A Novel Configurational-Bias Monte Carlo Method for Lattice Polymers: Application to Molecules with Multicyclic Architectures

Charlles R. A. Abreu and Fernando A. Escobedo*

School of Chemical and Biomolecular Engineering, Cornell University, Ithaca, New York 14853

Received April 7, 2005; Revised Manuscript Received July 12, 2005

ABSTRACT: A new configurational-bias Monte Carlo (CBMC) method is developed for lattice polymer simulation. In contrast to usual approaches, no particular structure is assumed a priori. Instead, any regular lattice (whose sites are connected to their neighbors by the same vectors) can be used. The method includes new features such as an algorithm to distinguish all conformations of short ideal chains in regular lattices and a simple selection scheme that can improve general CBMC methods. The method is suitable for simulating molecules with branches and/or closed cycles, allowing the study of polymers with complex, well-defined architectures. It was applied to simulate the conformational and thermal properties of multicyclic molecules in dilute solution. The results indicate systematic changes in the behavior of topologically constrained polymers, such as a shrinking of the radius of gyration, a slight reduction in scaling exponents, and an increase of the globule−coil transition temperatures.

1. Introduction

A major challenge in chemical synthesis is the assembly of macromolecules with precisely controlled architecture.[1,2] Considerable progress has been made in this area during the past years. Advanced techniques such as living free radical polymerization,[3] convergent iterative growth,[3,4] and covalent fixation after electrostatic self-assembly[5] now enable the synthesis of star polymers, comblike and hyperbranched polymers, dendrimers, cyclic and multicyclic polymers, and other topologically complex macromolecules.[6] Nucleic acids have been used to construct such intriguing structures as polyhedrons whose edges are DNA double-helix strands[1,7] and nanoscale fabrics formed by tiny RNA "jigsaw puzzle units".[8] As distinct molecular geometries entail distinct physical properties, these advances open a vast field for the discovery of novel materials, nanodevices, sensors, transportation agents, emulsifiers, and so forth.

Molecular simulations have played an important role in improving our understanding of the conformational statistics and structure−property relations of macromolecules. Most simulation studies to date have dealt with linear molecules, but this focus has been shifting with the increasing practical importance of other architectures. It is known that a topological change as simple as linking the ends of a linear molecule together can drastically alter some of its properties by reducing its conformational entropy.[9,10] Naturally, stronger departures from the linear chain behavior are expected for more intricate architectures. In this context, the present work has two basic purposes: (1) to develop a new, efficient Monte Carlo method to simulate chainlike molecules and (2) to investigate the effect of architectural complexity on the conformational statistics of macromolecules.

In a coarse-grained model, polymers can be viewed as sets of interconnected segments. Three classes of segments can be distinguished, based on the number of bonds: terminal (one bond), internal (two bonds), and cross-linking (three or more bonds) segments. A topological constraint is then defined as a feature that reduces conformational entropy, such as cross-links, closed cycles (absence of chain-ends), permanent knots, and permanent entanglements. Only the first two kinds of constraints, which define the "architecture" or "connectivity" of the chain, will be considered here. The presence of cross-linking segments is the distinguishing characteristic of branched polymers, while the absence of terminal segments is the one of cyclic polymers. It is then natural to anticipate that multicyclic polymers,[6] those which present both features simultaneously, are likely to exhibit unusual properties. This class involves fused rings (e.g., θ-shaped molecules), spiranic rings (e.g., 8-shaped molecules), and rings bridged by linear chains (e.g., manacle-shaped molecules).[6] Except for 8-shaped molecules,[11] only in the past few years have some multicyclic polymers been successfully synthesized,[5,12,13] and most of their conformational properties are yet to be characterized. Therefore, this class of architecture represents a good target for computer simulation studies, and the present work constitutes an initial step in this direction. For this, a family of molecules with increasing number of topological constraints is defined, as shown in Figure 1, by means of representations introduced by Tezuka and Oike.[6,14] The first member is a linear molecule, which has the most unhindered topology and is included as reference. The second is a simple ring, representing the initial step toward multicyclic architectures. The third member is a θ-shaped molecule, which is a ring with two of its segments bridged by a linear chain. Alternatively, on the basis of a route employed by Tezuka et al.[13] to

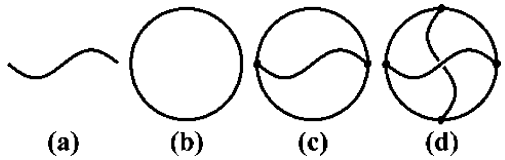

Figure 1. Family of molecules with increasing topological complexity: (a) linear molecule, (b) cyclic molecule, (c) theta-shaped molecule, and (d) crosshair-shaped molecule.

* To whom correspondence should be addressed. E-mail: fe13@cornell.edu.

10.1021/ma050725t CCC: $30.25
Published on Web 09/08/2005

synthesize θ-shaped polytetrahydrofuran, it can be conceived as a three-armed star polymer whose termini were connected together as a trifunctional cross-link. The fourth and most complex family member will be referred to as a "crosshair-shaped molecule". Although it appears that such architecture has not been synthesized thus far, it can be conceived as a three-armed star whose termini were bridged to each other by linear chains, summing up four trifunctional cross-links. In the notation introduced by Tezuka and Oike,[6,14] the family members are referred to as $A_3(2,0)$, $I_3(0,0)$, $II_4(0,2)$, and $III_4(0,4)[0^a\text{-}0^b\text{-}0^a\text{-}0^b]$.

The conformational behavior of a fully flexible linear chain in dilute solution is well-known.[15,16] In a good solvent, it behaves like a random coil, manifesting only excluded-volume effects, and its conformations are consistent with those of a self-avoiding random walk. In a poor solvent, the molecule collapses and generally becomes globular in shape; this can be modeled as a chain with strongly attractive segments, the solvent effects being implicit. In between these two situations a solvent quality exists in which excluded-volume and attraction effects cancel out, and the chain behaves as if no volume exclusion existed at all. In this state, often called the theta condition, the molecule can be suitably modeled as an ideal chain with conformational statistics of a random walk. It is expected that a molecule with topological constraints will certainly exhibit a different behavior, even with fully flexible bonds. For example, it is known that the mean-square radius of gyration of an ideal linear chain is twice as large as that of an ideal cyclic chain with the same number of bonds.[16] A question then arises as how the transition from good to poor solvent conditions (coil−globule transition) occurs for more complex molecules, such as multicyclic chains.

A common simplification in coarse-grained polymer modeling is to consider that segments are not free to move through coordinate space but that their positions are confined to sites on a lattice. With this approach, the possible conformations become enumerable, which considerably reduces the amount of computation necessary to obtain good statistics. The simple cubic (sc) lattice, which has found widespread applications,[17−19] has coordination number $Z = 6$ and can only support chains with fixed bond length. Besides, only three bending angles (0°, 90°, and 180°) and four torsional angles (0°, 90°, 180°, and 270°) are possible. Other lattices with fixed bond length but higher coordination numbers and more possible angles are also employed. A relevant example is the face-centered-cubic (fcc) lattice, sometimes referred to as the second-nearest-neighbor diamond (2nnd) lattice.[20] In fact, this lattice (with $Z = 12$) has been used to simulate melts and dilute solutions of polyethylene[21,22] and other real polymers[23,24] by means of the rotational isomeric state model,[15] where each chain segment corresponds to every other monomer of a real molecule. Moreover, it has been shown that the spatial distribution of residues in folded proteins presents regularities that conform to the fcc lattice,[25,26] suggesting that it could be suitable for coarse-grained model of proteins. Also, high-coordination lattices with variable bond length are often used to simulate polymer systems, the so-called bond fluctuation model (BFM)[27,28] being a notable example.

The lattice approach is adopted in the present work, but no particular structure is assumed a priori. Instead, the only imposed restriction is that all sites must be connected to their adjacent ones by the same fixed set of base vectors, $\{\mathbf{b}_j\}_{j=1,...,Z}$, where Z is the coordination number. Structures that conform to this restriction will be referred to as "regular lattices". Note that such base vectors are conceptually distinct from those known as "basis vectors" in crystallography. Unlike the latter, the former also incorporate concepts of connectivity and adjacency. Therefore, two sets of vectors are considered different here if they establish different connectivities, even if they generate the same lattice positions and are thus crystallographically identical. Note that the examples cited above constitute regular lattices as well as other common structures such as the bcc lattice and the cubic lattice with straight and diagonal bonds[29] ($Z = 26$). In the special case of proteins, most lattices cited in a recent review by Kolinski and Skolnick[30] are also regular. In this work, a general Monte Carlo method is developed for efficient conformational sampling in regular lattices of chains with cross-links and closed cycles. This method is then used to study conformational properties of the molecules of Figure 1 when isolated and confined to an fcc lattice. The choice of the fcc structure is based on a balance between simplicity and flexibility and also on the relative success of such lattice in modeling real polymers.[21−24]

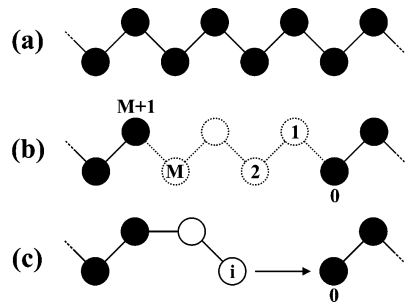

Figure 2. Scheme of the regrowth procedure characteristic of CBMC-FE methods.

The method advocated here belongs to the class of configurational-bias Monte Carlo methods with fixed end points (CMBC-FE),[31] which have been proposed for specific lattices[17,32] and for continuous-space systems.[33−36] In a method of this sort, conformational changes involve removing an internal section of the chain and then sequentially relocating its segments until the chain is restored in a new conformation. This procedure is called "rebridging" or "regrowing" and is illustrated in Figure 2. For efficiency, a biasing weight is introduced during the sequential relocation of segments to increase the probability that the chain connectivity is restored at the end. Given the similarity with the method of Chen and Escobedo,[36] the present method will be referred to as rebridging configurational-bias method for regular lattices, or RCB-RL method. It includes some new features such as an algorithm for distinguishing all possible conformations of (short) ideal chains confined to regular lattices and a scheme that can be used for increasing the efficiency of general CBMC methods.

2. Method Description

2.1. Regrowing Process. Let a regular lattice be defined by a set of Z base vectors $\{\mathbf{b}_j\}_{j=1,...,Z}$. Given a chain confined to this lattice, consider the situation in which an internal, linear section with M segments is randomly selected to be removed and then regrown. An index "1" is then assigned to one of the extreme segments of such a section (chosen at random), and

indexes "2" to "M" are assigned to the others. Additionally, indexes "0" and "$M + 1$" are assigned to the out-of-section segments connected to those with indexes "1" and "M", respectively. For the sake of generality, it is assumed that the chain is a heteropolymer and that every segment with index "i" has a type-distinguishing tag τ_i. The indexing scheme is illustrated in Figure 2b. Because the chain is constrained to a regular lattice with base vectors $\{\mathbf{b}_j\}_{j=1,...,Z}$, the positions of segments with consecutive indexes will always be linked by one of these vectors, that is

$$\mathbf{r}_i = \mathbf{r}_{i+1} + \mathbf{b}_{q_i} \tag{1}$$

where $\mathbf{r}_i$ and $\mathbf{r}_{i+1}$ are the positions of segments "i" and "$i + 1$", respectively, and q_i is some integer between 1 and Z. Note that, by convention, the vector $\mathbf{b}_{q_i}$ points from "$i + 1$" to "i". Thus, once one knows where segment "$M + 1$" is located, the positions of all segments of the section are fully determined by a set of M integers, $\{q_i\}_{i=1,...,M}$, all of them between 1 and Z. Of course, the fact that segment "1" is connected to segment "0" implies that such integers must satisfy the constraint

$$\mathbf{r}_{M+1} + \sum_{i=1}^{M} \mathbf{b}_{q_i} + \mathbf{b}_k = \mathbf{r}_0 \tag{2}$$

where k is also some integer between 1 and Z. Therefore, the regrowing process corresponds to replacing an original set $\{q_i^{\mathrm{o}}\}_{i=1,...,M}$ (the "old" conformation) by a new one, $\{q_i^{\mathrm{n}}\}_{i=1,...,M}$, whose elements must conform to eq 2. To select the elements q_i^{n} efficiently, a configurational-bias sampling is used similar to that proposed by Dijkstra et al.[17]

To simplify the notation, the regrowing is carried out in the reverse order of indexes; that is, q_M^{n} is chosen at first, then q_{M-1}^{n}, and so on. In other words, the new position for a segment "i", denoted by $\mathbf{r}_i^{\mathrm{n}}$, is proposed by adding one of the base vectors to the (already established) position $\mathbf{r}_{i+1}^{\mathrm{n}}$. As a consequence, after relocating a given segment "i", exactly i new bonds are needed to restore the chain. Let $g_i(\mathbf{r})$ be defined as the probability that a linear section formed by i bonds (base vectors) results in an end-to-end vector $\mathbf{r}$. Then, for a given integer $j \in [1,Z]$, $g_i(\mathbf{r}_{i+1}^{\mathrm{n}} + \mathbf{b}_j - \mathbf{r}_0)$ measures the likelihood of ultimately restoring the chain connectivity if $\mathbf{r}_{i+1}^{\mathrm{n}} + \mathbf{b}_j$ is chosen as the new position of segment "i". Therefore, the sampling can be improved if those integers which give the largest values of $g_i(\mathbf{r}_{i+1}^{\mathrm{n}} + \mathbf{b}_j - \mathbf{r}_0)$ are favored during the choice of q_i^{n}. Of course, $g_i(\mathbf{r})$ is generally not known a priori, but for now one can assume that some approximation is available (as discussed in the next section). Since conformations with high Boltzmann weights may also be preferred,[37] an integer j is then chosen as the value of q_i^{n} with probability

$$p_j^{(i)} = \frac{\exp[-\beta u(\tau_i,\mathbf{r}_{i+1}^{\mathrm{n}} + \mathbf{b}_j)] g_i(\mathbf{r}_{i+1}^{\mathrm{n}} + \mathbf{b}_j - \mathbf{r}_0)}{w_i^{\mathrm{n}}}, \quad j \in [1,Z] \tag{3}$$

where $\beta = (k_{\mathrm{b}}T)^{-1}$, k_{b} is the Boltzmann constant, T is the system temperature, $u(\tau,\mathbf{r})$ is the potential energy change associated with the placement of a segment of type τ in a position $\mathbf{r}$ (including interactions with previously restored segments "M" down to "$i + 1$"), and w_i^{n} is a normalization factor, given by

$$w_i^{\mathrm{n}} = \sum_{j=1}^{Z} \exp[-\beta u(\tau_i,\mathbf{r}_{i+1}^{\mathrm{n}} + \mathbf{b}_j)] g_i(\mathbf{r}_{i+1}^{\mathrm{n}} + \mathbf{b}_j - \mathbf{r}_0) \tag{4}$$

It is possible that, in the course of the regrowing process, no positions are feasible for placing a given segment "i" ($p_j^{(i)} = 0$ for all $j \in [1,Z]$). If so, the process is aborted and the old conformation is kept. Otherwise, once all $\{q_i^{\mathrm{n}}\}_{i=1,...,M}$ are determined, eq 2 is automatically satisfied since $g_1(\mathbf{r}) \neq 0$ only if $\mathbf{r}$ is a base vector. Because of this feature, $g_i(\mathbf{r})$ is called a "guiding weight function". Finally, detailed balance[31] is satisfied if the proposed move is accepted with probability

$$P_{\mathrm{acc}}(\mathrm{o} \rightarrow \mathrm{n}) = \min\left\{1, \frac{\alpha(\mathrm{n} \rightarrow \mathrm{o})}{\alpha(\mathrm{o} \rightarrow \mathrm{n})} \exp[-\beta \Delta U(\mathrm{o} \rightarrow \mathrm{n})]\right\} \tag{5}$$

where "o → n" and "n → o" respectively denote the proposed move and its reverse, α is the probability that a given move is proposed using the described procedure, and $\Delta U(\mathrm{o} \rightarrow \mathrm{n})$ is the potential energy variation caused by the proposed move. It is possible to calculate α(n → o) from eq 3, since

$$\alpha(\mathrm{o} \rightarrow \mathrm{n}) = \prod_{i=1}^{M} p_{q_i^{\mathrm{n}}}^{(i)} = \frac{\exp(-\beta U^{\mathrm{n}}) G^{\mathrm{n}}}{W^{\mathrm{n}}} \tag{6}$$

where

$$U^{\mathrm{n}} = \sum_{i=1}^{M} u(\tau_i,\mathbf{r}_i^{\mathrm{n}}),\ W^{\mathrm{n}} = \prod_{i=1}^{M} w_i^{\mathrm{n}},\ \text{and}\ G^{\mathrm{n}} = \prod_{i=1}^{M} g_i(\mathbf{r}_i^{\mathrm{n}} - \mathbf{r}_0) \tag{7}$$

To obtain α(n→o), the original configuration must be retraced.[31] For this, the regrowing procedure is repeated but always choosing q_i from the original set $\{q_i^{\mathrm{o}}\}_{i=1,...,M}$. One then finds

$$\alpha(\mathrm{n} \rightarrow \mathrm{o}) = \prod_{i=1}^{M} p_{q_i^{\mathrm{o}}}^{(i)} = \frac{\exp(-\beta U^{\mathrm{o}}) G^{\mathrm{o}}}{W^{\mathrm{o}}} \tag{8}$$

where U^{o}, W^{o}, and G^{o} are the analogues of U^{n}, W^{n}, and G^{n} for the retracing process. Note that $\Delta U(\mathrm{o} \rightarrow \mathrm{n}) = U^{\mathrm{n}} - U^{\mathrm{o}}$. Substituting (6) and (8) into (5) yields

$$P_{\mathrm{acc}}(\mathrm{o} \rightarrow \mathrm{n}) = \min\left\{1, \frac{W^{\mathrm{n}} G^{\mathrm{o}}}{W^{\mathrm{o}} G^{\mathrm{n}}}\right\} \tag{9}$$

Since the efficiency of the regrowing process relies on the ability to produce a new conformation as uncorrelated as possible to the old one, an alternative scheme is proposed that only considers trial moves in which all segments are displaced from their old positions. This

can be achieved by replacing eq 3 by

$$p_j^{(i)} = \begin{cases} \exp[-\beta u(\tau_i, \mathbf{r}_{i+1}^{\mathrm{n}} + \mathbf{b}_j)] \times \\ \qquad g_i(\mathbf{r}_{i+1}^{\mathrm{n}} + \mathbf{b}_j - \mathbf{r}_0)/w_i^{\mathrm{n}} & \text{if } \mathbf{r}_{i+1}^{\mathrm{n}} + \mathbf{b}_j \neq \mathbf{r}_i^{\mathrm{o}} \\ 0 & \text{if } \mathbf{r}_{i+1}^{\mathrm{n}} + \mathbf{b}_j = \mathbf{r}_i^{\mathrm{o}} \end{cases} \tag{10}$$

where $\mathbf{r}_i^{\mathrm{o}}$ is the position of segment "i" in the old conformation. The retracing must follow the same procedure to ensure detailed balance. This scheme can be occasionally too restrictive and be less efficient than the conventional one. Of course, less restrictive schemes can be achieved if only select segments are not allowed to occupy their old positions. Note also that similar schemes can be applied to (a) continuous-space systems by defining a threshold to avoid the choice of new positions in the close vicinity of the original ones and (b) to configurational-bias methods with no fixed end points.[37,38]

2.2. Guiding Weights. Chain restoration is guaranteed, after M regrowing steps are completed, if a function $g_i(\mathbf{r})$ is nonzero for any $\mathbf{r}$ obtainable by summing i base vectors or zero otherwise. The definition of $g_i(\mathbf{r})$ given in the previous section satisfies this condition, but other definitions could be adopted. In fact, although it has been thought that actual end-to-end vector probabilities might be optimal guiding weights, they could favor correlated conformations and reduce efficiency. On the other hand, finding optimal guiding weights can be impractical, and in this sense, it is reasonable to look for approximations based on the definition given above.

Different proposals of guiding weights have been reported in the literature. For the case of a simple cubic (sc) lattice, Dijkstra et al.[17] employed the number of distinct random walks with which one can go from the origin to a position $\mathbf{r}$ by using i steps, which is equivalent to define $g_i(\mathbf{r})$ as the probability that an ideal chain with $i + 1$ segments, confined to a sc lattice, has an end-to-end vector equal to $\mathbf{r}$. They provided an analytical formula for the number of random walks, but it is not trivial to generalize it for the case of an arbitrary regular lattice. Besides, while this approach is most suitable for simulating ideal chains, others can be more effective for chains with volume exclusion and interacting segments.

Let $P_i(\mathbf{r})$ be the analogue of $g_i(\mathbf{r})$ for continuous space systems, i.e., a probability density of end-to-end vectors for chain sections with i bonds. On the basis of the work of Dijkstra et al.,[17] Vendruscolo[34] proposed that $P_i(\mathbf{r})$ can be approximated by $P_i^{\mathrm{id}}(\mathbf{r})$, the probability distribution corresponding to ideal chains with fixed bond length. Analytical solution for $P_i^{\mathrm{id}}(\mathbf{r})$ does exist, but this function could be of limited usefulness for simulating nonideal chains. For the case of chains with strong interactions, Chen and Escobedo[36] suggested that $P_i(\mathbf{r})$ could be better approximated by $P_i^0(\mathbf{r})$, an ensemble-averaged probability distribution obtained from a preliminary simulation of a system with a simplified interaction potential. The authors obtained $P_i^0(\mathbf{r})$, for different numbers of bonds i, by collecting histograms of end-to-end distances of the corresponding chain sections. Once normalized, each histogram is an estimate of $\bar{P}_i^0(r)$, the probability distribution of end-to-end distances (not vectors) for sections with i bonds. Then, under the hypothesis of no directional preference, they obtained $P_i^0(\mathbf{r})$ by using the relation $\bar{P}_i^0(|\mathbf{r}|) = 4\pi|\mathbf{r}|^2 P_i^0(\mathbf{r})$.

Wick and Siepmann[35] proposed a method akin to that of Chen and Escobedo[36] but using an estimate of $\bar{P}_i(r)$ (which is analogous to $P_i(\mathbf{r})$ but based on end-to-end distances) directly as a guiding weight function. This is correct since $\bar{P}_i(|\mathbf{r}|)$ can properly guide the regrowing process to the fixed end point. A novel feature of Wick and Siepmann's method[35] is a periodic adjustment of the guiding weights throughout the equilibration part of a simulation. However, as mentioned earlier, the actual probabilities of the considered system may not be optimal guiding weights.

For an arbitrary regular lattice, the definition of $g_i(\mathbf{r})$ corresponds to

$$g_i(\mathbf{r}) = \left\langle \frac{n_i(\mathbf{r})}{N_i} \right\rangle \tag{11}$$

where N_i is the total number of linear sections with i bonds present in the system and $n_i(\mathbf{r})$ is the number of section with i bonds that have end-to-end vector equal to $\mathbf{r}$. The angled brackets denote ensemble average. Therefore, an approximation for $g_i(\mathbf{r})$ can be obtained by collecting end-to-end vector histograms during a preliminary simulation. However, it is more convenient to collect (one-dimensional) histograms of end-to-end distances or, even better, square distances (to avoid frequent calculation of square roots); the latter gives $\bar{g}_i(r^2)$, the probability of finding a section with square end-to-end distance equal to r^2 among all sections with i bonds:

$$\bar{g}_i(r^2) = \left\langle \frac{\bar{n}_i(r^2)}{N_i} \right\rangle \tag{12}$$

where $\bar{n}_i(r^2)$ is the number of linear sections present in the system with both i bonds and square end-to-end distance r^2. To relate $g_i(\mathbf{r})$ to $\bar{g}_i(r^2)$, let eq 11 be rewritten as

$$g_i(\mathbf{r}) = \left\langle \frac{n_i(\mathbf{r})}{\bar{n}_i(|\mathbf{r}|^2)} \frac{\bar{n}_i(|\mathbf{r}|^2)}{N_i} \right\rangle \tag{13}$$

where $|\mathbf{r}|$ represents the norm of $\mathbf{r}$. Assuming that the two ratios in the equation above are uncorrelated (null covariance) leads to

$$g_i(\mathbf{r}) = \left\langle \frac{n_i(\mathbf{r})}{\bar{n}_i(|\mathbf{r}|^2)} \right\rangle \left\langle \frac{\bar{n}_i(|\mathbf{r}|^2)}{N_i} \right\rangle \tag{14}$$

The rightmost bracketed term above is equal to $\bar{g}_i(|\mathbf{r}|^2)$. The other bracketed term represents the probability that a section with i bonds results in an end-to-end vector $\mathbf{r}$, given that its square end-to-end distance is $|\mathbf{r}|^2$. Assuming no directional preference, i.e., that all end-to-end vectors with the same square norm occur with the same probability, the referred term will no longer depend on details of the simulated chains, but only on the adopted lattice structure; a way of obtaining this term is discussed below.

For a lattice with coordination number Z, the number of possible sequences of i base vectors is Z^i, which corresponds to the number of distinct sequences of i integers between 1 and Z, i.e., $\{q_j\}_{j=1,\ldots,i}$ where $1 \leq q_j \leq$

Z for all j. By distinguishing every possible sequence and calculating its associated sum of base vectors, one can gather useful topological information about the lattice, such as the set of all possible distinct end-to-end vectors $\mathbf{r}$ and the set of all possible distinct square end-to-end distances r^2, both obtainable with i base vectors. Such sets will be denoted by $\{\mathbf{r}\}_i$ and $\{r^2\}_i$, respectively. In fact, because sequences formed by the same set of integers will always result in the same end-to-end vector (no matter the order), it is sufficient to distinguish the C_i^{Z+i-1} possible "combinations with repetitions" of i integers between 1 and Z instead of all Z^i different sequences. The number of such combinations is given by

$$C_i^{Z+i-1} = \frac{(Z+i-1)!}{i!(Z-1)!} \tag{15}$$

For each element of $\{r^2\}_i$, let $\{\mathbf{r}|r^2\}_i$ be a set of all vector elements of $\{\mathbf{r}\}_i$ whose square norms are equal to r^2 [the union of all such sets for a given i is equal to $\{\mathbf{r}\}_i$]. With this definition, "no directional preference" means that all elements of a given set $\{\mathbf{r}|r^2\}_i$ have the same probability of being sampled. This probability is equal to $1/N_{\text{elem}}(\{\mathbf{r}|r^2\}_i)$, the inverse of the number of elements in $\{\mathbf{r}|r^2\}_i$. Equation 14 can now be written as

$$g_i(\mathbf{r}) = \frac{\bar{g}_i(|\mathbf{r}|^2)}{N_{\text{elem}}(\{\mathbf{r}|r^2\}_i)} \tag{16}$$

With eq 16, approximations for the functions $g_i(\mathbf{r})$ can be obtained from histograms of r^2, collected during a preliminary simulation. For this, knowledge of all sets $\{\mathbf{r}|r^2\}_i$ corresponding to $i = 1, ..., M$ is necessary, thus requiring evaluation of the C_i^{Z+i-1} combinations of integers. Note that a previous knowledge of the sets $\{r^2\}_i$ can be very useful for the collection of histograms, since they provide all the discrete values of r^2 that can possibly be sampled in a simulation.

Before showing how to obtain the sets $\{r^2\}_i$ and $\{\mathbf{r}|r^2\}_i$, it is convenient to consider the case of ideal-chain guiding weights, denoted by $g_i^{\text{id}}(\mathbf{r})$. In an ideal chain all possible conformations are equally likely; i.e., the frequency with which a given r^2 occurs in an infinitely long simulation of an ideal chain with $i + 1$ segments is proportional to the number of possible conformations with such a r^2. Because of this fact, a function like $\bar{g}_i^{\text{id}}(r^2)$ [defined as in eq 12 for an ideal chain] is also a topological property of the considered lattice and, as such, depends only on the corresponding set of base vectors. Using $\eta_i(r^2)$ to denote the number of distinct sequences (not only combinations) of i integers between 1 and Z which result in a sum of base vectors with square norm r^2 and knowing that the overall number of integer sequences is Z^i, one can conclude that

$$\bar{g}_i^{\text{id}}(r^2) = \frac{\eta_i(r^2)}{Z^i} \tag{17}$$

Note that the number of sequences that come from a given "combination with repetitions" is often referred to as a "number of permutations with repetitions" and can be easily calculated. It is then sufficient to examine all distinct combinations (instead of all distinct sequences) of i base vectors for computing $\{r^2\}_i$, $\{\mathbf{r}|r^2\}_i$, and $\eta_i(r^2)$. An algorithm for this purpose, whose required inputs are just i, Z, and a set $\{\mathbf{b}_j\}_{j=1,...,Z}$, is shown in Figure 3.

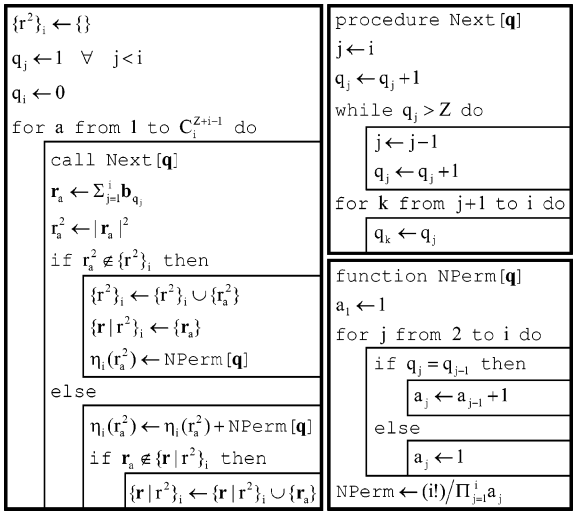

Figure 3. Algorithm for obtaining topological properties of a regular lattice. The symbol ← is an assignment operator. Blocks following "for a from b to c do" are executed only if $b \le c$.

Notice that the algorithm in Figure 3 enumerates exactly the whole conformational phase-space of an ideal chain with $i + 1$ segments that is confined to a regular lattice. If the ideal model is satisfactory for generating guiding weights, then no preliminary simulation is required. In such a case, one just needs to execute the algorithm once for every $i = 1, ..., M$, calculate $\bar{g}_i^{\text{id}}(r^2)$ by eq 17 and then $g_i^{\text{id}}(\mathbf{r})$ by eq 16, and finally save $g_i^{\text{id}}(\mathbf{r})$ for future simulations. If fact, if $\bar{g}_i^{\text{id}}(r^2)$ was used instead, the RCB-RL method would be equivalent to the method of Dijkstra et al.,[17] generalized for regular lattices. However, it should be stressed that using $g_i^{\text{id}}(\mathbf{r})$ seems more suitable and can be done with virtually no additional effort (a comparison of the two approaches is given in the Results section).

2.3. Extension to Molecules with Free Chain Ends and to Molecules with Cross-Linking Segments. As described hitherto, the RCB-RL method is only applicable to simple cyclic molecules because it overlooks the possibility of free chain-ends and/or cross-linking segments. For the first case, a method akin to the original CBMC method (with no fixed end point) should be used. For this, let segment "1" be a free chain end and segment "M" be attached to the "$M + 1$" segment. The regrowing process should proceed without a guiding function, i.e., $g_i(\mathbf{r}_{i+1}^{\text{n}} + \mathbf{b}_j - \mathbf{r}_0)$ appearing in eqs 3, 4, and 10 should be set to 1.

There are different possible ways of extending the method for rearranging branched sections (containing a cross-linking segment). The approach adopted here is based on that of Wick and Siepmann.[35] First, an index "0" is assigned to the cross-linking segment (the one attached to more than two other segments). Next, each branch connected to it is randomly assigned an index "k", which ranges from 1 to N_{B} (≥ 3). Then, for every branch "k", the segment connected to the cross-link receives an index "1", the next one receives an index "2", and so on. After that, the cross-linking segment plus M successive segments of each branch are removed, adding up to $1 + N_{\text{B}}M$ segments removed. After that, branch "1" is regrown by using an algorithm similar to

that in section 2.1. However, as there are $N_B - 1$ fixed end points to which the regrowing process must proceed, the probability of choosing a given base vector $\mathbf{b}_j$ to be the new link between segments "$i + 1$" and "i" of branch "1" is given by

$$p_j^{(i)} = \frac{\exp[-\beta u(\tau_i, \mathbf{r}^n_{i+1,1} + \mathbf{b}_j)] \prod_{k=2}^{N_B} g_{i+M+1}(\mathbf{r}^n_{i+1,1} + \mathbf{b}_j - \mathbf{r}_{M+1,k})}{w_i^n} \quad (18)$$

where $\mathbf{r}^n_{i+1,1}$ is the already chosen new position of segment "$i + 1$" of branch "1" and $\mathbf{r}_{M+1,k}$ is the position of the fixed segment "$M + 1$" of a branch "k". A process using eq 18 is carried out until the cross-linking segment (segment "0") is relocated. After that, the original method can be used to regrow the remaining branches, one after the other, always having segment "0" as a fixed end point. The alternative scheme introduced by eq 10 (i.e., avoiding the selection of old positions) can also be applied for sections with free chain ends or cross-linking segments.

The Appendix gives further details on the implementation of the proposed method.

3. Results and Discussion

The conformational behavior of homopolymers in dilute solutions is studied by means of single-molecule systems with implicit solvent effects. For this purpose, a computer code was written assuming contact potentials and integer base vectors as described in the Appendix. First, the validity and the efficiency of the proposed method are evaluated, and then results for the conformational statistics of multicyclic molecules are presented. Both ideal and self-avoiding chains are considered. Ideal chains are meant to represent molecules in theta-solvent conditions. In the case of self-avoiding chains, given the lattice structure and the molecule connectivity, the solvent quality is determined by the product $\beta\epsilon$, where β is the inverse temperature and ϵ is the effective energy of contact between segments. The number of RCB-RL moves necessary to yield reliable statistics varies from case to case. In this work, such a number was determined for every particular simulation by means of the "blocking method" described by Flyvbjerg and Petersen.[31,39] Typically, 10^5 moves were more than sufficient for small ideal chains, while up to 5×10^8 moves were necessary for long, strongly interacting chains.

3.1. Validity and Efficiency Tests. To evaluate the validity of the RCB-RL method, simulations of ideal linear chains confined in sc, bcc, and fcc lattices (as described in the Appendix) were carried out. The resulting histograms of square end-to-end distances of internal sections were compared to exact probability distributions provided by the algorithm of Figure 3 with eq 17. Total agreement was observed in all cases. Figure 4 contains the exact distribution of three-bonds square end-to-end distances in an fcc lattice as well as the corresponding histogram obtained for a chain with 30 segments. The length unit of the values shown in the horizontal axis are consistent with a fixed bond length equal to $2^{1/2}$, which is the one obtained by using the 12 permutations and sign combinations of (±1,±1,0) as base vectors. The same length unit will be used throughout this paper. For comparison, a biased simulation was performed by using an "uncorrected" Metropolis acceptance rule. The histogram obtained in this simulation is also shown in Figure 4. The observed discrepancy with respect to the exact distribution confirms the necessity of using the proper rule given by eq 9.

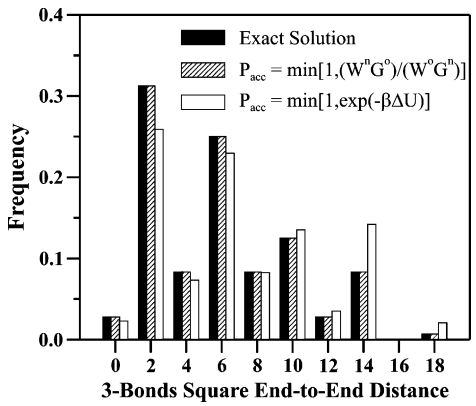

Figure 4. Comparison between simulation results and the exact solution for the probability distribution of three-bonds square end-to-end distances of ideal chains confined in the fcc lattice.

To evaluate the efficiency of the proposed method, a benchmark system is employed consisting of a cyclic chain of 30 segments and constrained to an fcc lattice. Three distinct solvent conditions are considered: (1) the theta condition, represented by an ideal chain; (2) a good solvent condition, corresponding to a self-avoiding chain with $\beta\epsilon = 0$ (athermal condition); and (3) a poor solvent condition, represented by a highly attractive ($\beta\epsilon = -1$) self-avoiding chain. A standard "flip" move was also implemented for comparison. In such a flip move, a segment is selected at random and a list of viable new positions is created. Such a list is formed by lattice sites which are adjacent to both bonded neighbors of the selected segment (excluding its current site and any occupied site for non ideal chains). Finally, one of the sites in the list (if any) is randomly chosen, and the move is accepted with probability given by min{1,exp($-\beta\Delta U$)}, where ΔU is the corresponding energy change. The efficiency of a method can be defined as being inversely proportional to the CPU time necessary to obtain uncorrelated conformations. It should be stressed that CPU time, as opposed to number of Monte Carlo steps, must be considered because different methods entail different computational effort per step. Such a "correlation CPU time" can be assessed by observing the autocorrelation function of half-chain end-to-end vectors,[36] defined as $\langle \mathbf{u}(t) \cdot \mathbf{u}(0) \rangle$, where t stands for CPU time and $\mathbf{u}$ is any unit vector connecting every segment "j" to segment "$j + N/2$" (for the benchmark system, $N = 30$ and j varies from 1 to 15). This function is unity when $t = 0$, and the rate with which it decays toward zero is a measure of the method's efficiency in generating uncorrelated conformations.[36] Exponential decay was observed in all tests performed, i.e.

$$\langle \mathbf{u}(t) \cdot \mathbf{u}(0) \rangle = A \exp(-\gamma t) \quad (19)$$

where A and γ are constants obtained by least-squares regression. In principle, A should be equal to unity if the decay were perfectly exponential. However, a faster decay was generally observed for very low values of t, and the regression quality improved by discarding a

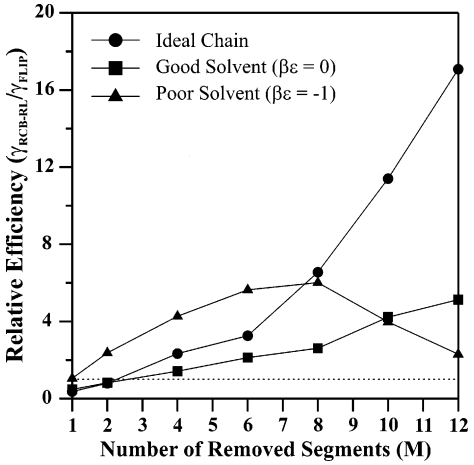

Figure 5. Effect of the number of segments removed per step (M) on the efficiency of the RCB-RL method (with eq 3 for $p_j^{(i)}$, $g_i^{id}(\mathbf{r})$, and $N_t = Z$) relative to flip moves.

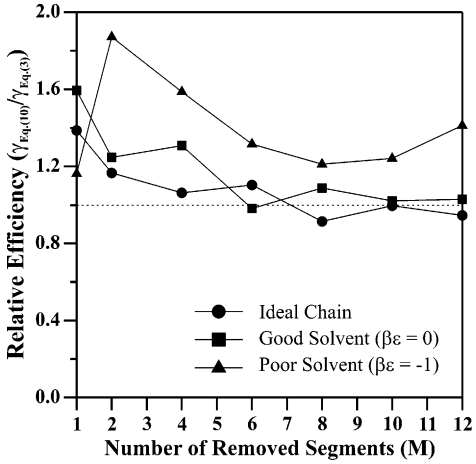

Figure 6. Analysis of the efficiencies attained with the RCB-RL method for the same systems as in Figure 5 but using eq 10 for calculating $p_j^{(i)}$.

small interval of t's close to zero and allowing A to be smaller than one. Since γ in eq 19 is the inverse of the correlation time, it provides a suitable measure of efficiency.

The efficiency γ was calculated for the benchmark system in all three solvent conditions by using either flip moves (γ_{FLIP}) or RCB-RL moves (γ_{RCB-RL}). Since the absolute values of γ are computer-dependent, all tests were carried out in the same machine (Athlon MP 2800) to allow direct comparison. The obtained values of γ_{FLIP} for the ideal, good, and poor solvent conditions were respectively 203.4, 49.15, and 3.612 s^{-1}, showing that decreasing $\beta\epsilon$ drastically increases the amount of CPU time required for accurate results.

Several parameters can affect the efficiency of the RCB-RL method, such as the number of removed segments (M), the guiding weight functions [$g_i(\mathbf{r})$], the number of trial base vectors in each step (N_t in the Appendix), and whether eq 3 or eq 10 is used to calculate placement probabilities. In what follows, each of these factors is studied in turn. First, the effect of M is evaluated. For this, the method is used with eq 3, ideal-chain guiding functions, and $N_t = Z$ (all base vectors). Figure 5 shows the relative efficiencies $\gamma_{RCB-RL}/\gamma_{FLIP}$ for $M = 1$ to $M = 12$. For the ideal-chain and the good-solvent condition, this ratio increases with M along this whole range, and RCB-RL moves are more effective than simple flip moves when $M \geq 4$. For $M \leq 2$, however, the favorable biases promoted by the RCB-RL method do not compensate the required additional effort. For the strongly attractive chain, $\gamma_{RCB-RL}/\gamma_{FLIP} > 1$ in the whole range of M, but a maximum is observed. This is because two factors compete when M increases. On one hand, more effective rearrangements are attained when the chain is restored. On the other hand, fewer successful restorations occur due to the compact structure of the attractive molecules. According to Figure 5, the best compromise between these two factors occurs when $M = 8$, where the RCB-RL moves were 6 times more effective than the flip moves.

In section 2.1, eq 10 was introduced as an alternative to eq 3 to help reduce the correlation between consecutive conformations and thus increase efficiency. To verify these expectations, simulations at the same conditions as those in Figure 5 were executed, but employing eq 10 instead of eq 3. The results are shown in Figure 6. For the ideal chain and the athermal self-avoiding chain, improvements take place for $M = 1$ up to $M = 4$. The superiority of flip moves observed in Figure 5 for $M \leq 2$ no longer exists, while eq 3 and eq 10 seem to be equally effective for $M > 4$. For the attractive chain, eq 10 provides an improvement over the whole range of M. Comparing Figure 5 and Figure 6, one concludes eq 10 with $M = 6$ is the most efficient case relative to flip moves ($\gamma_{RCB-RL}/\gamma_{FLIP} = 7.4$). In conclusion, Figure 6 shows that, at least for the isolated homopolymer under consideration, it is advantageous to replace eq 3 by eq 10 in the RCB-RL method. Thus, the upcoming tests were done using eq 10.

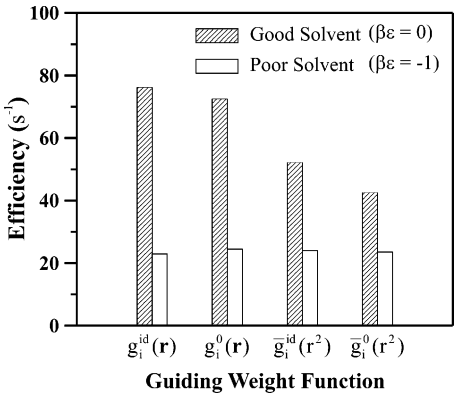

Figure 7. Effect of the guiding weight function on the efficiency of a RCB-RL simulation.

The effect of $g_i(\mathbf{r})$ over efficiency is evaluated next. For each system, both ideal-chain guiding functions, namely $g_i^{id}(\mathbf{r})$, and those obtained from a previous simulation of the same system, namely $g_i^0(\mathbf{r})$, are considered. These functions are based on distribution probabilities of end-to-end vectors. It was mentioned earlier that using such kind of distribution seems more reasonable than using end-to-end distance distributions. This idea is tested here by calculating the efficiency of simulations using $\bar{g}_i^{id}(r^2)$ and $\bar{g}_i^0(r^2)$. These are respectively the analogues of $g_i^{id}(\mathbf{r})$ and $g_i^0(\mathbf{r})$ that are based on end-to-end distance probabilities. Figure 7 shows efficiencies of simulations in either good or poor solvent conditions using the four mentioned guiding functions. In all cases, $M = 4$ and $N_t = Z = 12$. For the good-solvent condition, the simulation using $g_i^{id}(\mathbf{r})$ is slightly more efficient than that using $g_i^0(\mathbf{r})$. This result is specific to the system under consideration, but

it shows that the actual end-to-end vector probability density of the simulated system is not necessarily the best choice of guiding weight. In addition, simulations using functions based on end-to-end vectors are systematically more efficient than those using their end-to-end distance analogues. On the other hand, all four guiding functions result in nearly identical efficiencies in the case of poor-solvent condition. This happens because, during the selection of a new position for a segment, differences in Boltzmann factors are far more influential than differences in guiding probabilities. In this case, $g_i(\mathbf{r})$ merely ensures the restoration of the chain connectivity, with no influence on the method's efficiency.

The last parameter whose effect on efficiency was examined is N_t, the number of trial base vectors in each placement step. For this, several simulations were carried out using different values of N_t ranging from 1 to 12 (the coordination number). In all cases tested, the maximum efficiency was obtained when all base factors were tested, that is, when $N_t = 12$. However, this conclusion is not general. It is expected that intermediary values of N_t might be optimal when lattices with higher coordination numbers are considered.

To test the performance of the RCB-RL method when the regrowing beads encounter a heterogeneous local environment, we simulated a protein-like heteropolymer exhibiting a first-order folding transition. A chain with 48 segments whose interactions are designed by Abkevich et al.[40] to yield protein-like behavior is considered here (the first sequence in Table 1 of such reference, i.e., TSKRQQPYPMSLGSPFIRIPMIGPRPRMRLLILLMGYPKRGRSGGGLF). Unlike the previous systems, this chain is confined to a sc lattice and is made up of 12 different monomers. At its folding temperature, a simulation method must sample both the folded and unfolded states with the correct relative frequency; therefore, the efficiency of a method can be measured by its ability to decorrelate successive configurations (and overcome a large free energy barrier that separates the two states). It was found that flip moves do not have this ability, thus being unsuitable for efficiency comparisons. Hence, another standard procedure known as the Verdier−Stockmayer[41] (V−S) method was used instead. This method was specifically designed for sc lattices and consists of chain-end, corner, and crankshaft moves.[41] The 48-mer protein was then simulated at its folding temperature, by using both the V−S method and the RCB-RL method with ideal guiding functions $g_i^{id}(\mathbf{r})$. The inverse correlation time (eq 19) of the V−S simulation was 0.1114 s^{-1}. Relative efficiencies ($\gamma_{RCB-RL}/\gamma_{V-S}$) as a function of M are shown in Figure 8 for the two regrowing schemes, eq 3 and eq 10. In contrast to the homopolymer case, eq 3 leads to higher efficiency than eq 10 when $M \geq 4$. This occurs because eq 10 becomes too restrictive and results in very low acceptance ratios. Only RCB-RL moves with eq 3 were able to outperform V−S moves, and a maximum relative efficiency of about 2.5 was attained with $M = 8$. Although such an improvement is modest, one should bear in mind that the V−S moves are specialized for the sc lattice, while the RCB-RL method is very general with respect to lattice structure. Finally, the use of the alternative guiding functions alluded to in Figure 7 was also evaluated; as in the case of homopolymers in poor-solvent conditions, the applied guiding function had negligible effect on efficiency.

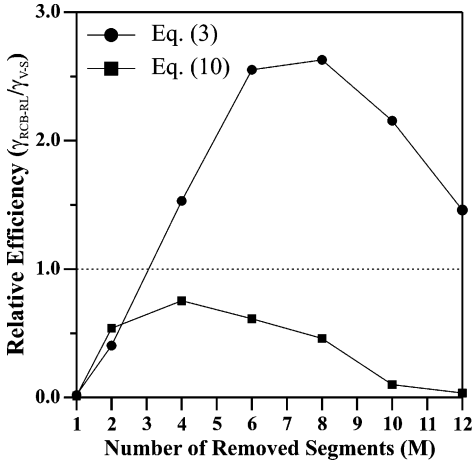

Figure 8. Efficiencies of the RCB-RL method relative to that of the Verdier−Stockmayer method in simulating a 48-mer lattice protein at the folding temperature.

In brief, the tests prove that the RCB-RL method is both valid and effective. Additional simulations of molecules containing chain ends and cross-linking segments (not reported) lead to the same conclusion.

3.2. Effects of Architectural Complexity on the Conformational Statistics of Macromolecules. The RCB-RL method is now used to investigate the conformational behavior of the structures referred to in the introductory section, namely linear, cyclic, theta-shaped, and crosshair-shaped molecules (see Figure 1). Again, polymers in dilute solution are modeled as isolated molecules with segments interacting via contact potential and confined to an fcc lattice. An important conformational property to be considered is the scaling of the radius of gyration with the number of segments (N). The scaling of linear and cyclic chains has been intensively studied, but neither theoretical nor experimental results were found for the scaling of multicyclic chains. For ideal off-lattice chains with fixed bond length, the root-mean-square radius of gyration (rms R_g) of both linear and cyclic architectures scales according to a power law $\langle R_g^2\rangle^{1/2} \propto N^{\nu}$ with exponent $\nu = 1/2$.[16] Yet, the rms R_g of an ideal linear chain is on average $2^{1/2}$ times larger than that of an ideal cyclic chain with the same number of bonds.[16] When excluded volume is considered, the theoretical scaling exponent for linear chains is about 3/5.[16] Monte Carlo simulations in the simple cubic lattice have shown that this exponent does not change for cyclic molecules.[42] For poor-solvent conditions, the globular nature of the conformations results in $\nu \approx 1/3$.

Here, fcc lattice simulations were carried out for linear and cyclic molecules as well as for the multicyclic (fused-ring) molecules shown in Figure 1. Before presenting the results, however, it is important to examine the applicability of CBMC-FE methods to molecules with closed cycles. One of the advantages of Monte Carlo methods is the possibility of performing unphysical moves in order to favor computational efficiency. CBMC-FE methods clearly take advantage of this possibility, given that a regrowing chain section can cross through another one in a single move. For linear and branched molecules, this enhances ergodicity. On the other hand, molecules with closed cycles possess "topologic memory", as properly discussed by Grosberg and Khokhlov.[16] This means that, for nonideal molecules and systems thereof, knots in closed cycles and entanglements among them are perpetual once they are formed during synthesis,

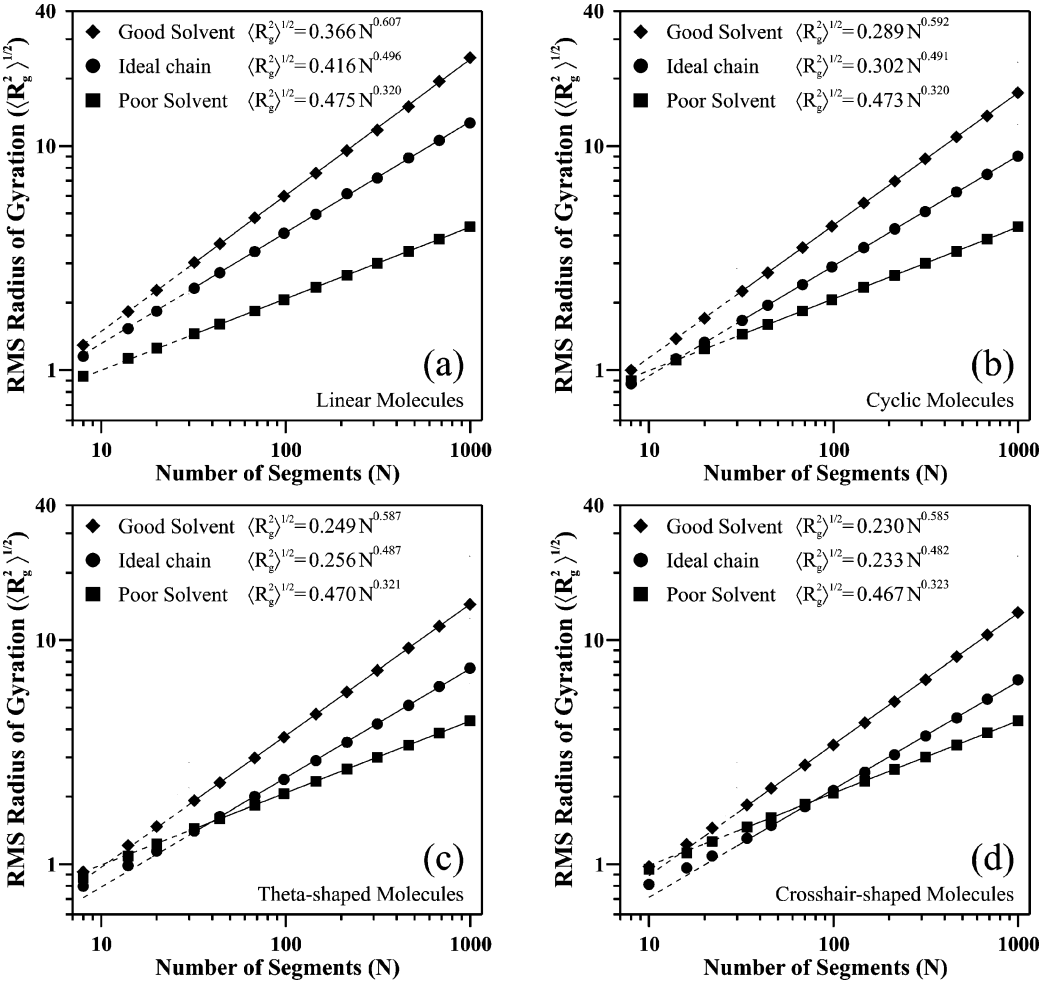

Figure 9. Size scaling of the molecular architectures of Figure 1 confined in an fcc lattice (with bond length equal to $2^{1/2}$) and subject to distinct solvent conditions.

and new knots and entanglements can never be formed after that. Clearly, the presence of such topological constraints might strongly affect the properties of a system, and these facts must be taken into consideration when conformations are sampled via Monte Carlo simulations. None of the published CBMC-FE methods are immune to this problem, and neither is the method proposed here. However, in simulations of isolated molecules with no original knots, eventually formed knots should be short-lived due to the much larger entropy of unknotted conformations. Zifferer and Presser[42] demonstrated this for the case of cyclic molecules in simple cubic lattices. Therefore, "topologic memory" issues will be neglected in this work. Nevertheless, one must be careful when applying CBMC-FE methods to cyclic molecules, especially in dense systems (e.g., polymer melts) if large values of M ($M > 10$) are used.

In all simulations, ideal guiding functions were applied and the number of removed segments (M) in each particular case was chosen based on the previous efficiency tests. The total number of segments varied from $N = 8$ to $N = 1000$. In the case of theta- and crosshair-shaped molecules, the values of N were carefully chosen to yield symmetric structures (same number of internal segments in all sections that bridge cross-linking segments). Again, three cases were considered: ideal chains, self-avoiding chains with $\beta\epsilon = 0$ (good solvent), and self-avoiding chains with $\beta\epsilon = -1$ (poor solvent). The results are presented in Figure 9. For every individual case, power-law parameters were calculated by least-squares regression using only results for $N > 30$. The resulting equations, along with their corresponding curves (straight lines in log−log plots), are also shown in Figure 9. Results for small molecules ($N < 30$) were not used in the regressions because deviations from power-law scaling are observed in this range for the theta- and crosshair-shaped architectures. These deviations are due to the spatial discretization (fcc lattice) as well as to the topological constraints imposed to these multicyclic architectures.

First of all, it is interesting to compare the scaling exponents obtained for the different situations. In the case of good solvent condition, all exponents are about the theoretical value for self-avoiding walks ($\nu = 3/5$). However, a systematic decrease in such exponent is observed as the level of topological constraint of the molecules increases from Figure 8a to 8d (0.607 > 0.592 > 0.587 > 0.585). The same occurs for ideal chains; i.e., all exponents are close to the theoretical value for random walks ($\nu = 1/2$), but a systematic decrease is observed for increasing complexity (0.496 > 0.491 > 0.487 > 0.482). This tendency seems to be inverted in the case of poor-solvent condition, but the small difference among exponents does not permit to draw a reliable conclusion. In fact, all fitted curves for the poor-solvent condition are very close to each other (including the value of the prefactor), indicating that the four kinds of molecules have similar globular structures. The relative sizes of the distinct molecules for the same

molecular weight and for a given solvent quality can be compared through the prefactors if the exponent differences are neglected. Hence, the rms R_g of linear ideal chains are approximately 1.38 times larger than those of their cyclic counterparts. This ratio is close to theoretical predictions ($2^{1/2} \approx 1.41$) for off-lattice ideal chains with the same number of bonds. According to Figure 9, the ideal cyclic molecules are 18% larger than the corresponding theta-shaped chains, while these are about 10% larger than the ideal crosshair-shaped chains. In the good-solvent condition, the rms R_g of molecules with a given N also decreases with the level of topological constraint, the size ratios being 1.27 (linear/cyclic), 1.16 (cyclic/theta), and 1.08 (theta/crosshair).

Ideal chains are often used as adequate models to represent the theta-solvent condition. However, a peculiar behavior can be observed for the crosshair-shaped molecules (Figure 9d). For N smaller than about 80, ideal chains are smaller than their attractive self-avoiding counterparts. As the globular state is expected to have the smallest possible size when excluded volume is considered, then ideal chains cannot represent any intermediary state between poor- and good-solvent conditions for such molecules. To a lesser degree, this also happens for theta-shaped and cyclic molecules. This underlines an important limitation of the ideal-chain model in describing "theta-solvent" conditions for multicyclic architectures.

To investigate the globule−coil transition of a given molecule, the variation of the rms R_g and the mean energy with temperature was examined. Marked changes in these properties for small changes in temperature signal a second-order globule−coil transition. The so-called multiple histogram reweighting (MHR) method[43] was employed to obtain such "thermal profiles". This method consists of first obtaining overlapping energy histograms by means of simulations at different temperatures. These histograms are then combined in a way that results can be interpolated or even extrapolated to other temperatures. Therefore, instead of representing a profile as a group of points, one can represent it as a continuous curve. The analysis can be simplified for the special case of homopolymers with contact interactions. First, the energy of such a molecule can be simply written as $E = \epsilon n_c$, where n_c is the number of contacts among nonbonded segments. Thus, one can deal directly with n_c profiles rather than E profiles. Furthermore, the effect of temperature on the conformational statistics depends on ϵ exclusively through the product $\beta\epsilon$. Therefore, it is possible to perform the analyses independently of ϵ by dealing with $\beta\epsilon$ rather than temperature. In a simulation where a certain value of the product $\beta\epsilon$ is assumed, the probability of sampling conformations with a given n_c is

$$P_{\beta\epsilon}(n_c) = \frac{\omega(n_c)\exp(-n_c\beta\epsilon)}{\sum_{n_c'}[\omega(n_c')\exp(-n_c'\beta\epsilon)]} \qquad (20)$$

where $\omega(n_c)$ is proportional to the total number of conformations with n_c contacts and the sum runs over all accessible numbers of contacts. If an estimate of $\omega(n_c)$ is available, then it is possible to calculate the ensemble average of any conformational property λ which depends directly on n_c, at any value of $\beta\epsilon$, by $\langle\lambda\rangle_{\beta\epsilon} = \sum_{n_c}[\lambda(n_c)P_{\beta\epsilon}(n_c)]$. Such an estimate can be obtained with the MHR method[43] by combining overlapping histograms of number of contacts instead of energy. Because R_g^2 is not a direct function of n_c, a special approach is needed for calculating its thermal profile. A new variable $\bar{R}_g^2(n_c)$ can be defined as the average of R_g^2 for all conformations with a given number of contacts. This variable can be easily estimated by a simple average taken for all conformations with n_c contacts which are sampled during the same simulations used to obtain the histograms. In this way, $\langle R_g^2\rangle^{1/2}$ can be obtained as a function of $\beta\epsilon$ since $\langle R_g^2\rangle_{\beta\epsilon} = \sum_{n_c}[\bar{R}_g^2(n_c)P_{\beta\epsilon}(n_c)]$.

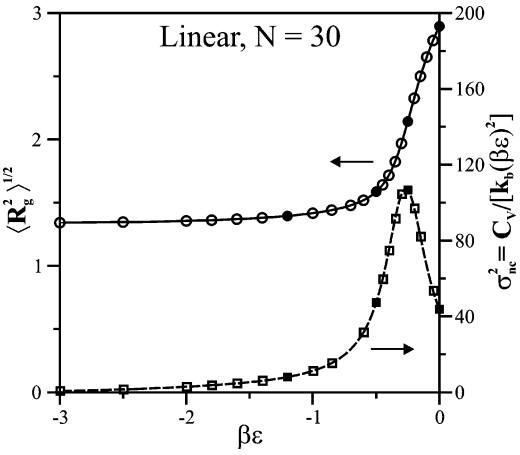

Figure 10. Thermal profiles of the root-mean-square radius of gyration and of the variance in the number of contacts for a linear chain with 30 segments confined to an fcc lattice.

A second-order transition is usually analyzed by tracing the curve of heat capacity (C_v) as a function of temperature. The transition point is then identified by a maximum in this curve. Since C_v is the derivative of the mean energy with respect to temperature, the temperature in which this maximum occurs, T^*, corresponds to an inflection point in the curve of $\langle E\rangle$ vs T, whose slope is given by the maximum C_v value. In other words, the higher the peak in the C_v curve, the "stronger" the transition in question. In the homopolymer case, because energy decreases with n_c ($\beta\epsilon$ is negative for attractive chains), an equivalent analysis is done by tracing a curve of $\sigma_{n_c}^2$ vs $\beta\epsilon$, where $\sigma_{n_c}^2$ is the reciprocal of the derivative of $\langle n_c\rangle_{\beta\epsilon}$ with respect to $\beta\epsilon$, that is

$$\sigma_{n_c}^2 = -\frac{\partial\langle n_c\rangle_{\beta\epsilon}}{\partial(\beta\epsilon)} = \langle n_c^2\rangle_{\beta\epsilon} - \langle n_c\rangle_{\beta\epsilon}^2 \qquad (21)$$

Notice that $\sigma_{n_c}^2$ is the variance in the number of contacts. Since $C_v = (\langle E^2\rangle - \langle E\rangle^2)/(k_bT^2)$, and in the present case $E = \epsilon n_c$, then $\sigma_{n_c}^2 = C_v/[k_b(\beta\epsilon)^2]$.

To evaluate the accuracy of interpolations and extrapolations obtained by the described approach, a linear fcc molecule with 30 segments was considered. Only four simulations with distinct values of $\beta\epsilon$ (0.0, −0.25, −0.5, −1.2) were sufficient to produce adequately overlapping histograms covering the whole accessible range of n_c (the maximum turned out to be 77 for such a molecule). Figure 10 contains the thermal profiles of $\langle R_g^2\rangle^{1/2}$ (solid line) and $\sigma_{n_c}^2$ (dashed line) calculated using the MHR method. The symbols in Figure 10 represent results of individual simulations (circles for $\langle R_g^2\rangle^{1/2}$ and squares for $\sigma_{n_c}^2$). Observe that only those histograms obtained in the simulations corresponding to the full

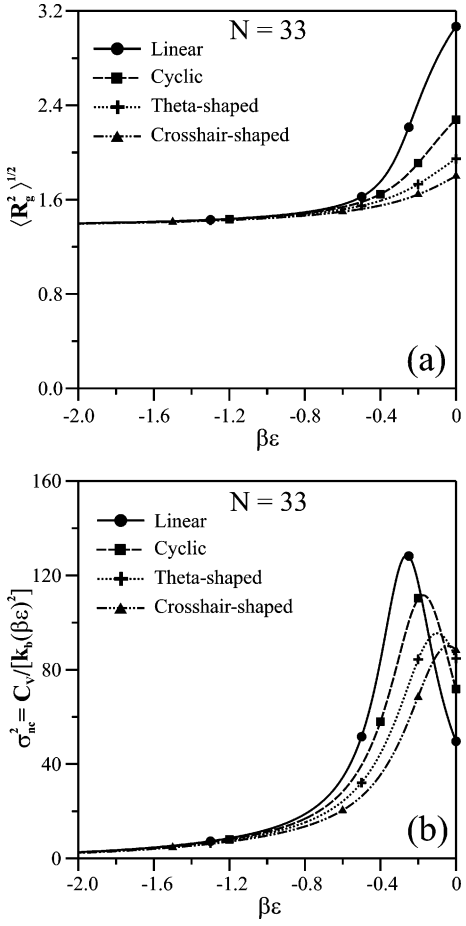

Figure 11. Thermal profiles for distinct molecules with 33 segments confined to an fcc lattice: (a) root-mean-square radius of gyration; (b) variance in the number of contacts.

symbols were used to obtain the curves. The excellent agreement between the open symbols and the curves attests to the reliability of the interpolations and extrapolations of $\omega(n_c)$ over the entire range of n_c.

A second-order transition can be clearly identified in Figure 10 from the sharp increase in the rms R_g profile and a peak in the ${\sigma_{n_c}}^2$ profile. The same profiles can be determined for the other molecular architectures. For a proper comparison, molecules with the same number of segments must be considered. Nevertheless, one cannot build a symmetric theta-shaped and a symmetric crosshair-shaped molecule both with the same N. A possible compromise is to choose N so as to have "quasi-symmetric" molecules (i.e., with ± 1 segment relative to symmetric ones). For example, since one could have a symmetric theta-shaped molecule with 32 segments or a symmetric crosshair-shaped molecule with 34 segments, $N = 33$ is chosen. From Figure 9c, one can see that $N = 33$ is close to the crossing point between the mean sizes of theta-shaped molecules in ideal and poor-solvent conditions. From Figure 9c, $N = 33$ is well below such a crossing point for crosshair-shaped molecules. Figure 11 contains the profiles obtained for the four architectures with $N = 33$. Four simulations were sufficient in each case to produce overlapping histograms in the whole n_c range. As it can be seen, Figure 11a confirms that the low-temperature globular state for all structures has the same rms R_g. In the athermal condition ($\beta\epsilon = 0$), the mean size diminishes as the number of topological constraints increases (linear → cyclic → theta-shaped → crosshair-shaped), which was already concluded from Figure 9. Besides, Figure 11a shows that the size differences continuously increase with temperature. As observed in Figure 11b, the globule−coil transition is markedly distinct in each case. First, the height of the peak in the ${\sigma_{n_c}}^2$ profile gets smaller as topological complexity increases, meaning that the transition becomes less pronounced (smaller slope at the inflection point of the $\langle n_c \rangle_{\beta\epsilon}$ curve). In fact, in the most constrained case the profile only slightly deviates from what would be a monotonic curve (with no second-order transition at all). Also note that the respective peaks shift to the right; i.e., the transition temperature $T^* \sim -(\beta^*\epsilon)^{-1}$ increases upon increasing chain complexity. This is due to the entropy reduction caused by the presence of topological constraints. T^* can be regarded as the one in which energetic effects (that favor compact conformations) and entropic effects (that favor expanded conformations) come to a balance. Below it, energy prevails over entropy, and vice versa. Therefore, if an overall reduction in entropy occurs, it is natural to expect an enlargement of the range in which energy prevails, i.e., an increase in T^*. One can finally observe that, for the studied structures, the fluctuations in the number of contacts at the athermal condition (${\sigma_{n_c}}^2$ at $\beta\epsilon = 0$) are larger for more constrained topologies.

To address the question of how the globule−coil transition occurs for larger molecules, simulations were carried out for $N = 99$ (which satisfies the "quasi-symmetry" criterion). This number is slightly above the one which gives the crossing point between ideal and poor-solvent chains with crosshair-shaped architecture (see Figure 9d). In this case, every structure required six simulations to produce the desired overlapping histograms covering the whole range of n_c. The resulting thermal profiles are shown in Figure 12. As observed in Figure 12a, the rms R_g profiles are similar to the ones for $N = 33$ but with greater differences among distinct architectures at high temperatures. For the most constrained molecules, the globule−coil transition is now more noticeable. This is confirmed by means of the ${\sigma_{n_c}}^2$ profiles in Figure 12b, all of which exhibit marked peaks. On the other hand, the shift of the peaks to the right is less pronounced in this case, meaning that T^* of different molecules gets closer to each other as N increases. This is consistent with the given explanation for the shift, since for larger multicyclic chains one would expect a smaller percentage difference in entropy relative to linear chains.

Comparison of the conformational entropies of the distinct molecules would give direct support to the argument provided above to explain the trends in R_g and ${\sigma_{n_c}}^2$. This would require knowledge of absolute entropies ($S_{\beta\epsilon}$), which is not accessible by Monte Carlo simulations. On the other hand, entropy differences of a given molecule at different conditions $\beta\epsilon$ can be obtained by a MHR analysis. Thus, assuming that differences in entropy of distinct molecules are much smaller at low temperatures than at high temperatures, a sensible comparison can be done at high temperatures by means of $S_{\beta\epsilon} - S_{\beta\epsilon\to-\infty}$, rather than the absolute entropies themselves. Let $\Delta S_{\beta\epsilon}$ be the entropy variation of a given molecule from the athermal condition ($\beta\epsilon = 0$) to an arbitrary condition $\beta\epsilon$. Then $\Delta S_{\beta\epsilon}/k_b = \langle \beta E \rangle_{\beta\epsilon} - \Delta(\beta F)_{\beta\epsilon}$, where F is the Helmholtz free energy and $\Delta(\beta F)_{\beta\epsilon}$ is the variation of βF from the athermal condition to $\beta\epsilon$. By definition, $\beta F = -\ln\{\sum_{n_c}[\Omega(n_c)\exp(-n_c\beta\epsilon)]\}$, where $\Omega(n_c)$ is the number of conformations with n_c

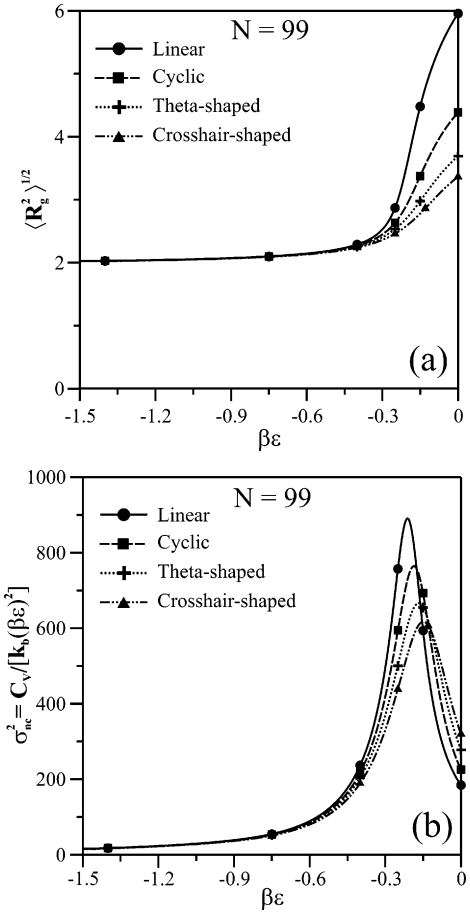

Figure 12. Thermal profiles for distinct molecules with 99 segments confined to an fcc lattice: (a) root-mean-square radius of gyration; (b) variance in the number of contacts.

contacts, proportional to $\omega(n_c)$. Therefore, one can conclude that

$$\frac{\Delta S_{\beta\epsilon}}{k_b} = \beta\epsilon\langle n_c\rangle_{\beta\epsilon} + \ln\left(\frac{\sum_{n_c}[\omega(n_c)\exp(-n_c\beta\epsilon)]}{\sum_{n_c}[\omega(n_c)]}\right) \qquad (22)$$

If multiple-histogram estimates of $\omega(n_c)$ are available, one can use eq 22 to obtain $\Delta S_{\beta\epsilon}$ for a given molecule from $\beta\epsilon = 0$ down to negative values of $\beta\epsilon$. As $\beta\epsilon$ decreases, $\Delta S_{\beta\epsilon}$ asymptotically approaches a minimum value, which is $\Delta S_{\beta\epsilon\to-\infty}$. Once this occurs, $S_{\beta\epsilon} - S_{\beta\epsilon\to-\infty}$ can be calculated for any condition $\beta\epsilon$, since it is equal to $\Delta S_{\beta\epsilon} - \Delta S_{\beta\epsilon\to-\infty}$. Entropy profiles for the simulated molecules are shown in Figure 13. These results are consistent with the explanations given for the observed profiles in Figures 10 and 11. Both for $N = 33$ or for $N = 99$, entropy systematically decreases as the topological complexity of the molecules increase. In addition, relative differences in entropy are smaller for $N = 99$ than for $N = 33$. For $N = 33$, the entropy of a crosshair-shaped molecule is about 29% smaller at the athermal condition than that of a linear molecule; for $N = 99$, the corresponding reduction is only 13%.

Recalling that $\beta\epsilon = -1$ and $\beta\epsilon = 0$ correspond to the poor- and good-solvent conditions represented in Figure 9, respectively, the corresponding values of $\langle R_g^2\rangle^{1/2}$ taken from the curves of Figures 10a and 11a can be compared to those obtained by taking $N = 33$ or N 99 in the proper power laws of Figure 9. Excellent agreement is obtained in such comparisons, with all relative deviations below

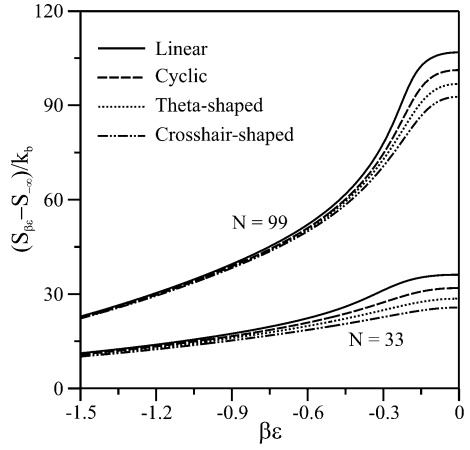

Figure 13. Entropy profiles of distinct homopolymer molecules confined to an fcc lattice.

1% and most below 0.5%. If ideal chains are assumed to be adequate models for the transition state between poor- and good-solvent conditions, then one might expect that the R_g of a self-avoiding molecule at its T^* is close to that of an equivalent ideal chain. To test this idea, the values of $\beta\epsilon$ which give the maximum $\sigma_{n_c}{}^2$ in Figures 10b and 11b are first identified for every molecule, and their rms R_g values are obtained from Figures 10a and 11a. These values are then compared to those obtained by the corresponding ideal power laws from Figure 9. Poor agreement is seen in all cases. If $N = 33$, the ideal-chain rms R_g deviate from those at T^* by 8.5%, −14.3%, −23.4%, and −28.8% for the linear, cyclic, theta-shaped, and crosshair-shaped molecules, respectively. If $N = 99$, such deviations are 20.0%, −6.1%, −16.7%, and −23.4%. Notice that the deviations are positive (ideal chain > T^* chain) for linear molecules and negative for all the others. At least for the systems studied here, the ideal-chain model does not quite represent the globule−coil transition states. Of course, this conclusion is subject to the definition of the transition state adopted here. To see how the rms R_g of molecules at T^* scales with N, the values for $N = 33$ and $N = 99$ were used to estimate the scaling exponents of 0.404, 0.408, 0.410, and 0.416 for the linear, cyclic, theta-shaped, and crosshair-shaped molecules, respectively. These exponents are close to 2/5, thus differing noticeably from the theoretical value of 1/2 for ideal chains. While a more reliable estimation would involve data for several N values over a larger range, the calculated exponents give a good indication of the range in which the true values might be.

Finally, it should be remarked that homopolymers in dilute solution can undergo a sort of solid−liquid (order−disorder) transition while in the globular state. This has been observed both experimentally[44] and via molecular simulations of off-lattice linear chains.[45,46] In this transition, stable spatially ordered (solidlike) globules give place to disordered (liquidlike) globules as temperature increases. However, if the model chain is confined to a crystalline lattice, it is not possible to have a spatially disordered compact globule. Hence, for the systems studied in this work, no evidence of such a "melting" transition was detected.

4. Conclusions

In this work, a new configurational-bias Monte Carlo method with fixed end points was developed for the

simulation of lattice polymers. Instead of assuming a particular structure, the method deals with arbitrary regular lattices whose sites are all connected to their neighbors by the same vectors. A "move" with such a method consists of removing an internal chain section and then sequentially relocating it until the original connectivity is restored. A key aspect is the use of weights that guide the process to chain closure. Ways of obtaining proper weights in the case of regular lattices were proposed. The advocated method was evaluated in several simulation conditions and shown to be more effective than "flip moves". In some cases, efficiency was further enhanced by incorporating a new, simple scheme for calculating probabilities, which corresponds to replacing eq 3 by eq 10 (see section 2.1).

The method was then used to investigate the effect of architectural complexity on the conformational statistics of polymers in dilute solution. For this, a family of molecular structures involving fused-ring multicyclic polymers (see Figure 1) was considered on the fcc lattice. The size scaling of the molecules in ideal, good-solvent, and poor-solvent conditions was studied for chains with up to 1000 segments. In all cases, the scaling exponents were observed to be close to the theoretical values for linear and cyclic chains, although a small but systematic decrease was noticed with increasing topological constraints, in both ideal and good-solvent conditions. R_g is also reduced by the presence of such constraints. Furthermore, it was observed that ideal multicyclic chains cannot represent well the transition state between poor- and good-solvent conditions (globule−coil transition), especially for low molecular weights.

The globule−coil transition of the structures in Figure 1 was investigated for $N = 33$ and $N = 99$. Thermal profiles of R_g and heat capacity were obtained with the MHR method. For $N = 33$, it was observed that the "strength" of the globule−coil transition is reduced and the transition temperature is increased as the architectural complexity increases. This is explained by the overall conformational entropy reduction in the molecules, relative to linear chains, associated with the topological constraints. For $N = 99$, similar thermal profiles were seen, but with smaller differences between molecule types. This is consistent with the fact that the percentage entropy reduction of a given architecture relative to linear chains diminishes as N increases.

Finally, it is important to mention that the method developed here could be used to simulate several other types of polymer systems, such as polymer melts, polymer networks, copolymers, and proteins. The results presented here constitute an initial effort in understanding the behavior of polymers with complex architecture.

Acknowledgment. The authors gratefully acknowledge the financial support of the National Science Foundation, Award ECS-030483.

Appendix

Some details concerning the implementation of the RCB-RL method are addressed in this Appendix. In principle, the method is applicable to any regular lattice. In practice, however, some difficulties arise when the coordination number is high. First of all, eq 3 or eq 10 must be evaluated several times during each regrowing process, which can become excessively expensive for a large Z. Moreover, instead of trying all base vectors when selecting a new position for a segment, it is possible to try only a few of them. In general, let N_t ($\leq Z$) be the number of trial base vector, whose indexes constitute a set of distinct integers $\{t_k\}_{k=1,...,N_t}$ between 1 and Z. To obey detailed balance, those indexes must be randomly chosen at each step, which can be done with one of the algorithms described by Vitter.[47] Then, the new position of a segment "i" is determined by selecting one of the elements of $\{t_k\}_{k=1,...,N_t}$ with probability given by making $j = t_k$ in eq 3 or eq 10. Naturally, the optimum number of trial vectors N_t is system-dependent. Another concern created by a large Z is that the number of combinations C_i^{Z+i-1} can become extremely large and compromise the feasibility of the algorithm in Figure 3. This issue also limits the maximum number of bonds that can be managed, since C_i^{Z+i-1} also increases with i.

Along with the ability of handling different lattice structures, the freedom in the choice of the interaction potential furthers the generality of the RCB-RL method. For computational efficiency, however, it is convenient to use contact potentials. In this case, two segments "i" and "j" interact only if they lie in adjacent sites, and the energy of interaction between them depends only on their types, τ_i and τ_j. Hence, if N_τ distinct segment types are considered, the interaction potential corresponds to a set of $N_\tau(N_\tau + 1)/2$ parameters. When searching for interacting partners of a given segment, one no longer needs to account for all other segments in the system, neither to maintain updated lists of neighbors, but only to verify the occupancy of adjacent sites. Adopting this approach in the RCB-RL method, the energy change associated with the placement of a segment "i" (of type τ_i) in a position $\mathbf{r}_i^n$ of the lattice is given by

$$u(\tau_i,\mathbf{r}_i^n) = \sum_{j=1}^{Z} \epsilon_{\tau_i,\tau_j} \tag{23}$$

where ϵ_{τ_i,τ_j} is the contact energy between types τ_i and τ_j, the latter being the type of the segment occupying the position $\mathbf{r}_i^n + \mathbf{b}_j$. If such position is empty, then $\epsilon_{\tau_i,\tau_j} = 0$. Moreover, knowing that terms like $u(\tau_i,\mathbf{r}_i^n)$ always appear as arguments of exponentials in the acceptance criterion, one can rewrite the equation above as

$$\exp[-\beta u(\tau_i,\mathbf{r}_i^n)] = \prod_{j=1}^{Z} \exp(-\beta\epsilon_{\tau_i,\tau_j}) \tag{24}$$

Therefore, for a fixed temperature, factors $\exp(-\beta\epsilon_{\tau_i,\tau_j})$ corresponding to all $N_\tau(N_\tau + 1)/2$ energy parameters can be evaluated in advance and stored and then eq 24 used to avoid frequent exponential evaluations during a simulation.

For the contact potential to be employed, some data structure is needed to store the occupancy status of the considered lattice domain. In general, a graph with the same topology as the lattice is a natural choice. However, a helpful simplification is possible if the range of possible lattices is restricted to those which can be defined by base vectors whose components are all integers. In this way, if the origin of the coordinate space is properly set, all site positions will be given by integer vectors and a simple array will be sufficient to store the lattice occupancy. All lattice structures mentioned in the introductory section can be defined by integer base

vectors. For example, all six permutations and sign combinations of (±1, ±1,0) define an fcc lattice. The same relation holds between (±1,0,0) and the sc lattice and between (±1, ±1, ±1) and the bcc lattice, and similar relations can be found for other structures.[28,30] For the sake of generality, the occupancy of a site located in (x,y,z) can be stored in the element (i,j,k) of a 3D array by simply making $i = x$, $j = y$, and $k = z$. However, depending on the lattice, such a mapping can give rise to a wasteful usage of computer memory. For example, half of the array elements would be redundant in the case of an fcc lattice. For this particular case, the problem is solved if every element (i,j,k) is related to a position (x,y,z) by making $i = x$, $j = y$, and $k = z/2$ if z is even or $k = (z - 1)/2$ if z is odd.

Regarding the guiding weights, suppose that a preliminary simulation is executed and histograms $\bar{g}_i^0(r^2)$ for $i = 1, ..., M$ are obtained and used to construct a guiding weight function $g_i^0(\mathbf{r})$. If, for the considered system, the actual probability of sampling a given combination of i (number of bonds) and r (end-to-end distance) is very small, then such a combination may not be sampled during the preliminary simulation and consequently $\bar{g}_i^0(r^2)$ will be zero for such i and r. Nevertheless, the same combination can be sampled in a further simulation using the guiding weights $g_i^0(\mathbf{r})$. For example, suppose that segments "3" and "4" are removed from a chain with five segments. After that, segment "3" is relocated, and its new distances to the segments "1" and "5" are r_{1-3} and r_{3-5}, respectively. This can occur as long as $\bar{g}_2^0(r_{3-5}^2) \neq 0$, regardless of the value of r_{1-3}. Then, once the whole move is accepted, the combination $i = 2$ and $r = r_{1-3}$ will have been sampled, even if $\bar{g}_2^0(r_{1-3}{}^2) = 0$. Now suppose that, in the next step, segments "2" and "3" are removed, and this is the first to be relocated. In such case, $\bar{g}_2^0(r_{1-3}{}^2)$ being zero implies that W^o can also be zero and consequently render $P_{acc}(o \rightarrow n)$ indefinite. Of course, such a problem would never occur if ideal guiding weights $g_i^{id}(\mathbf{r})$ were employed, since all accessible combinations of i and r have $g_i^{id}(\mathbf{r}) \neq 0$. Therefore, a possible way of avoiding this problem is to construct $g_i^0(\mathbf{r})$ by taking $\bar{g}_i^0(r)$ for all combinations of i and r which are sampled and $\bar{g}_i^{id}(r)$ for those which are accessible but not sampled during the preliminary simulation ($r = 0$ being an exception if volume exclusion is considered). Since this may occur only for combinations of i and r with very small probabilities, the effect of this approach on the efficiency of the method will be negligible.

References and Notes

(1) Seeman, N. C. *Annu. Rev. Biophys. Biomol. Struct.* **1998**, *27*, 225−248.
(2) Tomalia, D. A.; Frechet, J. M. J. *J. Polym. Sci., Part A: Polym. Chem.* **2002**, *40*, 2719−2728.
(3) Malmstrom, E. E.; Hawker, C. J. *Macromol. Chem. Phys.* **1998**, *199*, 923−935.
(4) Grayson, S. K.; Frechet, J. M. J. *Chem. Rev.* **2001**, *101*, 3819−3867.
(5) Tezuka, Y.; Oike, H. *Macromol. Rapid Commun.* **2001**, *22*, 1017−1029.
(6) Tezuka, Y.; Oike, H. *Prog. Polym. Sci.* **2002**, *27*, 1069−1122.
(7) Seeman, N. C. *Nature (London)* **2003**, *421*, 427−431.
(8) Chworos, A.; Severcan, I.; Koyfman, A. Y.; Weinkam, P.; Oroudjev, E.; Hansma, H. G.; Jaeger, L. *Science* **2004**, *306*, 2068−2072.
(9) McLeish, T. *Science* **2002**, *297*, 2005−2006.
(10) Bielawski, C. W.; Benitez, D.; Grubbs, R. H. *Science* **2002**, *297*, 2041−2044.
(11) Antonietti, M.; Folsch, K. J. *Makromol. Chem., Rapid Commun.* **1988**, *9*, 423−430.
(12) Oike, H.; Imaizumi, H.; Mouri, T.; Yoshioka, Y.; Uchibori, A.; Tezuka, Y. *J. Am. Chem. Soc.* **2000**, *122*, 9592−9599.
(13) Tezuka, Y.; Tsuchitani, A.; Yoshioka, Y.; Oike, H. *Macromolecules* **2003**, *36*, 65−70.
(14) Tezuka, Y.; Oike, H. *J. Am. Chem. Soc.* **2001**, *123*, 11570−11576.
(15) Flory, P. J. *Statistical Mechanics of Chain Molecules*; Hanser Publishers: Munich, 1989.
(16) Grosberg, A. Y.; Khokhlov, A. R. *Statistical Physics of Macromolecules*; AIP Press: New York, 1994.
(17) Dijkstra, M.; Frenkel, D.; Hansen, J. P. *J. Chem. Phys.* **1994**, *101*, 3179−3189.
(18) Tanaka, G.; Mattice, W. L. *Macromol. Theory Simul.* **1996**, *5*, 499−523.
(19) Sikorski, A. *Macromol. Theory Simul.* **2001**, *10*, 38−45.
(20) Rapold, R. F.; Mattice, W. L. *J. Chem. Soc., Faraday Trans.* **1995**, *91*, 2435−2441.
(21) Rapold, R. F.; Mattice, W. L. *Macromolecules* **1996**, *29*, 2457−2466.
(22) Doruker, P.; Mattice, W. L. *Macromolecules* **1998**, *31*, 1418−1426.
(23) Doruker, P.; Rapold, R. F.; Mattice, W. L. *J. Chem. Phys.* **1996**, *104*, 8742−8749.
(24) Chen, D. H.; Mattice, W. L. *Polymer* **2004**, *45*, 3877−3883.
(25) Bagci, Z.; Jernigan, R. L.; Bahar, I. *Polymer* **2002**, *43*, 451−459.
(26) Bagci, Z.; Kloczkowski, A.; Jernigan, R. L.; Bahar, I. *Proteins: Struct., Funct. Genet.* **2003**, *53*, 56−67.
(27) Carmesin, I.; Kremer, K. *Macromolecules* **1988**, *21*, 2819−2823.
(28) Deutsch, H. P.; Binder, K. *J. Chem. Phys.* **1991**, *94*, 2294−2304.
(29) Panagiotopoulos, A. Z.; Wong, V.; Floriano, M. A. *Macromolecules* **1998**, *31*, 912−918.
(30) Kolinski, A.; Skolnick, J. *Polymer* **2004**, *45*, 511−524.
(31) Frenkel, D.; Smit, B. *Understanding Molecular Simulation: From Algorithms to Applications*; Academic Press: San Diego, CA, 2002.
(32) Rane, S. S.; Mattice, W. L. *J. Chem. Phys.* **2005**, *122*.
(33) Escobedo, F. A.; dePablo, J. J. *J. Chem. Phys.* **1995**, *102*, 2636−2652.
(34) Vendruscolo, M. *J. Chem. Phys.* **1997**, *106*, 2970−2976.
(35) Wick, C. D.; Siepmann, J. I. *Macromolecules* **2000**, *33*, 7207−7218.
(36) Chen, Z.; Escobedo, F. A. *J. Chem. Phys.* **2000**, *113*, 11382−11392.
(37) Rosenbluth, M. N.; Rosenbluth, A. W. *J. Chem. Phys.* **1955**, *23*, 356−359.
(38) dePablo, J. J.; Laso, M.; Suter, U. W. *J. Chem. Phys.* **1992**, *96*, 6157−6162.
(39) Flyvbjerg, H.; Petersen, H. G. *J. Chem. Phys.* **1989**, *91*, 461−466.
(40) Abkevich, V. I.; Gutin, A. M.; Shakhnovich, E. I. *Folding Des.* **1996**, *1*, 221−230.
(41) Verdier, P. H.; Stockmayer, W. H. *J. Chem. Phys.* **1962**, *36*, 227-&.
(42) Zifferer, G.; Preusser, W. *Macromol. Theory Simul.* **2001**, *10*, 397−407.
(43) Ferrenberg, A. M.; Swendsen, R. H. *Phys. Rev. Lett.* **1989**, *63*, 1195−1198.
(44) Wu, C.; Wang, X. H. *Phys. Rev. Lett.* **1998**, *80*, 4092−4094.
(45) Zhou, Y. Q.; Hall, C. K.; Karplus, M. *Phys. Rev. Lett.* **1996**, *77*, 2822−2825.
(46) Zhou, Y. Q.; Karplus, M.; Wichert, J. M.; Hall, C. K. *J. Chem. Phys.* **1997**, *107*, 10691−10708.
(47) Vitter, J. S. *Commun. ACM* **1984**, *27*, 703−718.

MA050725T