Monte Carlo Conformational Sampling of the Internal Degrees of Freedom of Chain Molecules

Alfred Uhlherr*

CSIRO Molecular Science, Bag 10, Clayton South, Victoria 3169, Australia

Received June 1, 1999; Revised Manuscript Received September 8, 1999

ABSTRACT: The challenge of controlled sampling of the conformations of internal sections of chain molecules, subject to constrained interatomic bond lengths and angles, is central to many areas of macromolecular science. A new method for overcoming this challenge via an internal configuration bias (ICB) Monte Carlo algorithm is described. It is demonstrated that the algorithm obeys the detail balance (microscopic reversibility) criterion necessary for performing rigorous molecular simulations in equilibrium ensembles. The algorithm is applied to a study of the molecular conformations of cyclic alkane molecules in a vacuum, where it is shown to be up to ∼2 orders of magnitude more efficient than standard molecular dynamics simulation techniques. Qualitative transitions between constrained ring and flexible chain behavior are observed between 16 and 30 backbone atoms for local structure (torsion angle distribution) and between 30 and 50 backbone atoms for global ring dimensions.

Introduction

Progress in macromolecular science and technology, as with most chemical disciplines, is dependent on the knowledge of the atomic structure of a compound and the role of this structure in determining the properties of the compound. For molecules composed of flexible chains (polymers, proteins, surfactants, etc.), this interrelationship is complicated by the large range of conformational states accessible to such chains.[1] For this reason, molecular modeling and simulation techniques[2−5] are assuming increasing importance in macromolecular research and product development.

In many biological applications, it is often sufficient to perform a conformational search for the molecular system[6] in order to isolate one or more conformations of interest (e.g., minimum energy). More frequently in the physical sciences (condensed matter, materials, chemical physics, polymer science, petrochemical technology, etc.), it is necessary to characterize a compound and its properties by a representative subset of all the accessible conformational states, such that any given state occurs with the correct relative probability. Typically, this set of probabilities corresponds to a desired statistical mechanical ensemble.

Conformational analysis is widely utilized for characterizing individual chain molecules, typically under neutral (Θ) solvent conditions, or for bulk systems, such as polymer materials. The former may be studied analytically by the classical rotational isomer state (RIS) theory.[1,7,8] Alternatively, it is straightforward to apply a Monte Carlo (MC) simulation approach,[9−11] where the torsion angles in the chain are sampled using a Markov sequence of simple operations, such as a "pivot" move of the chain around a randomly selected bond.[12]

Bulk materials are much more awkward, as the additional constraints imposed by dense packing of chains greatly increases the time required to sample the accessible configuration space of the molecular system.[13] Molecular dynamics (MD) simulations are widely adopted for such studies, although they are computationally demanding. It would be highly desirable to study systems in full chemical detail, at realistic (large) chain lengths or molecular weights, over sufficiently long times to observe full global chain relaxation. In practice, currently and in the foreseeable future, only two of these three goals can be satisfied simultaneously.[14] Thus, all macromolecular MD studies to date have fallen into one of the following three categories: (i) atomistic systems of short chains (e.g., alkanes) over long times, (ii) atomistic systems of long chains over short times (e.g., nonequilibrium glassy polymers), and (iii) "coarse grained" systems of long chains over long times.

These limitations have been the major driving force for the development of new MC algorithms suitable for bulk systems of chain molecules.[13]

In this work, we consider the question of a Monte Carlo algorithm that displaces a specified variable number of atoms within a chain molecule while the other atom positions are held fixed. Several such algorithms have recently been devised for simple lattice and freely jointed chain models.[15−18] However, for more realistic chemically detailed chains, with constrained bond lengths and angles, such a requirement poses a surprisingly complex geometric problem. This was first addressed in the classic paper of Go and Scheraga[19] and led to the development of the statistically rigorous concerted rotation[20] (ConRot) and extended concerted rotation[21] (EcRot) MC algorithms. However, these algorithms operate on chain sections of fixed length (four atoms for ConRot, six or seven atoms for different versions of EcRot). This limits their flexibility and efficiency in vigorously sampling the torsional degrees of freedom in most chain molecules, given that these degrees of freedom are often subtly and extensively coupled.

To test the validity and performance of the new algorithm, the familiar "polybead" model was chosen. Each polybead molecule consists of a chain of spherical Lennard-Jones sites or "beads" with fixed bond lengths and angles.[22] Defining the beads to correspond to carbon-centered "united atoms", such as CH_2, CH_3, or CH groups, yields a reasonably good quantitative description of simple alkanes and polyolefin polymers. Note that the polybead model is strictly speaking not

* Telephone: + 61 3 9545 8107. Fax: + 61 3 9545 2446. E-mail: a.uhlherr@molsci.csiro.au.

10.1021/ma9908595 CCC: $19.00 Published 2000 by the American Chemical Society
Published on Web 01/31/2000

"fully atomistic", because the hydrogen atoms are not represented explicitly. However, the model provides a simple and general, yet still practical, representation of truly atomistic macromolecules, which are typically represented by a detailed "force field" of interatomic potentials containing stiff bond stretching and bond bending terms.[2]

A generalized internal sampling algorithm, of the type presented here, should prove valuable in the following applications: (i) conformational analysis of cyclic molecules, as explored briefly in this paper, (ii) conformational analysis of cyclic peptides and protein "loops",[23,24] (iii) conformations of long or heterogeneous chains adsorbed onto surfaces (e.g., alkyl dithiols, triblock and multiblock copolymers), and (iv) more efficient simulation of bulk atomistic polymer melts, including arbitrarily high molecular weights and cross-linked chains; this could be done, for example, by cooperatively displacing sections of two or more adjacent chains[25] or by using the algorithm in enhanced variable-connectivity schemes[26] for simulating bulk polymers in semigrand ensembles.[27]

Algorithm Derivation

The proposed algorithm is based on a combination of the (continuum) configuration bias or CBMC method[28,29] (here abbreviated to CB) with the internal geometric rebridging method;[27,20] hence, the choice of the name "internal configuration bias" (ICB). A selected set of atoms are removed from the chain and replaced one at a time. Initially, these new atom positions are generated by CB, until the geometric constraints restrict the remaining atoms to a finite set of new positions. Conceptually, the algorithm is virtually the reverse of that of Deem and Bader;[24] the latter generates cyclic peptide conformers by using a ConRot procedure to displace a set of four rigid units using a range of driver angles and then selects one of these new configurations by CB weighting.

Here, we describe the implementation of ICB sampling for a single linear chain molecule with fixed bond lengths and angles (i.e., polybead). This means that the internal energy $u^{\text{int}}(\mathbf{x}_i)$ of atom i at position $\mathbf{x}_i$ simply corresponds to the torsion energy of atom i. Extension to chains with variable bond angles or bond lengths, heteroatoms, and side groups are all straightforward, as with conventional CB.[4] Extension to sections of chains containing branch points (e.g., networks) is also feasible but is not considered here.

The basis of the ICB algorithm is depicted in Figure 1. A selected sequence of n_Ω atoms, Ω, is removed from the chain. These are then "regrown" one atom at a time, from one end of the sequence to the other. The order of regrowth may be selected at random to be either ascending or descending. The first n_χ atoms, χ, can each be regrown in one of an infinite number of possible new positions. The final n_ψ atoms, ψ, are constrained to a finite set of new positions. Thus, $n_\Omega = n_\chi + n_\psi$, where n_ψ is equal to 3 for linear chains with constrained bond lengths and angles.

Each atom i has an associated energy $u(\mathbf{x}_i)$ that is a function of its coordinate vector $\mathbf{x}_i$. This energy is summed over all interactions of atom i with those atoms that are not being moved or have already been regrown. It is convenient to split u $(\mathbf{x}_i)$ into two components,

$$u(\mathbf{x}_i) = u^{\text{int}}(\mathbf{x}_i) + u^{\text{ext}}(\mathbf{x}_i) \tag{1}$$

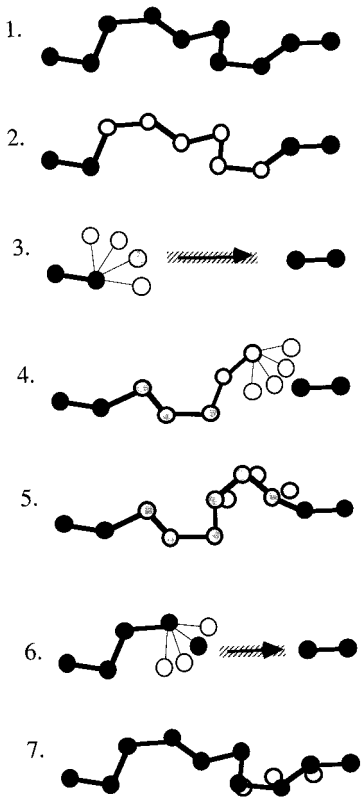

Figure 1. Pictorial two-dimensional representation of the internal configuration bias or ICB move (constraints not drawn to scale). An internal section of the chain molecule (1) is removed (2). Each removed atom i is regrown in turn by selecting one of a set new trial configurations j (3), using an energetic bias that directs the regrowth toward atom k. This is repeated up to and including the final three atoms (4). The coordinates of the final bridging trimer are then refined until they satisfy all bond length and angle constraints (5). The atoms are again removed and the original configuration is rebuilt, using the same sequence of configuration bias (6) and refinement (7) operations.

where $u^{\text{ext}}(\mathbf{x}_i)$ and $u^{\text{int}}(\mathbf{x}_i)$ correspond to nonbond and bonded (in this case torsional) interactions.

For each atom i in χ, we generate a set of n_i trial positions $\{\mathbf{x}_j\}$, $j = 1$ to n_i, with weighted probabilities given by

$$P(\mathbf{x}_j) = C \exp(-\beta v^{\text{int}}(\mathbf{x}_j)) \tag{2}$$

where $v^{\text{int}}(\mathbf{x}_j)$ is the applied internal energy bias, which may be conveniently chosen so that

$$v^{\text{int}}(\mathbf{x}_j) = u^{\text{int}}(\mathbf{x}_j) \tag{3}$$

The normalization constant C is chosen so that the sum of probabilities equals unity, and the reduced temperature β is simply related to the imposed temperature T via Boltzmann's constant k_B by the expression

$$\beta = \frac{1}{k_B T} \tag{4}$$

The new configuration $\mathbf{x}_i$ for atom i is chosen at random from the set of trial positions $\{\mathbf{x}_j\}$, with a weighted probability given by

$$P(\mathbf{x}_i) = \frac{\exp(-\beta v^{\text{ext}}(\mathbf{x}_i))}{\sum_{j=1}^{n_i} \exp(-\beta v^{\text{ext}}(\mathbf{x}_j))} \tag{5}$$

In conventional CB sampling, the biasing energy is given simply by the atom's external energy. However, for internal sampling to be effective, one needs to "encourage" the atom to regrow toward the other end of the chain. Otherwise, the newly grown subchain will tend to form a random walk that does not close the chain, a problem that becomes especially acute in a vacuum or low-density media. We handle this by noting that the bias energy is in fact completely arbitrary; as long as the bias is correctly accounted for in the final acceptance test (see below), then the algorithm's sampling will be rigorously correct (although not necessarily efficient). Note that a similar principle has been used in the context of transition pathway sampling by Chandler and co-workers.[30] Here, we choose the bias energy to take the form

$$v^{\text{ext}}(\mathbf{x}_j) = u^{\text{ext}}(\mathbf{x}_j) - \frac{a_{ik}}{b_{ik}} R_{ik}^{\,b_{ik}} \ln\left(1 - \left(\frac{r_{jk}}{R_{ik}}\right)^{b_{ik}}\right) \tag{6}$$

For polybead chains, $u^{\text{ext}}(\mathbf{x}_j)$ is just the total nonbond interaction energy of i at $\mathbf{x}_j$, discounting interactions with atoms that have yet to be regrown. The second term is a fictitious energy that directs the atom i toward the nearest fixed atom k at the other end of the removed chain section. A generalized finite extendable nonlinear elastic (FENE) "spring" appears to be a suitable functional form, as it is widely used to represent the extensional energy of coarse-grained chain segments. Here, $r_{jk} = |\mathbf{x}_i - \mathbf{x}_k|$ is the distance from trial position j to atom k. The user-defined constants, a_{ik}, b_{ik}, and R_{ik}, could in principle be defined uniquely for each pair of atoms i and k but in practice are conveniently defined for each value of l_{ik}, the number of bonds between i and k. Once the sequence of n_χ atoms, χ, has been regrown, one may evaluate the total (fictitious) bias energy,

$$V(\chi) = \sum_{i=1}^{n_\chi} (v^{\text{int}}(\mathbf{x}_i) + v^{\text{ext}}(\mathbf{x}_i)) \tag{7}$$

and the continuum Rosenbluth weight,[4,31]

$$W(\chi) = \prod_{i=1}^{n_\chi} \frac{1}{n_i} \sum \sum_{j=1}^{n_i} \exp(-\beta v^{\text{ext}}(\mathbf{x}_j)_i) \tag{8}$$

The next step is to identify new coordinates for the remaining three atoms. Two general approaches have previously been postulated. The first is to find all possible solutions to this geometric constraint problem and to randomly select one of these with a suitable (e.g., Boltzmann) weighting, which is incorporated into the acceptance criterion.[20] The second is to find *one* possible solution and enforce detail balance explicitly by applying the same search procedure in the reverse move and requiring that the trimer coordinates thus obtained give an exact match of the original coordinates.[27] Either of these approaches can readily be used as part of the ICB method; here, we use a variant on the latter approach.

The nine Cartesian coordinates of these atoms are subject to four nearest-neighbor (bond length) and five second-neighbor (bond angle) distance constraints. Figure 2 shows a convenient assignment of the atomic coordinates and constraint distances in Cartesian space. Note that the atoms whose coordinates have just been generated by configuration bias (eqs 1−7) are here regarded as "fixed", for example, atoms g and h. This allows us to define nine constraints σ_n in terms of the nine variables s_n and the nine distances d_n. For example, for $n = 1$,

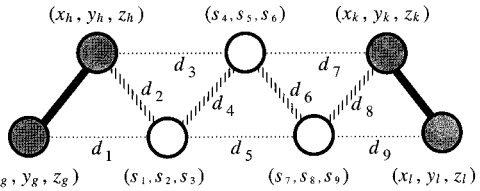

Figure 2. Solution of the geometric constraint equations for the last three atoms in an ICB move for a single linear chain. d_n denote the nine defined distance constraints for the nine variables s_n.

$$\sigma_1 = (s_1 - x_g)^2 + (s_2 - y_g)^2 + (s_3 - z_g)^2 - d_1^{\,2} \tag{9}$$

Each σ_n denotes the difference between the current separation of a pair of atoms and the desired (constraint) value, d_n. Hence a feasible solution for the bridging trimer may be generated by solving the simultaneous equations $\sigma_n = 0$ for $n = 1$ to 9. This can be achieved[27] by iteratively solving the linear system of equations

$$B_{mn}\delta s_n = -\sigma_m \tag{10}$$

where each element B_{mn} of the 9×9 matrix $\mathbf{B}$ is evaluated as

$$B_{mn} = \frac{\partial \sigma_m}{\partial s_n} \tag{11}$$

and δs_n is used to increment the coordinates,

$$s_n^{\,(\text{new})} = s_n^{\,(\text{old})} + \delta s_n \tag{12}$$

Equations 10−12, which form a Newton−Raphson procedure, can be solved repeatedly until convergence is attained,

$$\sum_{n=1}^{9} |\sigma_n| \le \epsilon_1 \tag{13}$$

where ϵ_1 is typically set to 10^{-10} Å. The Jacobian determinant for transforming a volume element from constraint space to Cartesian space[27] is then given by

$$J = \left|\frac{1}{\det \mathbf{B}}\right| \tag{14}$$

To perform such an iterative refinement, suitable arbitrary initial guess coordinates are required. Such a set of coordinates may be generated conveniently by the same configuration bias regrowth procedure as that used above, noting that the trimer atoms are *not* included in the Rosenbluth weight for the new subchain. A suitable form for the biasing potential is

$$v^{\text{ext}}(\mathbf{x}_j) = u^{\text{ext}}(\mathbf{x}_j) - a_{ik}(r_{jk} - R_{ik})^{b_{ik}} \tag{15}$$

In principle, this biased generation of the initial guess should encourage convergence toward energetically

favorable trimer solutions (alternatively, energy minimization could equally readily be used to generate energetically favorable initial coordinates). In practice, for polybead molecules, the correlation between initial guess and final trimer coordinates does not appear to be particularly close. Hence, it is straightforward to generate several sets of initial coordinates by "small" random displacements (<2.5 Å) that readily converge to different trimer solutions.

Such a procedure is used here to generate a set of trial solutions, $\{\psi_{j'}\}$, by independent Newton−Raphson refinement of n_{sol} different initial coordinate sets (typically 4−8). An arbitrary criterion for discarding all bar one of these solutions is needed. A convenient choice is to select a solution ψ at random using Boltzmann weighting, in other words,

$$P(\psi) = \frac{\exp(-\beta U(\psi))}{\sum_{j'=1}^{n_{sol}} \exp(-\beta U(\psi_{j'}))} \tag{16}$$

where the energy $U(\psi_{j'})$ of the j'th trimer solution $\psi_{j'}$ is a function of the coordinates $\{\mathbf{x}_{j'}\}$ of the atoms for that solution,

$$U(\psi_{j'}) = \sum_{i'=1}^{n_\psi} (u^{int}(\mathbf{x}_{j'})_{i'} + u^{ext}(\mathbf{x}_{j'})_{i'}) \tag{17}$$

Note that here n_{sol} is the number of solution *searches*. Hence, repeat occurrences of the same solution can occur, with equal weights, while unconverged solutions are assigned an infinite energy, corresponding to a weight of zero.

Overall, this approach to the trimer bridging problem is quite efficient, both for generating new trimer solutions and for recovering the original solution (to obey detail balance, as described below). It was adopted here largely because it is simpler to implement than a full solution search, not just for the linear homopolymer chains considered here but particularly for prospective extensions to, for example, branched or adsorbed chains, where many variants to the above constraint equations can occur. This approach to the trimer problem essentially reduces to that described by Pant and Theodorou[27] when $n_{sol} = 1$.

The original configuration of the selected atoms is then regrown, using the same sequence of CB additions (eqs 2−8) and the same trimer search procedure (eqs 9−17). For the CB additions, detail balance can be enforced via calculation of the ratio of probabilities for generating the old and new conformations. For the trimer coordinates, we enforce detail balance explicitly, by using the same procedure to identify one trimer solution for each of the forward and reverse transitions and rejecting the entire move if the reverse operation does not recreate the original trimer conformation.

For the CB regrowth, the Rosenbluth weight of the original conformation of the atoms χ is evaluated, via eq 8, by creating $n_i - 1$ (dummy) alternative positions for each atom i. The n_ith position is of course atom i's original coordinates.

Once the weights of the original conformation are evaluated by reconstruction of χ, a trial trimer is constructed by "blindly" following the steps in eqs 9−17. This creates n_{sol} candidates, but only one is selected via eq 16. If this does not match the original trimer, ψ, then the move must be rejected. This requirement is implemented by setting an upper limit to the difference between the original trimer coordinates s_n and the reconstructed coordinates s_n^{rec}, given by

$$\sum_{n=1}^{9} (s_n - s_n^{rec})^2 \le \epsilon_2 \tag{18}$$

where $\epsilon_2 = 10^{-10}$ Å^2 is a suitable choice. Note that typically the original trimer is recovered 50−80% of the time, due to its energetic favorability relative to competing trimer candidates.

We may now define an exact expression for the ratio of probabilities of suggesting the transition from the original state Ω to the new state Ω*, denoted $S_{\Omega\to\Omega^*}$, and the corresponding reverse transition, $S_{\Omega\to\Omega^*}$, in other words, constructing Ω from Ω*. This is given by the ratio of the product of probabilities for regrowing configurations χ and χ^*, via eq 5, times the ratio of volume elements in constraint space of the trimers ψ and ψ^*, via eq 14. Note that there is no energetic bias in the trimer selection, because we use a reversible procedure that requires the explicit unique selection of the correct old configuration. Hence, using the Rosenbluth weight definition of eq 8, we have

$$\frac{S_{\Omega^*\to\Omega}}{S_{\Omega\to\Omega^*}} = \frac{W(\chi^*)\exp(\beta V(\chi^*))J(\psi^*)}{W(\chi)\exp(\beta V(\chi))J(\psi)} \tag{19}$$

The new state may then be submitted to the familiar Metropolis trial,[4,32] which is used to accept or reject the complete move with a probability given by

$$\mathrm{Acc}_{\Omega\to\Omega^*} = \min\left(1, \frac{S_{\Omega^*\to\Omega}}{S_{\Omega\to\Omega^*}} \exp[-\beta(U(\Omega^*) - U(\Omega))]\right) \tag{20}$$

where $U(\Omega^*) - U(\Omega)$ gives the change in total energy of the system resulting from the MC move. Note that when eq 19 is substituted into eq 20, there is partial cancellation of the $U(\Omega)$ and $V(\chi)$ terms, leaving the trimer energies $U(\psi)$ and the fictitious "directing" contributions to $V(\chi)$. As a result, the acceptance of ICB moves is governed by the ratios of the Rosenbluth weights (as in CB),[28] trimer energies, and Jacobians (as in intramolecular rebridging),[27] as well as the fictitious contributions to the bias energy.

Molecular Model

The ICB algorithm was used to simulate alkane molecules in a vacuum at constant predefined temperatures. CH_2 and CH_3 functional groups were represented by equivalent united atom sites, with bond lengths and bond angles constrained to fixed values of 1.54 Å and 112°, respectively. The chosen interatomic potential[33] consisted of a Lennard-Jones interaction between nonbonded atoms, of the form

$$U(r_{ij}) = 4\epsilon\left(\left(\frac{\sigma}{r_{ij}}\right)^{12} - \left(\frac{\sigma}{r_{ij}}\right)^{6}\right) \tag{21}$$

where r_{ij} is the distance between atoms i and j. A Ryckaert−Bellemans form[22] was used for the torsion interaction as a function of dihedral angle ϕ,

$$U(\phi) = \sum_{n=0}^{5} A_n \cos^n(\phi) \qquad (22)$$

Values of the potential parameters σ, ϵ, and A_n are given in Table 1.

Table 2 lists the values of the adjustable coefficients used in the simulations. The number of trimer solution searches, n_{sol}, was set to 4, except where otherwise specified. Although these coefficients can be selected at the discretion of the user, considerable care is required to define an energetic field that leads to optimally vigorous sampling of configuration space.[30] The choice will most likely depend on the molecular system being studied, as well as the type of computer system that is used. Geometric rebridging operations are less vectorizable than configuration bias operations and occupy a larger fraction of the total CPU time for MC moves involving small chain sections. The values in Table 2 have been chosen heuristically as a suitable compromise for simulating alkane chains of varying length on Silicon Graphics R8000, Cray J90se, and NEC SX-4 computer systems.

Initial configurations for cyclic alkanes were generated by an iterative Cartesian refinement distance geometry procedure.[34] This rapidly generates independent molecule conformations that satisfy the bond length and bond angle constraints to within machine accuracy ($<10^{-12}$ Å).

Algorithm Validation

The ICB algorithm was tested by simulating a single "phantom" C_{16} chain, whereby all energetic interactions were switched off. The simulation consisted of 2×10^7 MC steps, of which 90% were ICB moves, with a (nonuniform) length distribution defined by random selection of the first and last atoms to be displaced. The remaining 10% were conventional CB moves of random length, to allow fluctuations of the chain ends. For such a system, we expect a uniform distribution of torsion angles. The actual distribution obtained from the simulation is shown in Figure 3.

Figure 4 shows the distribution of torsion angles obtained for a similar but shorter ICB/CB MC run, with a total of 1×10^6 MC steps, for an unperturbed chain with only torsional interactions. In the absence of nonbond interactions, it is also straightforward to derive analytically the predicted torsion angle distribution. Once again, the agreement is excellent, as required. This test, although not as sensitive as that in Figure 3, is important for ensuring that the real and fictitious contributions to the bias potential, and to the resultant Rosenbluth weights, are implemented in the correct balance.

It is evident that the ICB algorithm correctly samples the torsional degrees of freedom in these simple molecular systems, without any systematic bias. Equation 19 is thus demonstrated to rigorously obey the necessary detail balance condition for equilibrium Monte Carlo simulation.[4] The observed acceptance ratio for ICB Monte Carlo moves in the phantom chain calculations is about 5%. By contrast, pivot and configuration bias moves for such chains have 100% acceptance. Hence, it is already apparent that the ICB method is likely to be of only limited benefit in conformational sampling or simulation of isolated linear chains, except possibly in the long chain limit.

Table 1. Summary of Interatomic Potential Parameters for Polybead Model

	A_n/k_B (K)							
n	0	1	2	3	4	5	ϵ/k_B (K)	σ (Å)
	1082.5	−1418.1	−1530.7	356.9	3061.3	3674.3	49.1	3.94

Table 2. Values of Adjustable Coefficients Used in the ICB Algorithm[a]

l_{ik}	1	2	3	4	5	6	$x > 6$
R_{ik} (Å)	1.54	2.52	3.40	5.08	6.375	7.65	$1.275x$
a_{ik}/k_B (K)	10 000	10 000	50 000	10	10	10	10
b_{ik}	2	2	6	2	2	2	2
n_i	10	10	10	20	20	20	20

[a] l_{ik} denotes the number of atoms in the closed chain section from i (the atom currently being added) to k (the fixed atom at the other end).

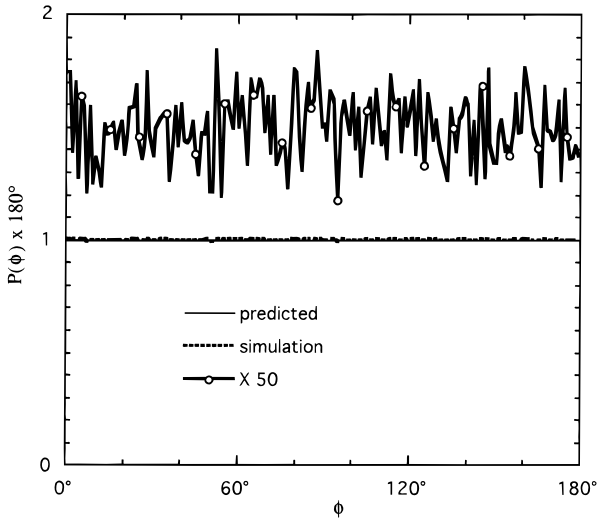

Figure 3. Predicted and simulated torsion angle distributions for a "phantom" linear C_{16} chain, with all interactions switched off. The simulation was performed using 2×10^7 MC moves (10% CB, 90% ICB). The expanded plot of the simulation data has been displaced vertically for clarity.

Performance aspects of the ICB algorithm were examined via a series of short benchmark runs (typically 10^6 MC steps) on isolated cyclic alkane molecules, with full interatomic interactions. It is worth noting that simulations of isolated flexible molecules can exhibit rapid and substantial atomic translations and bond vector reorientations, due to translation of, and rotation about, the molecule's center of mass. An appropriate measure of equilibration is thus to observe the internal relaxation via the decay of the torsional autocorrelation function (ACF). This may be defined as

$$f_{\cos\phi}(t) = \left\langle \frac{\langle \cos\phi(t)\cos\phi(0)\rangle - \langle\cos\phi(0)\rangle^2}{\langle \cos\phi(0)\cos\phi(0)\rangle - \langle\cos\phi(0)\rangle^2} \right\rangle \qquad (23)$$

where t represents the *simulation* time, in other words, the number of Monte Carlo steps, whereas the outer and inner angular brackets denote averages over different bonds and multiple time origins, respectively.

The benchmark results are summarized in Table 3. Note that where a range is given for n_Ω, this corresponds to a uniform distribution of move sizes within the specified range, inclusive of the lower and upper bounds. The 4-atom displacements correspond approximately to concerted rotation (ConRot) moves with one random "driver" angle; random angles are more suitable for

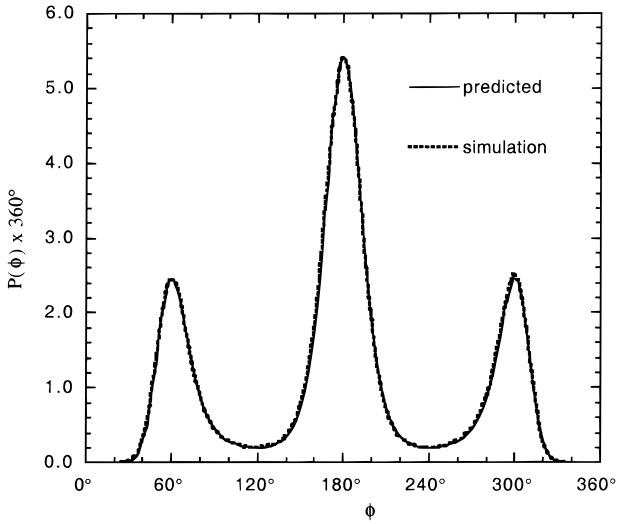

Figure 4. Predicted and simulated torsion angle distributions for a linear C_{20} chain at 450 K, in the presence of torsional interactions (eq 22), but with nonbond interactions switched off. The simulation was performed using 10^6 MC moves (10% CB, 90% ICB).

Table 3. Benchmarks for Simulation of Cycloalkane Molecules Pre-equilibrated at 300 K on SGI R8000, Cray J90se, and NEC SX-4 Computer Systems (single processor)[a]

N	n_Ω	n_{sol}	Acc %	δ_b	δ_m	τ_{SGI}/s	τ_{Cray}/s	τ_{NEC}/s
16	3	1	6.04	0.2	0.1	774	1498	1092
16	4	1	3.24	2.6	1.3	652	831	1443
16	4	4	4.57	2.9	1.4	574	466	961
16	7	4	0.26	3.7	1.1	299	326	335
16	10	4	0.01	0.8	0.1	303	527	204
16	4−10	4	1.09	4.0	1.2	261	221	257
16	4−10	1	0.79	2.0	0.7	312	176	384
16	4−10	8	1.81	6.5	2.0	170	541	350
100	4	4	6.29	4.4	1.9			2699
100	7	4	0.53	4.6	1.5			
100	10	4	0.11	2.7	0.8			
100	4−10	4	1.51	4.1	1.5			3331
100	4−30	4	0.40	0.9	0.3			4544

[a] N is the number of atoms in the molecule, n_Ω denotes the number of atoms displaced per MC move, and Acc denotes the overall ratio of accepted moves. δ_b and δ_m give the average number of transitions per 1000 MC steps between bond and molecular conformers, respectively. τ is the CPU time (in seconds) required for the torsional autocorrelation function to decay to a value less than 0.05, which provides a convenient comparative measure of the equilibration time.

isolated chains than the small driver angle range commonly used for bulk melts.[20]

The following general trends can be observed: (i) for C_{16} rings, the increased length ICB moves are 2−5 times more efficient than the 4-atom (ConRot-like) moves; (ii) for the less constrained C_{100} rings, 4-atom concerted rotation moves are more efficient; and (iii) the comparative advantage of displacements of large chain sections is more pronounced on computer systems with vector processors (although vectorization is still mediocre for these isolated molecule systems).

Interestingly, increasing the size of ICB displacements typically leads to the CPU time per MC step being *reduced*, because a higher proportion of moves can be rejected prior to the CPU-intensive trimer rebridging. This observation is examined in more detail in the next section.

Table 3 also details the rate of observing conformational transitions between the trans and gauche states for individual bonds and the resultant transitions between distinct conformers of the whole molecule. It can be seen that larger displacements lead to more bond transitions per molecular transition and, hence, more bond transitions per MC step. The increased efficiency of the larger displacements is likely to be due also to the reduced likelihood of reversal of such transitions. Given that algorithms such as ConRot work best when mixed with other MC moves[21] (simple translations, reptations, pivots, CB, etc.) it appears likely that ICB will be most useful in molecular systems where such moves are difficult or impossible to implement effectively, such as the cyclic alkanes considered here.

Table 4. Summary of Alkane MC Simulations[a]

	N	T	$t/10^6$	n_Ω	$\langle n_\Omega^* \rangle$	Acc	$\langle r_g^2 \rangle$	$\langle r_{1/2}^2 \rangle$
linear	16	300	10.0	1−13	6.7	0.7545	21.8	69.1
ring	16	300	9.0	4	4.0	0.0531	9.4	36.5
ring	16	300	10.0	4−10	4.7	0.0117	9.5	36.6
ring	16	450	1.0	4−10	5.3	0.0174	9.5	37.0
ring	30	300	10.0	4−24	4.8	0.0040	25.6	96.5
ring	50	300	4.0	4−30	5.1	0.0037	32.9	83.7
ring	50	300	3.0	4−44	5.3	0.0025	33.5	85.9
ring	100	300	10.0	4−30	5.0	0.0040	49.4	132.4
ring	100	300	3.0	4−94	5.2	0.0012	(54)	(124)

[a] N is the number of atoms in the molecule, T is the simulation temperature (in kelvin), and t is the number of MC steps. n_Ω denotes the number of atoms displaced per MC move, and $\langle n_\Omega^* \rangle$ is the average number of displaced atoms per successful MC move. Acc denotes the overall ratio of accepted moves. $\langle r_g^2 \rangle$ and $\langle r_{1/2}^2 \rangle$ represent the ensemble averaged values (in angstroms) of the radius of gyration squared and molecule half-length squared, respectively. The final run listed did not fully equilibrate.

Study of Cyclic Alkanes

Medium-length cyclic alkane molecules (C_{10}−C_{100}) are interesting in that they exhibit transitional behavior between that of small, highly constrained ring molecules (e.g., cyclohexane) and large, flexible polymer molecules (e.g., polyethylene). The segments in cyclic alkanes exhibit intramolecular constraints similar to those observed in bulk entangled chain systems, such as polymer melts, and in chains attached to surfaces. They thus provide a convenient subject for more detailed analysis of the performance of the ICB algorithm. Here, we examine isolated cyclic C_{16} (cyclohexadecane), C_{30} (cyclotriacontane), C_{50} (cyclopentacontane), and C_{100} (cyclohectane) molecules with full attractive nonbond interactions. The resultant conformations thus correspond to these molecules being in a vacuum or a poor solvent environment. The run details are given in Table 4. Where a range of move sizes, n_Ω, is specified, a uniform distribution between the two limiting values was used.

Figure 5 shows a typical plot of the ratio of accepted ICB moves as a function of the number of displaced atoms, n_Ω. Also shown is the fraction of moves where the CB regrowth phase of the move is completed for the first n_χ atoms; the other moves are prematurely terminated once it is apparent that the remaining atoms must be stretched beyond their maximum contour length in order to complete the intramolecular bridge. The gap between the two curves thus reflects the proportion of moves rejected due to either failed trimer rebridging or failure in the Metropolis trial condition (eq 20). The acceptance ratio is indeed a strong function of the number of atoms displaced. The drop is not simply exponential, as is observed for conventional CB or for ICB in bulk systems.[25] Nevertheless, it is apparent that

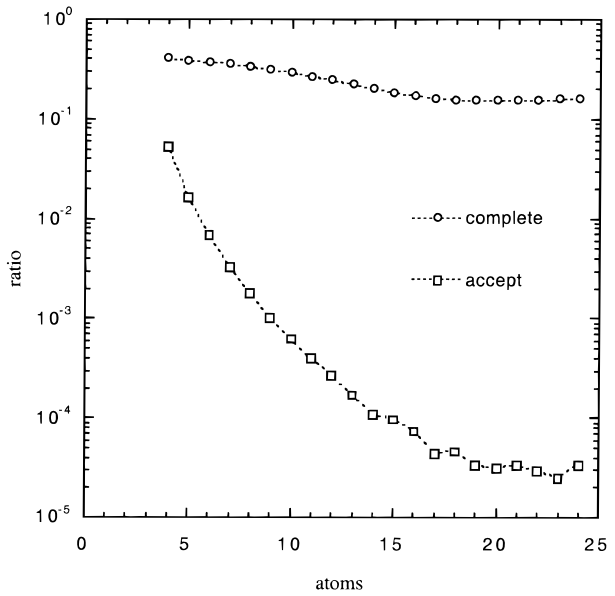

Figure 5. Acceptance ratio for ICB moves in cyclic C_{30} at 300 K, as a function of the number of displaced atoms, n_Ω (squares). The circles denote the fraction of moves for which the $n_\Omega - 3$ configuration bias operations are completed.

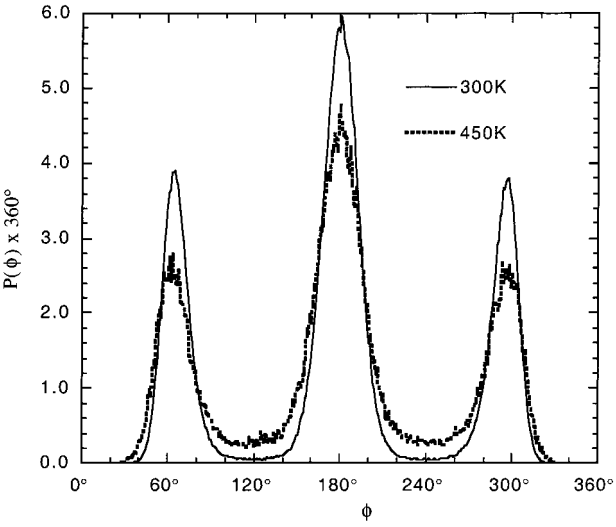

Figure 6. Distribution of torsion angles, ϕ, for cyclic C_{16} (cyclohexadecane) at 300 and 450 K.

moves involving more than about 20 atoms are marginally practical at best, given that present day computing resources are not suitable for simulations involving more than $\sim 10^9$ MC steps.

Figure 6 compares the torsion angle distribution functions for cyclohexadecane at 300 and 450 K. The latter agrees well with the MD results of Zhang and Mattice.[35] The relative ratios of the trans (t) and the two gauche (g^+, g^-) states are roughly the same at both temperatures. As a result, the radii of gyration at the two temperatures, listed in Table 4, are effectively the same. This is in marked contrast to the behavior widely observed in linear chains. The major effect of increasing temperature is thus to broaden the peaks in the torsion distribution and cause bonds to adopt conformations in the angle range corresponding to the gauche−trans energetic barriers.

Figure 7 displays the torsion angle distribution functions for the different ring molecules, in comparison with linear hexadecane. It is clear that the cyclic molecules

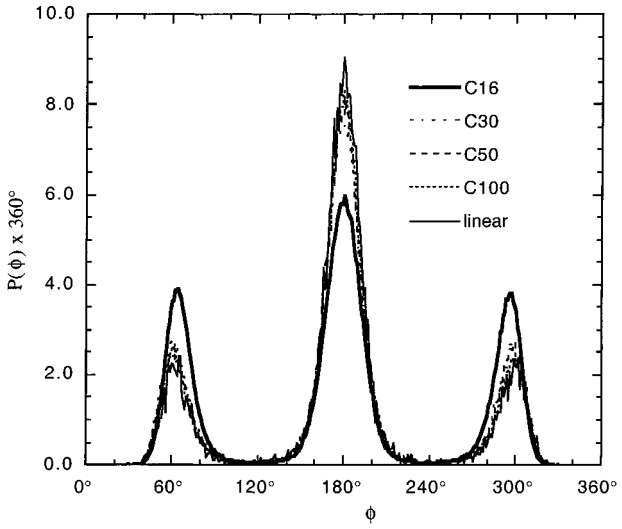

Figure 7. Distribution of torsion angles, ϕ, for cyclic C_{16}, C_{30}, C_{50}, and C_{100} molecules at 300 K. Also shown is the distribution for linear C_{16} at the same temperature.

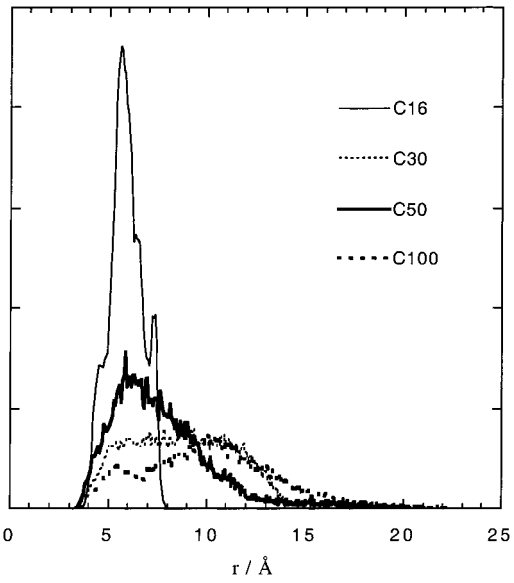

Figure 8. Distribution of molecule half-lengths, $r_{1/2}$, for cyclic C_{16}, C_{30}, C_{50}, and C_{100} at 300 K.

converge to the linear result in the large ring limit, as expected. Cyclohexadecane shows a ratio of gauche to trans states markedly larger than those of cyclotriacontane or the larger rings, and the gauche peaks are shifted slightly to less acute angles, due to the more strained molecular geometry imposed by the tighter ring constraint.

A familiar measure of the global chain extension characteristics of a linear chain molecule is the end-to-end distance. A suitable equivalent for cyclic molecules with n equivalent backbone atoms is the distance between atoms separated by $n/2$ bonds, $r_{1/2}$. When n is even, $r_{1/2}$ may be averaged over $n/2$ such pairs of atoms. Mean values of $r_{1/2}$ squared for different ring sizes are reported in Table 4. Figure 8 shows how the population distribution of $r_{1/2}$ varies for the different sized rings. Not surprisingly, the distributions become increasingly broad with increasing ring size. A striking feature is that the peak in the distribution at ~ 6 Å is much more pronounced for cyclopentacontane than for cyclotriacontane, although even cyclohectane has a pronounced

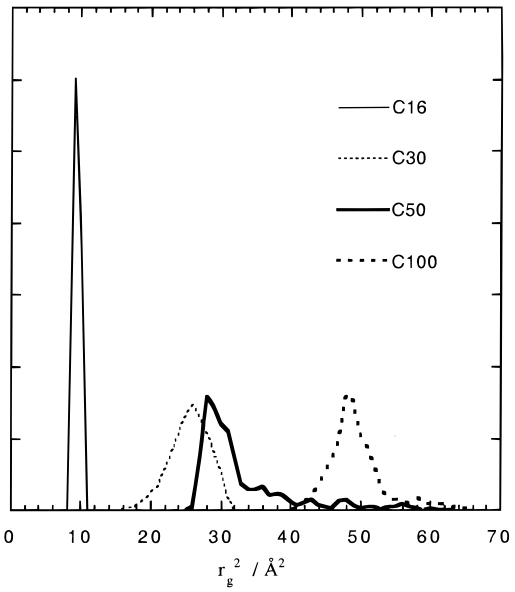

Figure 9. Distribution of radii of gyration, r_g, for cyclic C_{16}, C_{30}, C_{50}, and C_{100} molecules at 300 K.

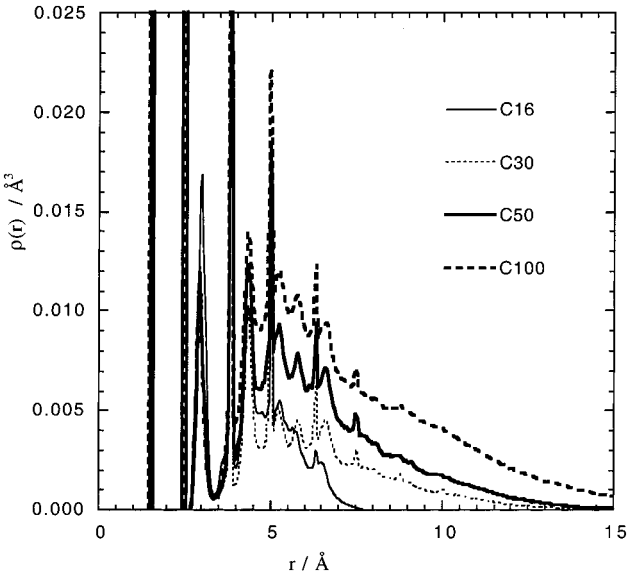

Figure 10. Pair density distribution functions for cyclic C_{16}, C_{30}, C_{50}, and C_{100} molecules at 300 K.

subpeak at this separation. This peak is close to the minimum separation of nonbonded united atoms, which is approximately 4 Å.

The corresponding distribution functions for the radius of gyration are depicted in Figure 9. Again, a qualitative change is apparent between C_{30} and C_{50}, in that the C_{50} ring is more compactly coiled relative to its contour dimensions. This behavior is consistent with a collapse transition for flexible ring molecules, as studied in more detail by Taylor et al[36] for ring polymers with square well interactions and noted briefly by Sundararajan and Kavassalis for cyclic alkanes.[37] The transition becomes sharper for larger rings and occurs at a higher temperature; we can deduce that for cyclopentacontane the transition temperature is in the vicinity of 300 K.

Figure 10 displays the pair density distribution function evaluated for each of the rings at 300 K. This function gives the average of the number density of atoms (absolute, not relative) at a given distance from an origin atom. The enhanced peak at approximately 3 Å for cyclohexadecane corresponds to the excess of gauche states for this molecule, as seen in Figure 7. Although obscured somewhat by the exaggerated nearest-neighbor peaks of the polybead model, a difference in overall density in the 4−8 Å range between cyclotriacontane and cyclopentacontane is also apparent. This is consistent with the difference in overall ring dimensions of the two molecules, as characterized in Figures 8 and 9.

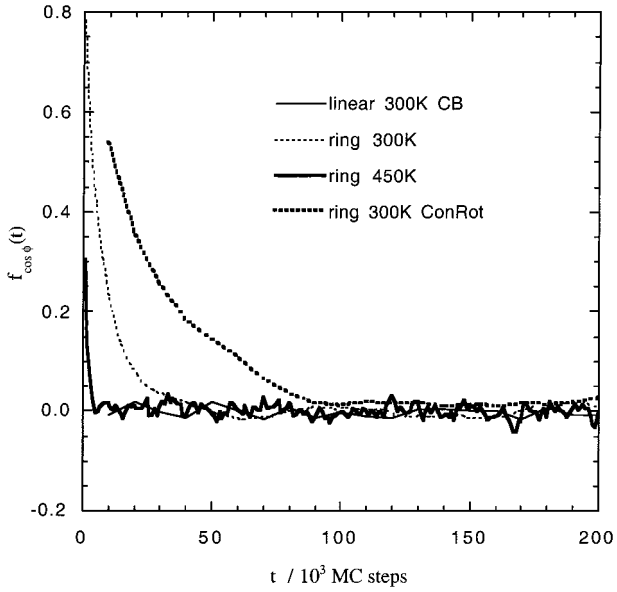

Figure 11. Torsion angle autocorrelation functions for cyclic C_{16} at 300 and 450 K, simulated by ICB Monte Carlo with the number of displaced atoms, n_Ω, varying uniformly between 4 and 10. The equivalent functions for $n_\Omega = 4$, corresponding to concerted rotation, and for linear C_{16}, simulated by conventional configuration bias Monte Carlo, are also shown.

Figure 11 compares torsion ACFs for ICB simulations of cyclohexadecane at 300 K. One of the runs was performed with 4-atom concerted rotation displacements ($n_\Omega = 4$) throughout, whereas the other run used displacements of 4−10 atoms. Decorrelation of the torsion angles requires roughly 4 times fewer MC steps when the larger range of displacements is used, even though the significantly lower acceptance ratio means that in fact fewer atoms are moved per MC step (see Table 4). Evidently, the larger displacements provide greater opportunity for rapid relaxation via cooperative torsional transitions about several atoms, in each accepted MC move. This effect counterbalances the decrease in acceptance ratio illustrated in Figure 5. Furthermore, the simulation with the larger range of displacements requires slightly less CPU time per step, because a significant proportion of moves are discarded before the computationally expensive trimer solution search is performed (see Figure 5). Clearly, the ICB algorithm with the range of atom displacements is more efficient for this application than 4-atom concerted rotation moves. It may be possible to further optimize the distribution of n_Ω for more efficient ICB simulation of this molecule; such further refinement has not been investigated in this work.

Figure 11 also depicts the torsion ACF for CB simulation of linear hexadecane. Here, the decay is virtually instantaneous, because the rotations performed by a successful CB move generate totally new and uncorrelated torsion angles. Also shown is the ACF for cyclo-

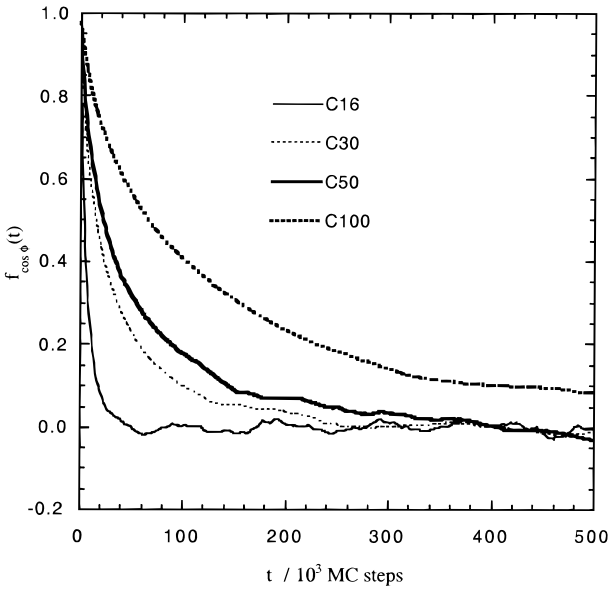

Figure 12. Torsion angle autocorrelation functions for cyclic C_{16}, C_{30}, C_{50}, and C_{100} at 300 K, simulated by ICB Monte Carlo. The n_Ω ranges are 4−10, 4−24, 4−30, and 4−30, respectively.

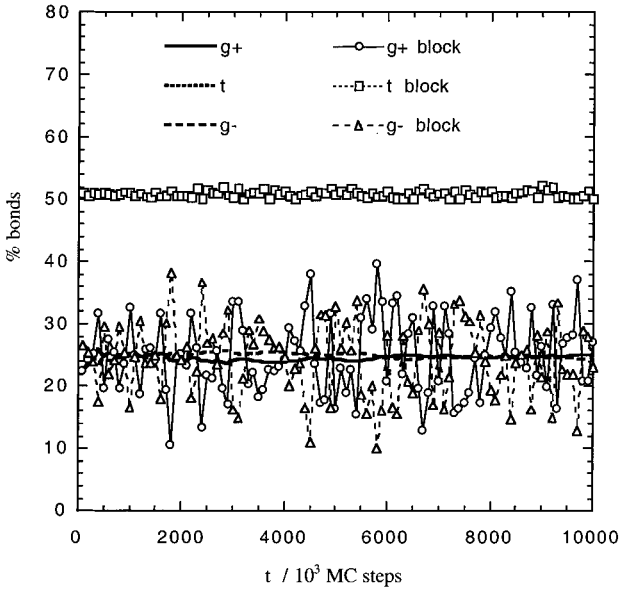

Figure 13. Running and block averages for the relative proportions of gauche and trans states in cyclic C_{16} simulated at 300 K by ICB Monte Carlo, using $n_\Omega = 4-10$.

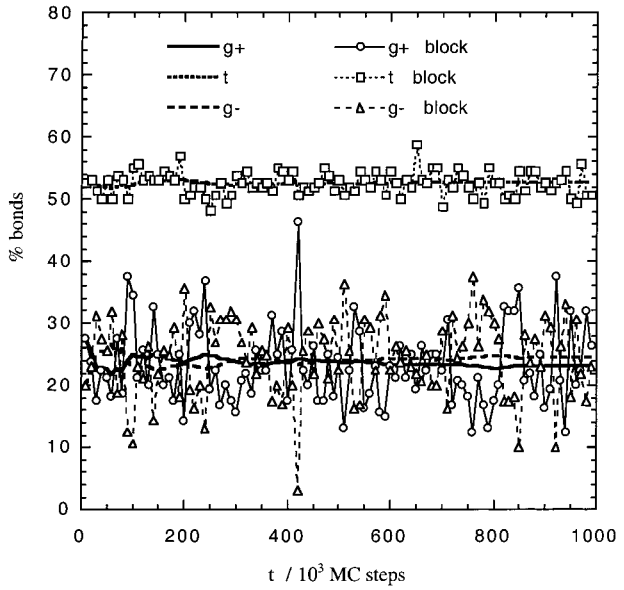

Figure 14. Running and block averages for the relative proportions of gauche and trans states in cyclic C_{16} simulated at 450 K by ICB Monte Carlo, using $n_\Omega = 4-10$.

hexadecane at 450 K, which vanishes to 0 in about 5000 MC steps. By contrast, the equivalent relaxation time for MD is greater than 1 ns,[35] which typically corresponds to 10^6 time steps. The ICB simulation runs at approximately 50 MC steps per CPU second on a Silicon Graphics R8000 workstation, whereas an equivalent MD simulation runs at approximately 100 time steps per CPU second. This means that ICB is approximately 2 orders of magnitude faster in generating independent (equilibrium) conformations for this ring molecule. Of course, the penalty for this speedup is that, unlike MD, the MC conformations are generated by artificial moves that cannot be compared with experimental dynamic quantities. The relative advantage of the MC algorithm also decreases with increasing ring size. This is shown clearly in Figure 12, which depicts how the torsion ACF varies for different sized rings at 300 K.

Another powerful measure of conformational sampling efficiency is to examine the variation in the population of trans and gauche states as a function of simulation time. This is shown for cyclohexadecane at 300 K in Figure 13. Clearly, a large range of different conformations is generated, as the block average proportions of the two gauche states ranges between 10% and 40%. However, the running average proportions reach and maintain their equilibrium values (±2%) after 10 000 MC steps. Interestingly, while the relative proportions of the two gauche states fluctuate markedly, the total proportion of trans states remains virtually static. The implication is that individual bonds can be efficiently transformed between g^+ and g^- states by ICB moves, despite the large energy barrier between the two states. During the entire simulation of 10^7 MC steps, a total of 43 640 transitions between the different bond conformation states were observed. The corresponding number of molecule conformer transitions was 11 998, implying that each conformational transition consisted on average of about 3.6 simultaneous transitions of individual bond states.

Similar behavior is observed at 450 K, as is seen in Figure 14. At this temperature, 31 392 bond transitions and 7892 molecular transitions were observed in 10^6 MC steps. After the first 10 000 MC steps, the running average populations of each state again vary by less than 2% from their long time values. By contrast, MD simulation at 450 K results in large, long-lived variations in the population distributions and takes at least 10 ns to attain a comparable degree of equilibration.[35] To the author's knowledge, nobody has attempted to perform a similar MD simulation at room temperature. By this measure, ICB requires approximately 2−3 orders of magnitude less CPU time than MD to generate a fully representative set of equilibrium conformations of this medium-sized cyclic molecule. In such comparisons, it should be emphasized that the performance difference will depend on the algorithm implementation (e.g., interatomic potential, constrained versus variable bond lengths/angles) as well as on the system under investigation (ring size, temperature, etc.).

The equivalent data for the largest ring studied, cyclohectane, is given in Figure 15. The relative propor-

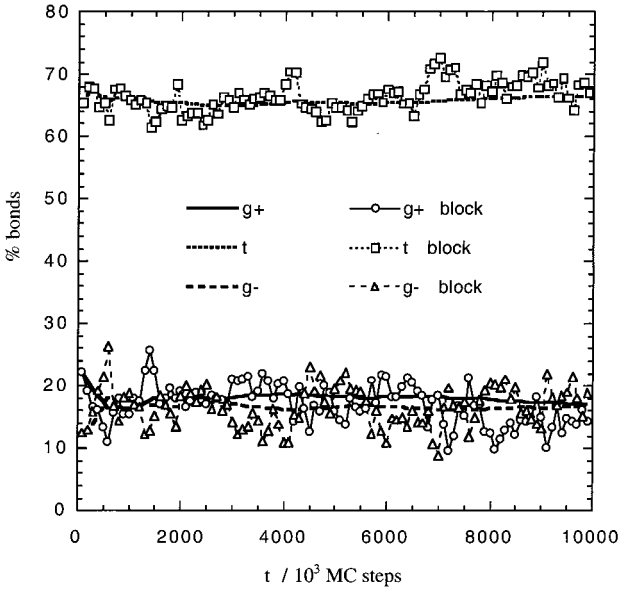

Figure 15. Running and block averages for the relative proportions of gauche and trans states in cyclic C_{100} simulated at 300 K by ICB Monte Carlo, using $n_\Omega = 4-30$.

tions of the two gauche states are less than those for cyclohexadecane (as expected from Figure 6), and the spread of populations is smaller, most likely due simply to the larger number of individual bonds. Conversely, there is an increased spread in the relative proportion of trans states in the large ring. Not surprisingly, the running averages of the populations take somewhat longer to equilibrate than those for the small ring.

ICB MC simulation is clearly a suitable method for more detailed investigation of the dimensions of realistic cyclic molecules. An obvious extension would be to consider ring conformations under good or neutral solvent conditions, by implementing increased repulsion contributions to the interactions between nonbonded atoms.[10] More accurate interatomic potentials, incorporating for example explicit hydrogen atoms and variable bond angles, would also be worth investigating.

Conclusions

We have seen how the internal configuration bias (ICB) Monte Carlo algorithm can be used to rigorously and efficiently sample the internal degrees of freedom of flexible chain molecules with bond length and bond angle constraints. As such, the algorithm is a suitable component for conformational searches and molecular simulations involving many important classes of compounds that have previously posed considerable difficulties. ICB can be regarded as a generalization of the successful concerted rotation (ConRot) algorithm, extended to chain sections of arbitrary length. The range of section lengths used in a particular calculation can then be chosen for a particular molecular system, based on some foreknowledge of the degree of intramolecular cooperativity required (e.g., coupled torsional transitions) to most efficiently sample the configuration space of that system. Furthermore, given that concerted rotation forms the basis of a range of more elaborate algorithms for such applications as cyclic peptides[24] and polydisperse polymers,[27] it is likely that ICB will similarly find application as a key element in future MC algorithms for a wide variety of covalently bonded molecular systems.

Acknowledgment. The author thanks Tom Spurling, Graeme Moad, Billy Todd, and Daan Frenkel for valuable discussions. Portions of the code used in the simulation described in Figure 4 were kindly provided by Vlasis Mavrantzas and Doros Theodorou. The financial support of the Cooperative Research Centre for Polymers (CRC-P) and extensive access to supercomputer facilities at the High Performance Computing and Communications Centre (HPCCC), are gratefully acknowledged.

References and Notes

(1) Flory, P. J. *Statistical Mechanics of Chain Molecules;* Interscience: New York, 1969.
(2) Leach, A. R. *Molecular Modelling: Principles and Applications;* Addison-Wesley Longman: Harlow, 1996.
(3) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids;* Clarendon Press: Oxford, 1987.
(4) Frenkel, D.; Smit, B. *Understanding Molecular Simulation;* Academic Press: San Diego, 1996.
(5) Sadus, R. J. *Molecular Simulation of Fluids;* Elsevier: Amsterdam, 1999.
(6) Blaney, J. M.; Dixon, J. S. *Rev. Comput. Chem.* **1994**, *5*, 299.
(7) Mattice, W. L.; Suter, U. W. *Conformational Theory of Large Molecules;* Wiley: New York, 1994.
(8) Rehahn, M.; Mattice, W. L.; Suter, U. W. *Adv. Polym. Sci.* **1997**, *131/132*, 1.
(9) Yoon, D. Y.; Flory, P. J. *Macromolecules* **1976**, *9*, 294.
(10) Destrée, M.; Lyulin, A.; Ryckaert, J.-P. *Macromolecules* **1996**, *29*, 1721.
(11) Honeycutt, J. D. *Comput. Theor. Polym. Sci.* **1998**, *8*, 1.
(12) Lal, M. *Mol. Phys.* **1969**, *17*, 57.
(13) Leontidis, E.; Forrest, B. M.; Widmann, A. H.; Suter, U. W. *J. Chem. Soc., Faraday Trans.* **1995**, *91*, 2355.
(14) Uhlherr, A.; Theodorou, D. N. *Curr. Opin. Solid State Mater. Sci.* **1998**, *3*, 544.
(15) Dijkstra, M.; Frenkel, D.; Hansen, J.-P. *J. Chem. Phys.* **1994**, *101*, 3179.
(16) Escobedo, F. A.; de Pablo, J. J. *J. Chem. Phys.* **1995**, *102*, 2636.
(17) Yong, C. W.; Clarke, J. H. R.; Freire, J. J.; Bishop, M. *J. Chem. Phys.* **1996**, *105*, 9666.
(18) Vendruscolo, M. *J. Chem. Phys.* **1997**, *106*, 2970.
(19) Go, N.; Scheraga, H. A. *Macromolecules* **1970**, *3*, 178.
(20) Dodd, L. R.; Boone, T. D.; Theodorou, D. N. *Mol. Phys.* **1993**, *78*, 961.
(21) Leontidis, E.; de Pablo, J. J.; Laso, M.; Suter, U. W. *Adv. Polym. Sci.* **1994**, *116*, 283.
(22) Ryckaert, J.-P.; Bellemans, A. *Chem. Phys. Lett.* **1975**, *30*, 123.
(23) Palmer, K. A.; Scheraga, H. A. *J. Comput. Chem.* **1991**, *12*, 505.
(24) Deem, M. W.; Bader, J. S. *Mol. Phys.* **1996**, *87*, 1245.
(25) Uhlherr, A. *Comput. Theor. Polym. Sci.*, in press.
(26) Uhlherr, A. *Proceedings of the 37th International Symposium on Macromolecules;* Gold Coast, Australia, 1998; p 621. Uhlherr, A.; Mavrantzas, V. G.; Theodorou, D. N., to be published.
(27) Pant, P. V. K.; Theodorou, D. N. *Macromolecules* **1995**, *28*, 7224.
(28) Frenkel, D.; Mooij, G. C. A. M.; Smit, B. *J. Phys.: Condens. Matter* **1992**, *4*, 3053.
(29) de Pablo, J. J.; Laso, M.; Suter, U. W. *J. Chem. Phys.* **1992**, *96*, 2395.
(30) Dellago, C.; Bolhuis, P. G.; Csajka, F. S.; Chandler, D. *J. Chem. Phys.* **1998**, *108*, 1964.
(31) Rosenbluth, M. N.; Rosenbluth, A. W. *J. Chem. Phys.* **1955**, *23*, 356.
(32) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. *J. Chem. Phys.* **1953**, *21*, 1087.
(33) Dodd, L. R.; Theodorou, D. N. *Adv. Polym. Sci.* **1994**, *116*, 249.
(34) Uhlherr, A. to be published.
(35) Zhang, R.; Mattice, W. L. *J. Chem. Phys.* **1993**, *98*, 9888.
(36) Taylor, M. P.; Mar, J. L.; Lipson, J. E. G. *J. Chem. Phys.* **1997**, *106*, 5181.
(37) Sundararajan, P. R.; Kavassalis, T. A. *Macromolecules* **1997**, *30*, 5172.

MA9908595