lennard-Jones potential given by

$$U^{NB}(r_{ij}) = \begin{cases} U^{LJ}(r_{ij}) - U^{LJ}(r_c), & 0 < r_{ij} \le r_c, \\ 0, & \text{otherwise}, \end{cases} \quad (1)$$

where

$$U^{LJ}(r_{ij}) = \epsilon\left[\left(\frac{\sigma}{r_{ij}}\right)^{12} - 2\left(\frac{\sigma}{r_{ij}}\right)^{6}\right]. \quad (2)$$

r_{ij} is the distance between a pair (i,j) of nonbonded monomers, ϵ is the well depth, and σ is the distance at which the potential energy has its minimum. Dimensionless units are used, where $\epsilon=1$, $\sigma=2^{-1/6}$, and the interaction cutoff distance r_c is 3σ.

Interactions between two bonded monomers consists of a combination of finite extensible nonlinear elastic (FENE) [16,17] and Lennard-Jones potentials

$$U^{B}(l_i) = \begin{cases} U^{\mathrm{FENE}}(l_i) + U^{LJ}(l_i), & 0 < l_i \le R_o, \\ 0, & \text{otherwise}, \end{cases} \quad (3)$$

where the Lennard-Jones potential $U^{LJ}(l_i)$ is of the same form as described above and the FENE potential $U^{\mathrm{FENE}}(l_i)$ is given by

$$U^{\mathrm{FENE}}(l_i) = -0.5kR_o^2 ln[1 - (l_i/R_o)^2]. \quad (4)$$

R_o is the finite extensibility, k is the stiffness constant, and l_i is the bond length. Dimensionless units are again implemented, where $R_o=1.2$ and $k=2$. Analyzing the derivative of $U^B(l_i)$ with the above values allows this potential to be set such that its minimum is zero and occurs at an equilibrium value $l_o=1$. Note, the parameter σ found in the $U^{LJ}(l_i)$ term in Eq. (3) is used to adjust the distance at which the minimum occurs and it has a different value than that applied in Eq. (2).

All parameters for U^{NB} and U^B were chosen such that the minima of both these potentials occur at a dimensionless distance of 1. For comparison, these potentials are plotted in Fig. 1. Other recent studies of continuous homopolymers [3,4,6,7,18] have used an alternate implementation of U^B compared to that described above. In these studies, the nonbonded Lennard-Jones potential is still applied to bonded monomers, meaning that their models contain $N-1$ more nonbonded interactions compared to our model. Results are still comparable since these extra interactions are simply superimposed on an already dominant bonded interaction.

III. METHODS

A. Wang-Landau sampling

The temperature dependence of thermodynamic properties can be calculated from the partition function, which, in terms of a density of states $g(E)$, is

$$Z(T) = \int_E g(E)e^{-E/k_BT}dE \approx \sum_E g(E)e^{-E/k_BT}, \quad (5)$$

where $g(E)$ is the density of states, E is the energy range of a particular system, T is the dimensionless temperature, and k_B is Boltzmann's constant. The $\int_E$ represents the continuous representation, while the Σ_E represents an approximation often used in simulations (see below). While $g(E)$ is generally quite difficult to determine, the Wang-Landau algorithm is now known to be quite efficient in iteratively estimating the density of states $g(E)$ for both discrete and continuous distributions of energy. Details used in the iteration process can be found elsewhere [19,20]. In this study, the modification factor is reduced to $ln(f) \to 10^{-6}$ and we require a flatness of $p=0.4$ for the histogram, where p is the minimum histogram entry divided by its average. We have chosen $ln(f) \to 10^{-6}$ based on other studies of similar systems [5,6,10,11,21] and also through our own testing of the effects of varying this parameter [15]. Recent studies [1,22] have shown that $ln(f) \to 10^{-6}$ can be insufficient, meaning that the modification factor must be reduced further (or in a different way all together [23,24]). For the majority of our chains, we found very little improvement in results when reducing the modification factor below 10^{-6}, with the exceptions being our largest chains, $N=500$ and 561, where fluctuations in thermodynamic properties were improved by reducing $ln(f) \to 10^{-7}$. However, these two chains pose the most challenges to our simulation techniques and general conclusions regarding $ln(f)$ should not be taken from these cases alone. Our choice of flatness criteria was based on our own testing and offers a balance between the most stringent case ($p \ge 0.8$) and cases where little or no flatness criteria is used at all ($p \le 0.1$). In Fig. 2, the logarithm of the density of states is plotted versus energy for chain lengths ranging from $N=50$ to $N=500$. Each curve is the result of a single Wang-Landau simulation and the maximum of each curve is set to zero for clarity. The most noticeable aspects of Fig. 2 are the large differences in relative height between the highest and lowest energy states. The challenges associated with analyzing these data will be discussed shortly.

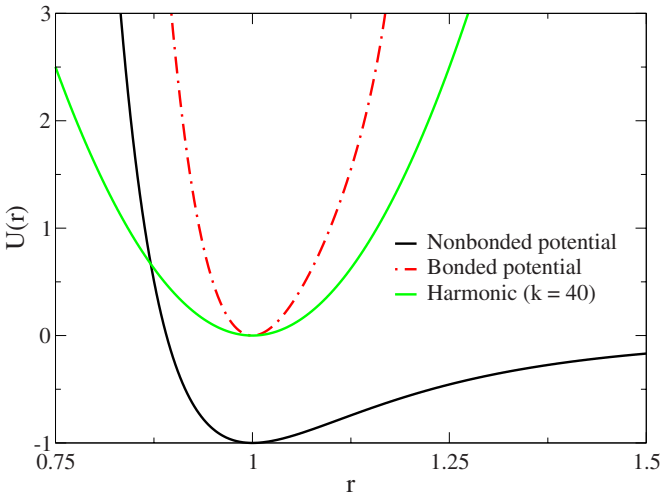

FIG. 1. (Color online) Bonded and nonbonded potentials versus the distance between two monomers. A harmonic potential is plotted as a reference for the bonded potential (FENE-Lennard-Jones combination). The nonbonded potential is plotted to a distance of only 1.5 for clarity (the actual cutoff occurs at 3).

Another feature of Fig. 2 is that the minimum energy $(E_{\min}/N)$ is set according to chain length, while the maximum energy $(E_{\max}/\mathrm{N})$ is set to $E_{\max}/N=3.0$ for all chain lengths. In our previous studies [14,15], we discussed how

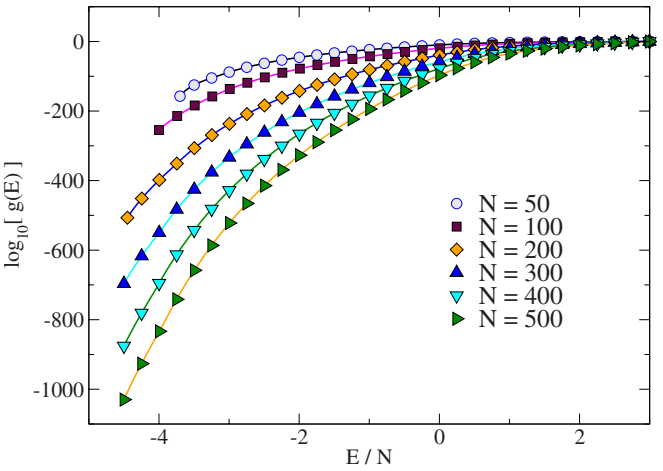

FIG. 2. (Color online) The relative density of states versus energy per monomer for a number of chain lengths. The maximum of each curve is set to zero for plotting purposes. Values for the highest and lowest energy states for the longest chain (N=500) differ by ≈1000 orders of magnitude. (Each curve represents a single run and the majority of symbols were omitted for clarity.)

the energy range over which the random walk is performed can dramatically affect the location of transitions in thermodynamic quantities, leading to inaccurate estimations of transition temperatures. Since the minimum energy and transition temperatures vary with chain length, care was taken to run a number of different simulations, varying the energy range for each chain [14,15], with our goal being to attain consistent behavior for each chain. Energy range effects have been seen in other studies as well [11] and are likely to be the leading contributor to inconsistencies seen in recent work on continuous homopolymers [6,7]. Use of lower $E_{\min}/N$ values requires longer CPU times; hence, all minimum energies are set both to allow determination of low temperature thermodynamic behavior and to consume reasonable amounts of CPU time. The behavior of the coil-globule transition is considered in the same way, but this transition was found to be less susceptible to energy range effects [15]. A constant value of $E_{\max}/N$=3.0 was found to be appropriate.

Equation (5) requires a continuous energy range, but in practice, the energy range must be discretized such that E is divided up into a number of energy bins $E_i = idE + E_{\min}$, where dE is the binning resolution, i is the bin index, and $E_{\min}$ is the minimum energy sampled. Typically, bin widths dE=0.1 were used for chain lengths <100, while chain lengths ≥100 were simulated with dE=1.0. Exception was taken with the largest chain lengths, $N \geq 500$, where dE =2.0. A number of binning resolutions were used, and for the range discussed above, we found that the greatest effect was on computational efficiency, where the number of bins is proportional to both N and CPU time. Minor differences in thermodynamic quantities can be observed when varying dE, but these are generally within statistical errors. An important note is that implementing a large number of bins can lead to excessively large CPU times; the nature of the algorithm is to visit all binned energies a significant number of times in each simulation. Therefore, we attempted to keep the number of bins in each simulation from growing larger than ≈4000, but in some cases this could not be avoided. Similar descriptions of the implementation of the Wang-Landau algorithm in continuous systems can be found elsewhere [21,25,26]. All simulations were run on a single CPU (one core) using the KISS random number generator [27] and error bars were determined from Jackknife analysis of results from multiple, independent runs. The results were unchanged when tests with other random number generators were performed.

Structural quantities were also calculated, including the radius of gyration, the end-to-end distance, and the core density. For example, the radius of gyration is

$$R_g^2 = \sum_i^N (\vec{r}_i - \vec{r}_{cm})^2/N, \tag{6}$$

where the sum is over all N monomers, $(\vec{r}_i - \vec{r}_{cm})$ is the relative separation between monomer i and the center of mass of the chain. Ideally, one of these quantities would be incorporated as a second sampling direction [25], but this becomes unrealistic when multiple structural quantities are desired. Therefore, we calculated the average value of these quantities as a function of energy, e.g., $\overline{R_g^2}(E)$, $\overline{R_e}(E)$, or $\overline{\rho}_c(E)$. This gives the canonical form

$$\langle R_g^2 \rangle(T) = \frac{1}{Z(T)} \sum_E g(E) \overline{R_g^2}(E) e^{-E/k_B T}, \tag{7}$$

where $Z(T)$ is again the partition function for a given temperature, the Σ_E is over all energies E, $g(E)$ is the "final" density of states, $\overline{R_g^2}(E)$ is the average value of the radius of gyration for each energy, and the exponential is the Boltzmann factor. This approach has been used in other studies as well [11,12].

The average values of these structural quantities are calculated after each simulation is complete. Because of inefficiencies in sampling at low values of f [15,23], a saved copy of $g(E)$ from a previous iteration (around $ln[f] \approx 10^{-3}$) is sampled with normal Wang-Landau transition rates. Then, while resampling this iteration, we calculate the average values of structural quantities for each bin visited, making these calculations whether the new trial state is accepted or rejected, and requiring that each bin be visited at least 10^6 times. Using Eq. (7), these structural averages are then applied to the final $g(E)$, which is unaffected by this procedure. In Fig. 3, the average squared radius of gyration is plotted versus energy per monomer. These data show the general collapse behavior of this system, where at high E/N the chain is unwound, and at low E/N the chain becomes more compact. These curves are all monotonic, which is most discernible at high E/N, where the value of $\overline{R_g^2}(E)$ increases systematically with N. To check the accuracy of our work, we also applied this technique to the final iteration of $g(E)$ instead of a previous iteration ($ln[f] \approx 10^{-3}$), with the only differences being that using the final iteration required longer simulation times. We also kept track of the structural averages over entire simulations and found no differences in these techniques.

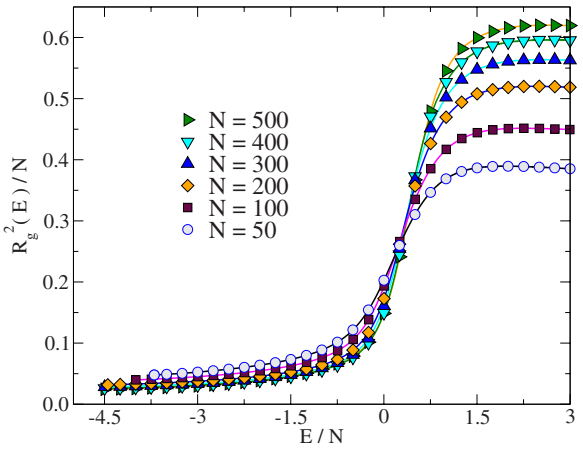

FIG. 3. (Color online) The radius of gyration versus energy per monomer for chain lengths 100, 200, 300, 400, and 500. All curves are monotonic at both low and high E/N. (The majority of symbols were omitted for clarity and error bars are smaller than the symbol size.)

B. Analysis

All thermodynamic and structural quantities are averaged over multiple, independent simulations (typically 20); however, care must be taken when averaging data whose exponent value can be out of the range of double precision arithmetic. The $N=500$ chain, for example, can easily give an underflow when working with $g(E)$. In order to avoid this issue (which typically occurs for sizes $N \geq 100$ in this study), we calculate thermodynamic and structural properties for a given temperature T using only the relevant portion of $g(E)$, which is found by analyzing the canonical probability distribution $P(E,T)$. This analysis allows us to sum over only the most important portion of each $g(E)$, minimizing the sum over the partition function [see Eq. (5)], hence, minimizing the underflow issues. Two criteria are applied when analyzing the importance of $g(E)$ for a particular T, both considering the relative weight of $P(E,T)$ with the maximum of the distribution $P_{\max}(E,T)$. The first is that for each bin considered in the Σ_E, the relative weight must be $P(E,T)/P_{\max}(E,T) > 10^{-16}$. This restriction ensures that extremely small contributions to $Z(T)$ are not considered, e.g., low temperatures where the behavior of $P(E,T)$ becomes similar to that of a delta function, there is no need to sum over the entire $g(E)$. The second criterion ensures that each T has a complete distribution, where $P(E,T)/P_{\max}(E,T) < 0.001$ on both sides of the distribution. This restriction avoids the inclusion of thermodynamic values near the boundaries where often $g(E)$ can be incomplete due to energy restrictions (a similar scheme was implemented in Ref. [6]). When considering k multiple runs, each distribution $P_k(E,T)$ is analyzed and their relevant energy range stored. The maximum energy range encompassing all of these individual ranges is then applied to the summation in Eq. (5). Each $g(E)$ is then normalized within this newly defined range and these curves are averaged, giving $\overline{g(E)}$ for that range. Thermodynamic and structural quantities are then calculated using a resolution of $dT=0.001$ for temperatures between $0.001 \leq T \leq 5.0$. The second constraint gives an applicable temperature range that mainly depends on $E_{\min}/N$. In this study, the lowest applicable temperature is for the smallest length ($N=10$) and is $T=0.036$. This lower limit increases with N, and is $T=0.511$ for our largest chain ($N=561$). Care is taken to incorporate as much of the low temperature behavior as is possible within our limits of CPU time. Thermodynamic properties are not only difficult to determine at these temperatures but also exhibit rich phenomena in homopolymers, heteropolymers, and clusters [1,3,4,22,28,29]. The error bars of all quantities are calculated using a Jackknife analysis [30]. Quantities such as the heat capacity C_V are calculated after $g(E)$ is averaged.

Recent studies of single homopolymer chains [6,7,10,11] have all reported transition temperatures for the coil-globule transition using Wang-Landau simulations. Another study using lattice models and multicanonical simulation techniques gave an in depth analysis of this transition [2]. In each study, the following size dependence was considered:

$$T(N) - T_\Theta = a_1/\sqrt{N} + a_2/N, \tag{8}$$

where $T(N)$ is the transition temperature for each chain length and T_Θ is the Θ temperature [31].

An infinite-chain length prediction of the liquid-solid transition has also been offered [10,11],

$$T(N) - T_{ls} = a/N^{1/3}, \tag{9}$$

where $T(N)$ is again the transition temperature for each chain length and T_{sl} is the liquid-solid transition temperature for an infinite chain. Infinite-chain estimates have been reported [6,7,10,11] for the liquid-solid transition, but with the recent understanding of chain length dependencies [2–4] in these types of systems, more care must be taken when analyzing this transition. One question that has also been addressed in many of the recent homopolymer publications [2,6,7,10] is whether these two transitions merge in the infinite-chain limit. Simulations of the bond-fluctuation model [11] originally showed these two transitions approaching each other, but a later study [10] showed that by extending the range of this interaction, the transitions remained separated in the infinite-chain limit. Studies of lattice homopolymers [1,2] have also shown these two transitions to remain separated, along with simulation of continuous models [3,4,6,7,14,15]. Whether these two transitions merge or remain separated is still a relevant issue. A recent study, detailing how crystallization may prevent the collapse if the range of the attractive part of the interaction is sufficiently short, has been used to model protein phenomena [5].

IV. SIMULATION RESULTS

A. Heat capacity

In Fig. 4, the heat capacity is plotted versus temperature for chain lengths between $100 \leq N \leq 500$ (heat capacity was chosen for visual clarity). Two distinct features are observed, the most discernible being sharp (weakly rounded first-order) peaks at low temperatures ($\approx 0.3-0.6$) representing the liquid-solid transition, and the other being shoulders at

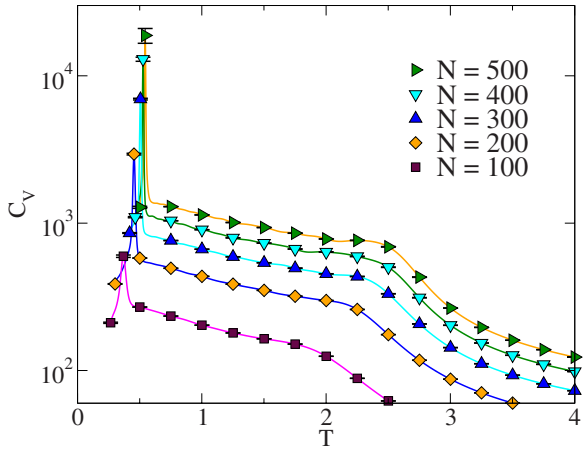

FIG. 4. (Color online) The heat capacity for chain lengths 100, 200, 300, 400, and 500. Two noticeable features appear: large peaks at low T representing the liquid-solid transition, and shoulders at high T representing the coil-globule transition. Each curve is the average of at least 20 independent simulations, where the error bars are calculated using a Jackknife analysis (dT=0.001).

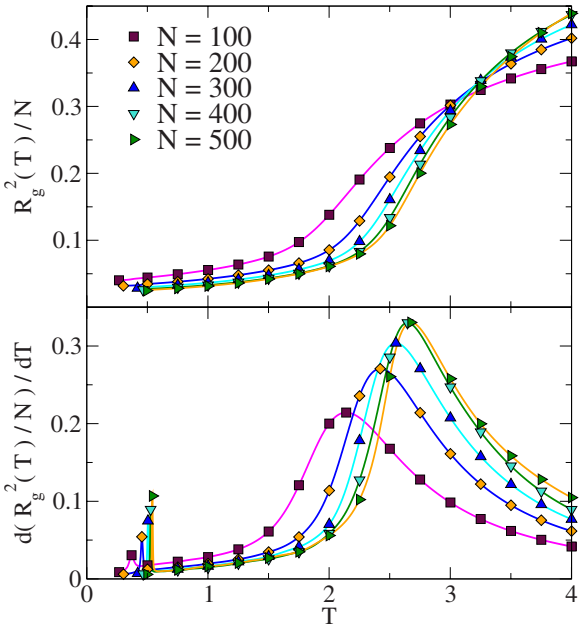

FIG. 5. (Color online) The radius of gyration (top) and its derivative (bottom) versus temperature for chain lengths 100, 200, 300, 400, and 500. Features that were seen in the heat capacity are also present in this figure. The derivative indicates the coil-globule transition by large, broad peaks at high T. The liquid-solid transition is also represented at low T, but the peaks are much smaller in height and less distinguishable. (The majority of symbols were omitted for clarity and error bars are smaller than the symbol size, dT=0.001.)

higher temperatures (≈1.75–2.75) indicating the coil-globule transition. The two transitions are well separated, with no indication that they would merge in the infinite-chain limit (in view of the arguments [5,10] about when such a merging could occur, it is not expected for the present model). The region between these two transitions is also of interest since one study claims that a liquid-liquid transition occurs [6,7], while another has hypothesized that "interesting behavior" may occur [10]. Our results show no evidence of a third independent transition, and analysis of structural quantities supports this conclusion. The heat capacity, together with visualizations of the homopolymer, describes the general trends discussed previously, i.e., as T decreases, a single chain undergoes a coil-globule transition, followed by a liquid-solid transition. However, we will show that certain chain lengths exhibits nonmonotonic finite-size effects in the liquid-solid temperature regime. Estimation of infinite-chain behavior will be discussed in detail later in this section.

B. Structural quantities

In Fig. 5, we plot the squared radius of gyration per monomer versus temperature, along with its derivative with respect to T, for the same chain lengths considered previously. The radius of gyration behaves monotonically at both low and high temperatures, showing the relation between overall size and chain length. The derivatives with respect to temperature provide more illuminating results. Temperatures between ≈0.3–0.6 show five sharp, individual peaks, each representing the liquid-solid transition for a particular chain length. The temperatures at which these peaks occur correspond with those seen in Fig. 4. Clear indications of the coil-globule transition are the most noticeable feature, indicated by the large, broad peaks between temperatures of ≈1.75–2.75. The similarity between the N=400 and N =500 curves suggests that we are reaching chain lengths in the asymptotic regions, however, we are also pushing the limits of our ability to sample these large systems. The liquid-solid peaks however remain well separated, but are still less distinguishable compared to the coil-globule peaks. The radius of gyration gives an excellent description of the coil-globule transition, as does the end-to-end distance, which is not shown because the two quantities give nearly identical results. However, another quantity is perhaps needed in order to better characterize the liquid-solid transition.

One such quantity is the core density, which is plotted in Fig. 6, along with the negative of its derivative with respect to T, for the chain lengths previously considered. This quantity counts the number of monomers within a spherical shell whose origin is at the center of mass and radius is the maximum interaction distance of the nonbonded potential, 3σ. For chain lengths ≥200, large peaks similar to those seen in the heat capacity form in the derivative at T≈0.3–0.6, giving clear indications of the liquid-solid transition. Since this quantity simply counts the number of monomers within an interaction sphere around the central monomer, the N=100 length chain has only a small peak, due to the fact that nearly all monomers are already within the interaction sphere at the temperature where the liquid-solid transition occurs. The largest chains show very similar behavior in the core density since they are simply reaching the limit of the number of monomers that can be packed into this single interaction sphere. However, the derivatives still describe the liquid-solid transition quite well, where large peaks can be seen steadily increasing in height and temperature as N increases.

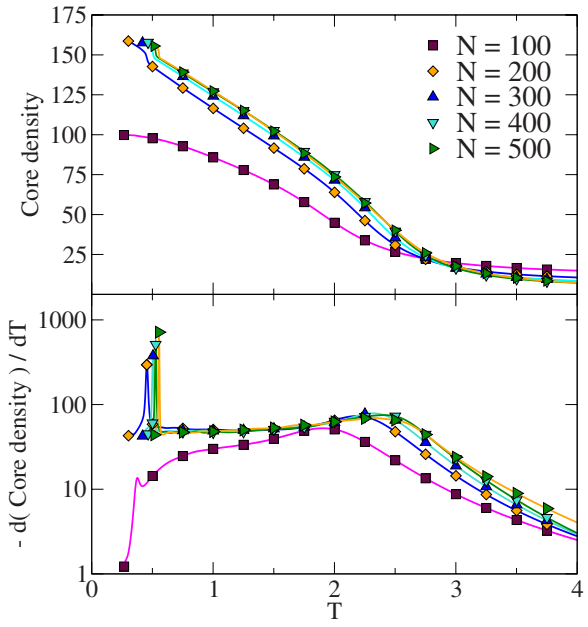

FIG. 6. (Color online) The core density (top) and its derivative (bottom) versus temperature for chain lengths 100, 200, 300, 400, and 500. Features that were present in previous figures are also found here, namely, peaks at low T indicative of the liquid-solid transition and broad peaks (similar to the shoulder seen in the specific heat) at high T indicative of the coil-globule transition. (The majority of symbols were omitted for clarity and error bars are smaller than the symbol size, $dT=0.001$.)

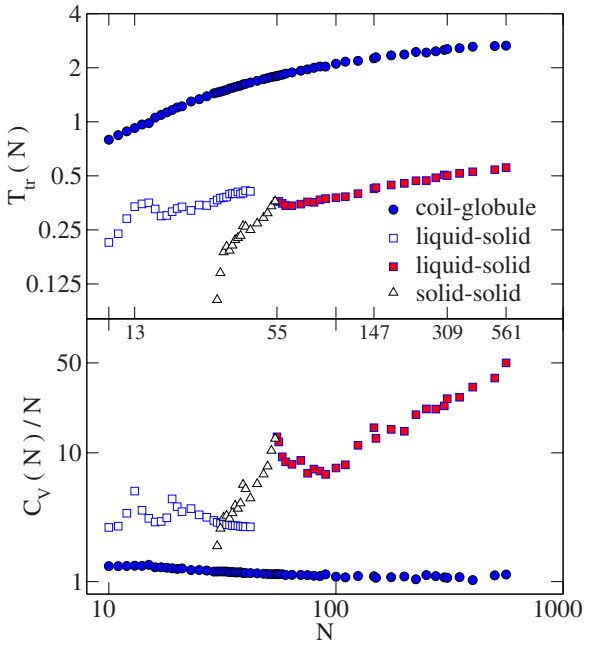

FIG. 7. (Color online) Transition temperatures (T_{tr}) from the specific heat (top), along with the corresponding specific heat (C_V/N) values (bottom), plotted versus chain length N. The coil-globule (circles), the liquid-solid (squares), and solid-solid (triangles) transitions are all represented. The coil-globule and liquid-solid transitions are present in nearly all chain lengths, while the solid-solid transition appears between $30 \leq N \leq 54$. The small gap between $42 \leq N \leq 54$ is a result of the difficulty in distinguishing between the liquid-solid and solid-solid transition. This plot has striking similarities compared to recent studies of classical Lennard-Jones clusters (Ref. [28,29,32]). (Error bars are on the order of or smaller than the symbols.)

Furthermore, there are indications of the coil-globule transition at high T, but they are not as discernible as those seen in Fig. 5.

The results for the chain lengths discussed so far offer a qualitative description of the collapse behavior in our system and can be compared directly with work seen in other homopolymer studies using similar chain lengths. Results from the bond-fluctuation model [10–12] show qualitative agreement with the behavior of our model, while results from off-lattice models remain mixed [3,4,6,7]. The comparison with off-lattice cases is especially important, particularly since our models are so similar. Quantitative comparisons will be made later in this section.

C. Transitions and clusters

All of the quantities discussed above offer insight into the nature of the coil-globule and liquid-solid transitions. Analyzing these quantities gives the location of transition temperatures $T_{tr}(N)$ for each chain length, which then leads to estimations of infinite-chain transition temperatures. However, a more detailed resolution in N reveals complex finite-size effects associated with the liquid-solid transition. These effects must be addressed before estimates of the infinite-chain transition temperatures can be calculated. In Fig. 7, transition temperatures $T_{tr}(N)$ (top) determined from the specific heat, along with the subsequent peak values $C_V(N)$ (bottom), are considered for a number of chain lengths between $10 \leq N \leq 561$ (numerical data can be found in Table I). For such small systems in the canonical ensemble, no sharp phase transition can occur, and only a rounded transition region appears. By "transition temperature" we then denote the temperature at which the specific heat maximum occurs, marking the center of the region. Three distinct trends in peak behavior can be found in these data: the coil-globule transition, and two transitions found in the liquid-solid regime. The coil-globule transition is represented by filled circles, the liquid-solid transition is represented by filled and unfilled squares, and the solid-solid transition (not reported in many previous studies [6,7,10,12] with which we have compared) appears for chain lengths $30 \leq N \leq 54$ and is represented by unfilled triangles. The discussion begins by considering the coil-globule transition, which is found by analyzing the shoulder seen at high temperatures in the specific heat (see Fig. 4). In Fig. 7, each chain length has an estimate of T_{tr} and fluctuations are small relative to neighboring chain lengths. The largest fluctuations are seen for the largest chain lengths, where the shoulders representative of this transition can become increasingly difficult to analyze and error becomes more relevant. The coil-globule transition has been analyzed in detail in other studies [2,6,7,10,12] of homopolymers and our results compare qualitatively well with those works. However, the quantitative description of the infinite chain still differs according to model (which will be discussed shortly).

TABLE I. Liquid-solid (T_{ls}) and coil-globule (T_{cg}) transition temperatures along with corresponding C_V/N values for magic number chain lengths. Data corresponds to Figs. 7 and 8. (Error estimates were calculated using a Jackknife analysis.)

N	T_{ls}	$C_V(T_{ls})/N$	T_{cg}	$C_V(T_{cg})/N$
13	0.33679 ± 0.00007	5.037 ± 0.003	0.9230 ± 0.0005	1.3294 ± 0.0009
55	0.3614 ± 0.0003	13.31 ± 0.03	1.799 ± 0.001	1.145 ± 0.001
147	0.425 ± 0.001	15.7 ± 0.4	2.258 ± 0.002	1.101 ± 0.003
309	0.5024 ± 0.0006	26.4 ± 0.2	2.542 ± 0.006	1.071 ± 0.009
561	0.556 ± 0.002	50.0 ± 1.7	2.65 ± 0.01	1.13 ± 0.03

The most striking feature seen in Fig. 7 is the low temperature behavior in the "liquid-solid" region. The belief regarding various homopolymer models has been that these systems undergo only two distinct transformations, namely, the coil-globule transition and a single liquid-solid (crystallization) transition. What was found by Schnabel and coworkers [3,4], and verified by our results, is that for certain chain lengths, a number of unique structural transformations, not limited to a single crystallization, can be found at low temperatures. The data plotted in Fig. 7 for low temperature transitions ($T \leq 0.6$) has remarkable similarities to the behavior of Lennard-Jones clusters [28,29,32,33], and Schnabel and co-workers [3,4] have confirmed that there is a near one-to-one correspondence between these two systems. To fully understand the behavior of our model, general features of classical Lennard-Jones clusters must be considered.

In Table II, we list features which have been reported in recent cluster studies [28,29,32] regarding the ground state and transition behavior of cluster sizes up to $N=309$ (these features are also present in [3,4] with only a few exceptions). The magic number set $N=13$, 55, 147, 309, and 561 all have icosahedral geometries and smaller magic number configurations remain as cores of the longer chains. The specific heat versus temperature from our own homopolymer simulations of the magic set lengths can be found in Fig. 8 (numerical data can be found in Table I). These chains offer the previously expected transition behavior, namely, a single liquid-solid transition and a coil-globule transition. However, they also account for the plethora of unexpected ground state and transformation behaviors occurring for chain lengths between these numbers. The general descriptions of anti-Mackay and Mackay arrangements [32,34,35] are used to classify the ground state configurations of clusters. In Fig. 9, the anti-Mackay and Mackay overlayers are represented on one face of a $N=13$ icosahedral core. These terms describe packing in the outer shell of a cluster, where typically an icosahedral core (based on the magic set) forms and additional monomers begin arranging on its surface. In the Mackay case, enough monomers are present that the next layer of a magic number icosahedron forms. Anti-Mackay packing can form a variety of geometries where monomers locate the most energetically favorable place to attach to the core's surface. In clusters, sizes between $14 \leq N \leq 30$ have anti-Mackay ground states and a single peak associated with melting is visible in the specific heat. At $N=31$, the ground state behavior changes to the Mackay type and a peak forms well below the melting transition. This new peak corresponds to a transformation in the outer shell of the cluster

TABLE II. Table of transformations seen in studies of classical Lennard-Jones clusters (Refs. [28,29,32]) for various cluster sizes. These transformations are typically identified using peaks in the specific heat. A recent study of homopolymers (Refs. [3,4]) with Lennard-Jones interactions revealed striking similarities. The peaks identified here are found near the liquid-solid transition.

Cluster size	Ground state	Peaks in C_V
13,55,147,309	Complete icosahedral	1
$14 \leq N \leq 30$	Anti-Mackay	1
$31 \leq N \leq 54$	Mackay	2
$56 \leq N \leq 81$	Anti-Mackay	1
$82 \leq N \leq 146$	Mackay	2
Exceptions		
38,75–77	Non-icosahedral	2
98,102–104		

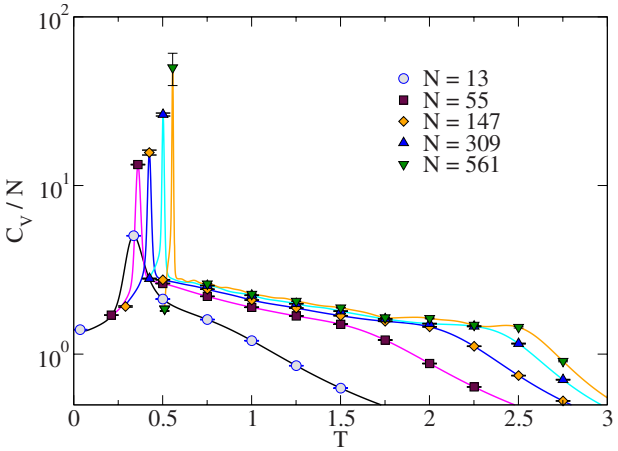

FIG. 8. (Color online) The specific heat versus temperature for the magic set of chain lengths, where minimum energy configurations for each chain correspond to an icosahedral geometry. These chain lengths offer two generic features: a shoulder in C_V/N indicative of the coil-globule transition (high temperatures) and a first-order like peak indicative of a liquid-solid transition (low temperatures). They also offer a basis for understanding the behavior of chain lengths between two magic lengths.

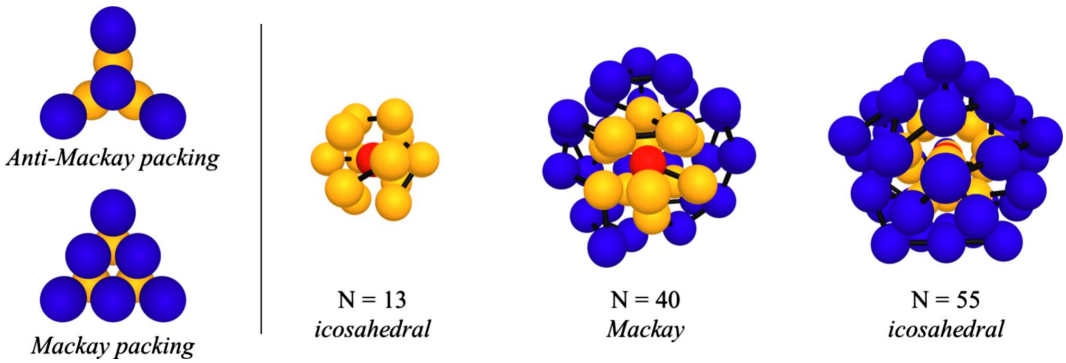

FIG. 9. (Color online) Illustration of Mackay and Anti-Mackay packing, along with minimum energy states for $N=13$, $N=40$, and $N=55$. $N=13$ is the first icosahedral geometry in the magic set, and $N=55$ is the next. The $N=40$ state represents a Mackay ground state, where a $N=13$ icosahedral core can be found at its center. This illustration represents a class of minimum energy states where the outer layer surrounding the icosahedral core begins forming the overlayer for the next magic number chain length. (The central monomer is red, the first monomer layer is yellow, and the outer monomer layer is blue.)

and is representative of a change from anti-Mackay to Mackay states (aM/M). This transition has been well studied in cluster systems [29,32,36], but has not been considered in great detail in homopolymer studies until recently [3,4].

In Fig. 7, we see very similar behavior compared to the cluster studies [[28,29,32]] discussed above, as well as the recent homopolymer study [3,4]. In particular, chain lengths between $30 \leq N \leq 42$ exemplify the finite-size effects seen in these studies. Chains below this range have a single peak

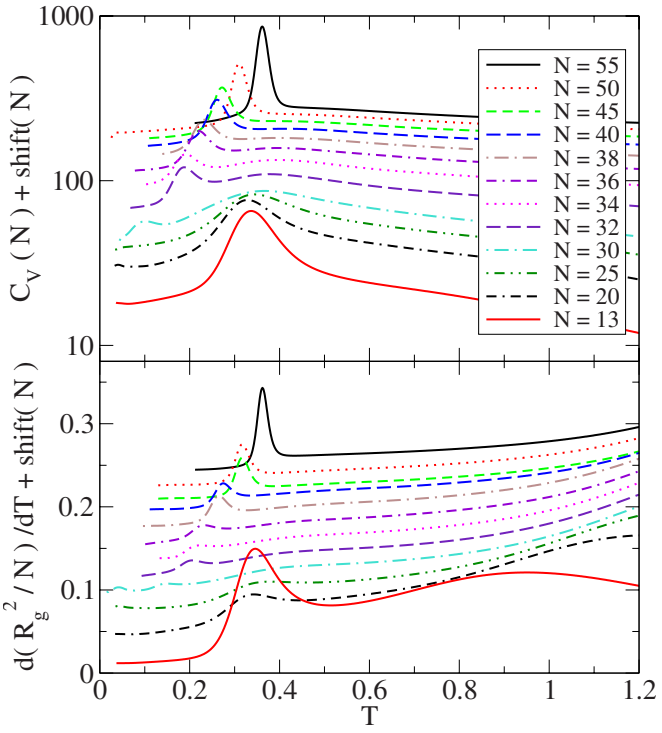

FIG. 10. (Color online) The heat capacity (top) and the derivative of the radius of gyration (bottom) versus temperature for chain lengths between $13 \leq N \leq 55$. These chain lengths offer unique behavior in the liquid-solid regime, where for some chains two peaks can be seen in the heat capacity and a peak and shoulder in the radius of gyration. Each curve is shifted by an arbitrary amount for visual clarity. ($dT=0.001$)

representative of the liquid-solid transition (unfilled squares), the most prominent being the magic length $N=13$, whose ground state is a perfect icosahedron. At $N=30$, a small peak forms (open triangles) and it continues to grow as N increases, eventually merging with the liquid-solid transition. This is contrary to both recent cluster studies and Schnabel *et al.*, where in those studies the small peak indicative of a transformation in the outer shell of the homopolymer appears for $N=31$ rather than $N=30$ (which we currently attribute to the minor model differences discussed in Sec. II). In Fig. 10 we provide more detailed evidence of this behavior by plotting C_V/N and $d(R_g^2)/dT$ as functions of temperature. In our work, the solid-solid (aM/M) transition first appears for $N=30$ at $T=0.1$ in both quantities. Chain lengths $30 \leq N \leq 34$ offer the most discernible indication of the solid-solid (aM/M) transition and the liquid-solid transition existing in a single chain. As N increases, the solid-solid transition becomes more defined, while the liquid-solid transition becomes broader, appearing to eventually overcome the solid-solid transition near $N=42$. A shoulder still exists for many larger lengths in the $42 < N < 55$ range, but this is not easily analyzed within our current framework. The merger of these two transitions is an interesting, but also challenging topic. Cluster simulations [29,32], as well as the recent homopolymer study [3,4], have given details regarding behavior in this region and our results are in good agreement. Cluster studies have shown these two transitions to merge just below the $N=55$ Mackay icosahedron [37], where the Mackay ground states become much more favorable compared to the anti-Mackay states as they approach $N=55$. What these chain lengths indicate, along with the studies discussed above, is that the interplay between Mackay and anti-MacKay states can lead to significant thermodynamic behavior.

An important issue brought about by these effects is whether or not these types of transformations occur in larger chain lengths. The range of sizes between the two magic number $N=55$ and 147 have shown similar Mackay and anti-Mackay structure in the outer shell of particular cluster sizes (as represented in Table II) and Schnabel and co-workers [3,4] have found similar behavior in this range (with a few exceptions). This becomes more interesting for the next magic number range $147 < N < 309$, where in Fig. 7 one

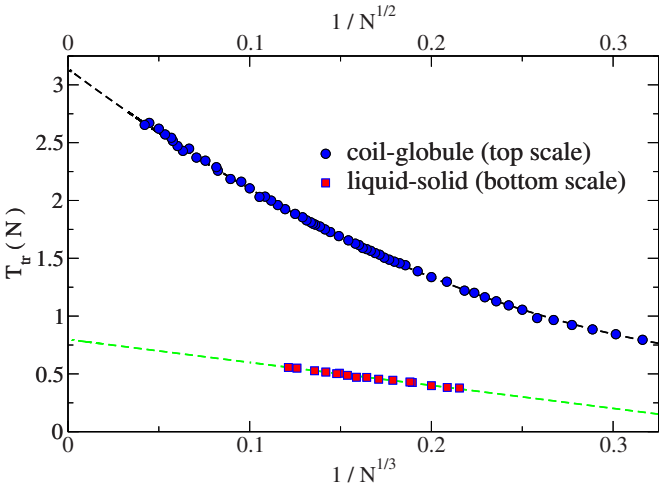

FIG. 11. (Color online) Analysis of the liquid-solid and coil-globule transitions using transition temperatures provided by features in the specific heat. The liquid-solid transition data (squares) are plotted with the lower-horizontal axis, and the coil-globule transition data (circles) are plotted with the upper-horizontal axis. Dashed lines represent the fits from Eqs. (8) and (9).

might expect these effects to have become insignificant because of core-surface arguments. However, cluster studies still show signs of multiple transitions at low temperatures. In fact, one study [35] has shown two-peak behavior within the magic number $N=309$. This implies that these effects may continue well past the range of chain lengths considered in most recent studies. This could all be further complicated by transformations in larger clusters >1690, where systems prefer decahedral geometries over icosahedral [3,4,38]. The results shown in Figs. 7 and 10 reveal that transition behavior in homopolymers is not as ideal as has been presented in previous studies. The coil-globule transition is still rather consistent, but the liquid-solid transition is far more susceptible to finite-size effects. Real polymer chains are much more complicated than the present model, and optimal packing near the liquid-solid transition will subsequently depend on chemical structure.

In Fig. 11, the predictions discussed in the methods section are considered, where Eqs. (8) and (9) are applied to results (seen in Fig. 7) from the analysis of the specific heat and the radius of gyration. The coil-globule transition behaves as is expected, and the fit is seen to coincide with the data for all chain lengths. In Table III, we present the fitting results for the coil-globule transition using both specific heat and radius of gyration data, where the infinite-chain prediction of the Θ temperature is $T_\Theta=3.156\pm0.007$ for C_V/N and $T_\Theta=3.176\pm0.004$ for R_g^2. This fit was applied to chain lengths between $10\leq N\leq561$ and the results agree quite well with the prediction in Eq. (8). There are some fluctuations, especially at large N, but these fluctuations do not fall outside the bounds of statistical error. In comparing our fitting results with those seen in the work from Parsons and Williams, we find that our estimate of T_Θ is significantly larger. This is most likely due to our sampling of a broader energy range, where $E_{max}/N=3.0$ rather than zero. It is also noted that when the shortest chains are excluded from the fitting procedure, the estimates for T_Θ increase slightly, but systematically. In all cases the "best fit" for the R_g^2 data yield a slightly larger value of T_Θ than is obtained from the specific heat. Thus, the error estimates for the values in Table III may be underestimated due to possible, residual finite-size effects for our smallest chain lengths.

The analysis using Eq. (9) to describe the liquid-solid transition data is also shown in Fig. 11; however, finite-size effects previously discussed must be considered. The variations with size below $N=100$ makes it hard to include these results for extrapolation to the infinite-chain limit. In Fig. 7, the $N=55$ specific heat peak is uncharacteristically high compared to the other chain lengths, as well as other magic numbers. This large peak is not novel and can be seen in many similar analyses of clusters [28,29,32]. Understanding this peak may take more analysis of the liquid-solid and solid-solid transitions, and in particular, how these two merge before the $N=55$ chain length. Because of the finite-size effects discussed above, we restrict our analysis to $100\leq N\leq561$ when applying the prediction of Eq. (9). The data for $N\geq100$ appear to be rather smooth and magic number spikes decrease as N increases (see points $N=147$, $N=309$, and 561). Results for this fit are presented in Table III, where $T_{ls}=0.790\pm0.007$ using the C_V/N data and $T_{ls}=0.79\pm0.01$ using the R_g^2 data. These results still offer perspective into previous issues in homopolymer studies. The first is that the coil-globule and liquid-solid transition remain separated in the infinite-chain limit. The second is that our estimate of the liquid-solid transition is significantly lower than that given by Parsons and Williams, providing further evidence that sampling a restricted energy range will yield imprecise results. However, this analysis is further complicated by finite-size effects that have been shown to occur for chain lengths greater than $N=55$ and which are currently beyond our ability to characterize.

V. CONCLUSIONS

Results describing the general behavior of single flexible homopolymer chains have been obtained for a wide range of

TABLE III. Table of transition temperatures for both the liquid-solid (T_{sl}) and the coil-globule (T_θ) transitions. These transitions have been analyzed using both specific heat and radius of gyration data.

Fit function	Data	$T(N\rightarrow\infty)$	a_1, a_2	N
$T(N)-T_\Theta=a_1/\sqrt{N}+a_2/N$	C_V/N	3.156 ± 0.007	$-11.90, 13.98$	$10\rightarrow561$
	R_g^2	3.176 ± 0.004	$-11.56, 12.77$	$10\rightarrow500$
$T(N)-T_{ls}=a_1/N^{1/3}$	C_V/N	0.790 ± 0.007	-1.94	$100\rightarrow561$
	R_g^2	0.79 ± 0.01	-1.9	$100\rightarrow500$

chain lengths. The application of the Wang-Landau algorithm has proven to be an efficient and straightforward method for studying thermodynamic and structural properties of homopolymers. Nonetheless, the few parameters involved require a careful adjustment in order to yield an optimal balance between simulation accuracy and computational effort, e.g., the density of states must be obtained for a sufficiently broad range of energies. Unlike earlier studies of continuous models [6,7], we find only two transitions for similar chains, and as has been shown in studies of a bond-fluctuation model [10], these two transitions remain distinct in the limit of infinite size. We also see a significantly higher T_Θ compared to these off-lattice studies [6,7], most likely due to our broader sampling of energy. The liquid-solid transition is more difficult to analyze because of finite-size effects occurring for chain lengths $30 \leq N \leq 54$, where additional solid-solid transitions associated with transformations in the outer shell of a collapsed chain are visible in thermodynamic and structural quantities. Such effects have been seen in a very recent study of Lennard-Jones homopolymers [3,4], with indications of such behavior in longer chain lengths. However, for $N \geq 100$ variations in behavior with chain length become smoother and we were able to estimate the infinite-chain transition temperature. Our value of T_{ls} is significantly lower than that found in previous off-lattice studies, which we again attribute to energy range effects. Results also show that size dependent effects are prominent and do not behave in a simple, monotonic manner as chain length increases. These finite-size effects must be considered in future studies of single homopolymer chains. Our simulations are at the limit of what is currently possible with current computing resources, and adaptations to our method will be needed in order to perform further analysis.

ACKNOWLEDGMENTS

We thank C. Gervais, S. Tangirala, S.-H. Tsai, S. Schnabel, M. Bachmann, W. Paul, and K. Binder for helpful discussions. This work was supported by Grant No. DMR-0810223.

[1] T. Wüst and D. P. Landau, Phys. Rev. Lett. **102**, 178101 (2009).
[2] T. Vogel, M. Bachmann, and W. Janke, Phys. Rev. E **76**, 061803 (2007).
[3] S. Schnabel, M. Bachmann, and W. Janke, J. Chem. Phys. **131**, 124904 (2009).
[4] S. Schnabel, T. Vogel, M. Bachmann, and W. Janke, Chem. Phys. Lett. **476**, 201 (2009).
[5] M. P. Taylor, W. Paul, and K. Binder, Phys. Rev. E **79**, 050801(R) (2009).
[6] D. F. Parsons and D. R. M. Williams, Phys. Rev. E **74**, 041804 (2006).
[7] D. F. Parsons and D. R. M. Williams, J. Chem. Phys. **124**, 221103 (2006).
[8] B. Xue, J. Wang, and W. Wang, J. Chem. Phys. **119**, 7534 (2003).
[9] F. Calvo, J. P. K. Doye, and D. J. Wales, J. Chem. Phys. **116**, 2642 (2002).
[10] W. Paul, T. Strauch, F. Rampf, and K. Binder, Phys. Rev. E **75**, 060801(R) (2007).
[11] F. Rampf, K. Binder, and W. Paul, J. Polym. Sci., Part B: Polym. Phys. **44**, 2542 (2006).
[12] F. Rampf, W. Paul, and K. Binder, EPL **70**, 628 (2005).
[13] T. Wüst, D. P. Landau, C. Gervais, and Y. Xu, Comput. Phys. Commun. **180**, 475 (2009).
[14] D. T. Seaton, T. Wüst, and D. P. Landau, Comput. Phys. Commun. **180**, 587 (2009).
[15] D. T. Seaton, S. J. Mitchell, and D. P. Landau, Braz. J. Phys. **38**, 48 (2008).
[16] K. Kremer and G. S. Grest, J. Chem. Phys. **92**, 5057 (1990).
[17] C. Bennemann, W. Paul, K. Binder, and B. Dünweg, Phys. Rev. E **57**, 843 (1998).
[18] A. Milchev, A. Bhattacharya, and K. Binder, Macromolecules **34**, 1881 (2001).
[19] D. P. Landau, S.-H. Tsai, and M. Exler, Am. J. Phys. **72**, 1294 (2004).
[20] D. P. Landau and F. Wang, Braz. J. Phys. **34**, 354 (2004).
[21] C. Gervais, T. Wüst, D. P. Landau, and Y. Xu, J. Chem. Phys. **130**, 215106 (2009).
[22] T. Wüst and D. P. Landau, Comput. Phys. Commun. **179**, 124 (2008).
[23] R. E. Belardinelli and V. D. Pereyra, J. Chem. Phys. **127**, 184105 (2007).
[24] R. E. Belardinelli, S. Manzi, and V. D. Pereyra, Phys. Rev. E **78**, 067701 (2008).
[25] C. Zhou, T. C. Schulthess, S. Torbrügge, and D. P. Landau, Phys. Rev. Lett. **96**, 120201 (2006).
[26] M. S. Shell, P. G. Debenedetti, and A. Z. Panagiotopoulos, Phys. Rev. E **66**, 056703 (2002).
[27] G. Marsaglia, J. Mod. Appl. Stat. Methods **2**, 2 (2003).
[28] D. D. Frantz, J. Chem. Phys. **115**, 6136 (2001).
[29] V. A. Mandelshtam and P. A. Frantsuzov, J. Chem. Phys. **124**, 204511 (2006).
[30] B. A. Berg, *Markov Chain Monte Carlo Simulations And Their Statistical Analysis: With Web-based Fortran Code* (World Scientific, Hackensack, NJ, 2004).
[31] I. M. Lifshitz, A. Y. Grosberg, and A. R. Khokhlov, Rev. Mod. Phys. **50**, 683 (1978).
[32] P. A. Frantsuzov and V. A. Mandelshtam, Phys. Rev. E **72**, 037102 (2005).
[33] D. D. Frantz, J. Chem. Phys. **102**, 3747 (1995).
[34] J. A. Northby, J. Chem. Phys. **87**, 6166 (1987).
[35] E. G. Noya and J. P. K. Doye, J. Chem. Phys. **124**, 104503 (2006).
[36] V. A. Sharapov, D. Meluzzi, and V. A. Mandelshtam, Phys. Rev. Lett. **98**, 105701 (2007).
[37] A. L. Mackay, Acta Crystallogr. **15**, 916 (1962).
[38] J. P. K. Doye and F. Calvo, J. Chem. Phys. **116**, 8307 (2002).