2-6-2018

Innovative Monte Carlo Methods for Sampling Molecular Conformations

Aliasghar Sepehri
Louisiana State University and Agricultural and Mechanical College, asepeh1@lsu.edu

INNOVATIVE MONTE CARLO METHODS FOR SAMPLING MOLECULAR CONFORMATIONS

A Dissertation

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfilment of the
requirements for the degree of
Doctor of Philosophy

in

The Department of Chemistry

by
Aliasghar Sepehri
B.S., Amirkabir University of Technology, 2007
M.S., Amirkabir University of Technology, 2010
May 2018

ACKNOWLEDGEMENTS

I would like to take this opportunity to express my gratitude and appreciation to all people who helped me reach my objectives in the PhD program. First of all, I would like to thank my advisor Dr. Bin Chen for mentoring me for more than five years. Second, my colleague Dr. Troy Loeffler who had many valuable conversations with me that led to novel ideas for developing new methodologies. Third, my advisory committee, Dr. Kenneth Lopata, Dr. Daniel Kuroda, and Dr. Dorel Moldovan, for their advice in my general exam as well as physical chemistry seminars. I also like to thank National Science Foundation for financial support, Louisiana Optical Network (LONI) and Louisiana State University High Performance Computing (LSU-HPC) for supplying me with the computational resources, and my family members for their emotional support.

It is also important for me to remember those people who helped me during my Bachelor and Master of Science in Amirkabir University of Technology, in particular, Dr. Hamid Modarress who taught me physical chemistry, quantum chemistry, statistical mechanics, etc., so I had a good background to enter PhD program.

TABLE OF CONTENTS

ABSTRACT

Sampling molecular conformations is an important step in evaluating physical, mechanical, hydrodynamic, and optical properties of flexible molecules especially polymers. One powerful method for this purpose is configurational-bias Monte Carlo in which one random segment of a molecule is chosen, all segments toward one random end are removed, and then regrown segment by segment to produce a new geometry to be accepted/rejected according to probability laws. The advantage of this method is the ability to generate acceptable conformations that are favorable for intra- and intermolecular energies to save computational costs. However, when there are several interdependent energetic terms, trial generation can be very time consuming because a trial must be generated that is satisfactory for all energetic terms. There are two important cases where a number of intramolecular energies are coupled: bending angle energies in a branched point, and bending and torsional angle energies for growing segments between two fixed points.

According to probability laws, if trials are generated according to their probability density function, all trials will be accepted. The basic idea of the methods, which have been developed for the two above cases, is to generate trials that are close to the Boltzmann distributions of intramolecular energies. It has been proved that new methods are faster and more efficient than traditional methods. One of the methods for generating bending angle trials have been used in nucleation simulations of flexible amine molecules which accelerates simulation process by four to five folds.

CHAPTER 1. INTRODUCTION

1.1. Importance of molecular conformation

Molecular conformation or molecular geometry is the arrangement of atoms in a molecule. This arrangement can be defined according to bond lengths, bending angles, and torsional angles. In addition to properties that define molecular dimensions, such as end-to-end distance[1] and radius of gyration,[2] there are many physical, mechanical, hydrodynamic, and thermodynamic properties that are highly dependent on molecular conformations. Some of these properties are dipole moment,[3] light scattering,[4] X-ray scattering,[5] NMR spectroscopy,[6] viscosity,[7] elasticity,[8] diffusion,[9] pH,[10] and second virial coefficient.[11] In biological molecules, such as proteins, the conformation of an antibody is crucial for targeting diseases.[12] Statistical mechanics provides computational tools for evaluating these properties by averaging over molecular conformations. Thus, sampling molecular conformations is a key step in accurate calculations.

Molecular simulation methods, such as molecular dynamics (MD) and Monte Carlo (MC),[13] are utilized for sampling a system. In MD, the equation of motion is solved numerically to calculate the position and the velocity of each particle at each time. In MC, positions of particles are sampled by proposing random moves that are accepted or rejected according to probability laws. According to statistical mechanics, the probability density function of a system is proportional to its Boltzmann distribution, $e^{-\beta U}$, where $\beta = (k_B T)^{-1}$ (k_B is the Boltzmann constant and T is the temperature) and U is the potential energy of the system.

MD has the advantage to simulate time-dependent and nonequilibrium phenomena. On the other hand, MC moves are more efficient to jump between different energy regions. Jorgensen and Tirado-Rives demonstrated[14] that for sampling molecular conformations, MC is 1.6-3.8 times faster than MD because MD is very likely to be trapped by internal energy barriers such as

torsional energy. It must be noted that in their MC simulation, traditional MC moves have been implemented whereas in this study, advanced MC methods have been developed that are much more efficient.

1.2. Background in probability and Monte Carlo

A probability density function, f, for a continuous variable, x, has three properties[15]

1. For each value of x, $f(x) \geq 0$

2. The probability density function is normalized

$$\int_{-\infty}^{+\infty} f(x)dx = 1 \tag{1.1}$$

3. The probability of finding x between two values, x_1 and x_2, where $x_1 \leq x_2$, is calculated by

$$P(x_1 \leq x \leq x_2) = \int_{x_1}^{x_2} f(x)dx \tag{1.2}$$

The cumulative distribution function is defined as

$$F(x) = \int_{-\infty}^{x} f(t)dt \tag{1.3}$$

According to Eq. (1.1), for each value of x, we have

$$0 \leq F(x) \leq 1 \tag{1.4}$$

Thus, Eq. (1.2) can be written as

$$P(x_1 \leq x \leq x_2) = F(x_2) - F(x_1) \tag{1.5}$$

In order sample the probability density function, random variable x must be generated from $f(x)$. The most straightforward method for this purpose is inverse transform. This method, which is based on Eq. (1.4), generate a uniform random number R on (0, 1) and calculate its corresponding random variable as

$$x = F^{-1}(R) \tag{1.6}$$

where F^{-1} is the inverse function of F. This method is applicable for simple probability density functions, such as sin x, where F and F^{-1} are calculated easily.

For more complicated functions, accept-reject scheme[16] can be used. In this algorithm, in order to sample probability density function f, a random variable x is generated from a simpler function g (e.g., uniform function), which is called generation function, and a uniform random number R on (0, 1) until $\frac{f(x)}{cg(x)} \geq R$, where c is a constant to ensure that the fraction $\frac{f(x)}{cg(x)}$ is between 0 and 1.

The accept-reject method is very efficient for probability density functions with few variables. But, when there are many variables that can affect each other, this method becomes very time-consuming to generate an acceptable set of variables. For instance, in a liquid system with many molecules, random generation of all particles positions is very likely to produce a system with high energy because of probable molecular overlaps. Consequently, $e^{-\beta U}$ is very low that leads to trial rejection. In order to sample these systems, Markov chain process[17] is used. In this process, it is assumed that the probability of a system to be at each state only depends on the previous state and it is independent of states prior to the previous states. In other words, if the current state of the system is called $\mathbf{x}_o$ and a new state, which is called $\mathbf{x}_n$, is generated from $\mathbf{x}_o$. The probability ratio of accepting the forward move to accepting the reverse move is

$$\frac{acc(\mathbf{x}_o \to \mathbf{x}_n)}{acc(\mathbf{x}_n \to \mathbf{x}_o)} = \frac{\frac{f(\mathbf{x}_n)}{g(\mathbf{x}_o \to \mathbf{x}_n)}}{\frac{f(\mathbf{x}_o)}{g(\mathbf{x}_n \to \mathbf{x}_o)}} \tag{1.7}$$

where arrow $\to$ means from one state to the other, f is the probability density function, g is the generation probability, and acc is the accepting probability. Eq. (1.7) can be written in the following form that is called detailed balance condition or microscopic reversibility

$$f(\mathbf{x}_o)\pi(\mathbf{x}_o \to \mathbf{x}_n) = f(\mathbf{x}_n)\pi(\mathbf{x}_n \to \mathbf{x}_o) \tag{1.8}$$

where π is the transition probability which is the product of generation and accepting probabilities. The acceptance rate, which is also the ratio of accepted moves to attempted moves, is defined as

$$P_{acc} = \min \left[1, \frac{acc(\mathbf{x}_o \to \mathbf{x}_n)}{acc(\mathbf{x}_n \to \mathbf{x}_o)}\right] \tag{1.9}$$

A uniform random number R is generated on (0, 1). If $P_{acc} \geq R$, the system goes to state $\mathbf{x}_n$, otherwise, it stays at state $\mathbf{x}_o$. According to Eqs. (1.7) and (1.9), if the new states are generated according to the probability density function, all moves will be accepted.

In order to generate new states, several algorithms have been proposed. One of the traditional algorithms is Metropolis sampling[18] where the generation probability is symmetric, i.e., $g(\mathbf{x}_o \to \mathbf{x}_n) = g(\mathbf{x}_n \to \mathbf{x}_o)$, and the Boltzmann distribution describes the probability density function. So, the acceptance rate is

$$P_{acc} = \min \{1, \exp(-\beta[U(\mathbf{x}_n) - U(\mathbf{x}_o)])\} \tag{1.10}$$

Two common moves in Metropolis sampling are translation and rotation where one molecule is chosen randomly and translated by a random displacement or rotated by a random angle to generate a new state. The potential energies of new and old states are calculated to accept or reject the move according to Eq. (1.10).

1.3. Flexible molecules

Metropolis algorithms are efficient for sampling positions of molecules in a system where a whole molecule is moved. However, it is not efficient for sampling conformations of a molecule because random displacement of one atom or one segment in a molecule can cause huge energy penalty due to the intramolecular interactions inside the molecule.

One of the earliest methods that has been proposed to sample conformations of a linear chain is the self-avoiding walk (SAW)[19-20] on square or cubic lattices for two or three dimensions

respectively. In this model, bond lengths are constant and equal to lattice constant. Each segment of the chain can occupy one lattice site. So, in order to generate conformations of a chain, the chain walks randomly on the lattice by occupying lattice sites segment by segment. Thus, bending angles can be either 90^{o} or 180^{o}. If the chain crosses an occupied site, the conformation will be rejected due to excluded volume repulsions. As the chain length increases, more attempts are likely to be rejected (attrition problem). One solution to this problem was proposed by Rosenbluth and Rosenbluth[21] to avoid occupied sites at each step by choosing one of the available sites. In this method, each grown chain is weighted to count all conformations equally (unbiased sampling). Another solution to the attrition problem is the enrichment method.[22-23] In this approach, walking a long chain is done in n steps where in each steps, p short chains with lengths s are generated. Successfully grown chains are attempted to grow further for another s walks in the next step. Grassberger[24] combined Rosenbluth-Rosenbluth and enrichment methods, so that very long chains can be generated in a lattice.

After generating a configuration for a long chain, the molecule can relax to generate other configurations through Markov chain processes. In these Monte Carlo moves, such as end rotation, kink jumping,[25] crankshaft,[26] slithering snake,[27] and pivot algorithm,[28] one or a few segments are relocated to new lattice sites in a way that fixed bond lengths are preserved to generate a new valid configuration. The new configuration is accepted according to the detailed balance condition.

1.4. Configurational-bias Monte Carlo

Another method that can be used in both lattice and off-lattice model is configurational-bias Monte Carlo (CBMC).[13] This method was first proposed to calculate chemical potential[29] using particle insertion method[30] in a lattice model.[31] The first version was similar to Rosenbluth-

Rosenbluth SAW with this difference that CBMC satisfies detailed balance condition to yield unbiased sampling. Since molecules in lattice models can only take fixed bond lengths and few bending angles, CBMC was extended to the off-lattice (or continuous) model[32-35] to consider strong intramolecular interactions. In a CBMC move, a random segment of a random molecule is chosen; all segments toward one end are removed, and then, regrown segment by segment to generate a new conformation. In the growth of a segment, l, K_{Trial} trials are generated and one of them (say ith trial) is selected with this probability

$$P_{\text{select}}(i) = \frac{\exp(-\beta U(i))}{W_l} \tag{1.11}$$

with

$$W_l = \sum_{i=1}^{K_{\text{Trial}}} \exp(-\beta U(i)) \tag{1.12}$$

The Rosenbluth weight for growing N segments is

$$W = \prod_{l=1}^{N} W_l \tag{1.13}$$

The new conformation is accepted with the probability of min[1, $W(n)/W(o)$] where o and n stand for old and new conformations respectively.

Since calculating intramolecular interactions is inexpensive in comparison with intermolecular interactions, it is computationally efficient to decouple them.[36] The probability density function of intramolecular interactions can be written as $J\exp(-\beta U^{\text{intra}})$, where J is the Jacobian factor and U^{intra} is the sum of all intramolecular energies. Each trial is generated according to this function using accept-reject method, which is also known as Boltzmann rejection method,[37] and one trial is selected according to intermolecular energies with this probability

$$P_{\text{select}}(i) = \frac{\exp\left(-\beta U^{\text{inter}}(i)\right)}{W^{\text{inter}}} \tag{1.14}$$

with

$$W^{\text{inter}} = \sum_{i=1}^{K_{\text{Trial}}} \exp\left(-\beta U^{\text{inter}}(i)\right) \tag{1.15}$$

where U^{inter} is the intermolecular energy. This method had been used to study linear[38] and branched[39-42] molecules until Vlugt et al. showed[43] that at a branch point, all branches must be grown simultaneously to yield correct distributions because several bending angles are coupled together. Their solution to this problem was to run a minor internal MC simulation to generate the positions of the branches simultaneously.

Macedonia and Maginn[44] proposed a branch point sampling method to deal with this problem where a set of correctly distributed molecular fragments are prepared and stored in advance to be used during the simulation. Apart from the large memory requirement, this method may not be used to generate any geometry outside of these pre-tabulated ones.

Another solution to this problem is coupled-decoupled CBMC (CD-CBMC)[45] where K_{Trial} trials are generated according to the Jacobian term (i.e., the $\sin\theta$ term), which is close to a uniform distribution, and one of them is selected based on the following Boltzmann probability distribution of intramolecular energies:

$$P_{\text{select}}(i) = \frac{\exp\left(-\beta U^{\text{intra}}(i)\right)}{W^{\text{intra}}} \tag{1.16}$$

with

$$W^{\text{intra}} = \sum_{i=1}^{K_{\text{Trial}}} \exp\left(-\beta U^{\text{intra}}(i)\right) \tag{1.17}$$

The growth of a branch point in CD-CBMC consists of two parts; in part 1, all bending angles of each branch with previously grown atoms are chosen; in part 2, the dihedral angles between planes made by growing angles are determined. These two parts are implemented independently and the product of their Rosenbluth weights appears in the detailed balance

condition. Since trials are generated almost uniformly despite the fact that bending and dihedral distributions are very narrow and nonuniform, many trials are required to be generated to produce an acceptable conformation. Thus, trial generation can become the most time-consuming part as revealed from a recent profiling of our nucleation MC code, where the size of the nucleation system is fairly small, not more than 100 molecules typically.[46] The angle generation was also found to be the most expensive component for the Gibbs ensemble MC (GEMC)[47-49] in phase equilibrium calculation[50-52] even for systems containing a few hundred molecules. For example, we repeated the phase equilibrium calculation reported for an n-heptane system[53] (with 300 molecules and a liquid box of 40Å) and found that more than 60% of the computer time was spent on the generation of the intramolecular angles. For an isolated molecule in a gas phase, the angle generation consumed over 99% of the computer time. In addition, since the dihedral angle distributions depend on the selected bending angles, while the two parts are performed independently, high acceptance rates cannot be obtained for highly branched molecules. For example, the acceptance rate for growing 2,2-dimethylpropane does not exceed 65% even with 10000 trials.[54]

Martin and Frischknecht[54] offered a solution, which is based on the energy bias scheme by Snurr et al.,[55] to the problems of CD-CBMC by generating trials according to an arbitrary distribution, such as Gaussian,[56] whose parameters are calculated during the simulation. An appropriate fitting approach is essential for this approach to achieve efficient sampling. They also coupled parts 1 and 2 to attain fairly good acceptance rates for branched molecules. This method works well for linear molecules, but needs many trials for branched molecules because Gaussian distribution cannot predict dihedral distribution correctly.

In chapter 2, two novel methods, the density-guided and the Jacobian-Gaussian, are explained for generating bending angle trials for linear and branched molecules.

Although CBMC can sample oligomers with a small number of segments successfully, it cannot be applied to polymers and cyclic molecules. In the case of polymers, growing many segments reduces the acceptance rate. Thus, CBMC can be used for the segments close to the ends, but any regrowth involving inner segments is very likely to be rejected. For cyclic molecules, since CBMC regrows the molecule segment by segment and does not determine the position of the last segment at the beginning of the growth procedure, it is very unlikely to generate a cycle with acceptable conformation. For this problem to be overcome, techniques have been proposed and they can be categorized into three groups depending on how the intramolecular interactions are treated.

In the first group, intramolecular interactions are ignored and all segments are connected to each other with fixed bond lengths. Thus, only nonbonded or intermolecular interactions are considered. One of the earliest methods uses the so-called crankshaft move[57-59] in which one segment is chosen randomly and then rotated by a random angle around the line passing its two neighboring segments to produce a new conformation. Escobedo and Pablo[60-61] developed extended continuum CBMC methods for linear, branched, and cross-linked molecules. In these methods, two segments for linear molecules and more than two segments for branched and cross-linked molecules are chosen randomly, and the segments between them are removed and regrown to produce a new conformation. When growing each segment, the direction of the growth (i.e. the polar and the azimuth angles) is generated uniformly from the available space, which is determined using geometrical equations, to ensure the closure of the chain. Another approach is to use the biasing probability function for each growth direction that guides the

growing segment toward the final segment. Because of the absence of the intramolecular interactions, the biasing probability function can be counted[62] for a lattice or calculated by integration[63] over the continuous space.

In the second group, bond lengths and bending angles are fixed at their equilibrium values, while torsional angles are allowed to vary under a given torsional potential. In each move, at least three segments[64] are relocated to generate a new configuration. One method in this group is called concerted rotation (CONROT),[65] where several sequential segments are chosen randomly. The torsional angles of the segments before and/or after the selected segments are changed to random values. The constraint equations (i.e., defined by the fixed bond lengths and fixed bending angles) are then solved numerically to find the new positions for these selected segments. Wu and Deem[66] showed that there are at most 16 solutions for these equations, and all answers must be calculated to satisfy the detailed balance condition. CONROT can be combined with CBMC[67-68] for cyclic peptides where the cyclic backbones are sampled using CONROT and the side chains are regrown with regular CBMC. Uhlherr[69] developed the internal configurational bias method in which a finite, extendable, nonlinear, and elastic biasing probability function is utilized between the growing and the final segment. The last three segments are regrown using the CONROT move to close the chain.

In the third group, a semiflexible model is used where bond lengths are fixed at their equilibrium values and bending and torsional angles are allowed to vary according to certain potential functions. Shah and Maginn[70] utilized a fragments library containing different conformations of cyclic fragments (e.g., cyclohexane and methylcyclohexane) in trial generation. This method requires large memory storage. Ulmschneider and Jorgensen[71] extended CONROT to flexible bending angles where several sequential segments are chosen randomly and their

bending and torsional angles are perturbed to create a new conformation. Because of the huge bending energy penalty and requirement for chain closure, only small angle perturbations can lead to acceptable conformations which increase sampling time. Rebridging configurational bias[72] and self-adapting fixed end points CBMC[73-74] extended regular CBMC to the regrowth of inner segments. In both methods, each trial is weighted by a biasing probability function that is assumed to be a function of the distance between the growing and the last fixed segment calculated either before or during the course of simulation. Despite the use of a large number of trials, the acceptance rates of these two methods for growing two, three, and four segments are approximately 40%, 20%, and 10%, respectively, and are even lower for higher number of segments. A low acceptance rate occurs for two reasons. First, the positions of the last few segments determine several tightly coupled bending and torsional angles, e.g., relocating even just two sequential segments in a linear chain, can lead to the change of up to four bending and five torsional angles. These two methods cannot include all these energetic terms in the trial generation step, and trial selection of each segment is performed sequentially, which ignores the coupling (or interdependencies) between these angles. Low acceptance rates were also observed for regular CBMC when using it on a highly branched molecule for similar reasons (i.e., the position of one branch simultaneously determines several tightly coupled intramolecular angles). Second, since each trial is weighted by both the Boltzmann factor and the biasing probability function, the final weight of each trial is not necessarily energetically favorable which may lead to a selection of inappropriate trial positions and consequently further decrease in the acceptance rate. In addition, because the biasing probability function must be evaluated between each pair of segments with different number of growing segments between them, this evaluation becomes

more computationally expensive and requires higher memory storage, in particular, for the case of polymers with different sequential orders of segments (e.g., proteins).

In chapter 3, a novel method is explained to improve the efficiency of fixed end points CBMC.

One of the developed methods for bending angle trial generation is tested in nucleation simulation of amines in chapter 4.

CHAPTER 2. BENDING ANGLE TRIAL GENERATION

2.1. Introduction

In this chapter, we explain the density-guided and the Jacobian-Gaussian methods for bending angle trial generation. A harmonic bending angle potential is used for angle θ as follows:

$$U_{\text{bend}}(\theta) = \frac{1}{2} k_\theta (\theta - \theta_0)^2 \tag{2.1}$$

where θ_0 and k_θ are the equilibrium bending angle and the force constant, respectively. These force field parameters are chosen from the transferable potential for phase equilibria-united atom (TraPPE-UA)[45, 75] and listed in Table 2.1 for different bending angle types in different molecules. The temperature of each simulation is T = 300 K.

2.2. Density-guided method

Here we introduce the density-guided method that attempts to use the exact probability density function so that each generated geometry can be accepted. In actual practice, due to the complexity of this probability density function, a numerical representation of this distribution function would be required. This numerical table can be generated either a priori from the distribution function or on-the-fly in a self-adapting manner. This method has been tested on propane, 2-methylpropane, and 2,2-dimethylpropane, that are good representatives of both linear and branched molecules. It has been shown from these test cases that reasonable approximations can be made (especially for the highly branched molecules) to drastically reduce the dimensionality and correspondingly the amount of the tabulated data that is needed to be stored, while the dependencies between the various geometrical variables can be still well considered so that a great acceptance rate can be achieved.

Table 2.1. Bending angle force field parameters

Molecule	Bending angle	θ_0 (degree)	k_θ/k_B (K)
Propane	CH_3-CH_2-CH_3	114	62500
2-Methylpropane	CH_3-CH-CH_3	112	62500
2,2-dimethylpropane	CH_3-C-CH_3	109.47	62500
Acetone	CH_3-C-CH_3	117.2	62500
Acetone	CH_3-C=O	121.4	62500

Described below are the details of this method. We only focus on the generation of bending angles for molecules in a gas-phase where the expected distribution of these angles can be numerically obtained via integration but extension of this method to other geometrical variables (including bond length and torsional angles) in other environment (with external interactions that can be tabulated in advance) is straightforward.

2.2.1. Regrowth of a one-branched (linear) molecule.

This case is shown in Fig. 2.1a where P and C are previous and current segments which have already been grown and segment G must be grown. The true normalized probability distribution of the bending angle can be described as follows:

$$f(\theta) = \frac{\sin\theta \exp(-\beta U_{\text{bend}}(\theta))}{\int_0^\pi \sin\alpha \exp(-\beta U_{\text{bend}}(\alpha))\mathrm{d}\alpha} \tag{2.2}$$

From this distribution, the cumulative probability distribution can be obtained

$$P(\theta) = \int_0^\theta f(\alpha)\mathrm{d}\alpha \tag{2.3}$$

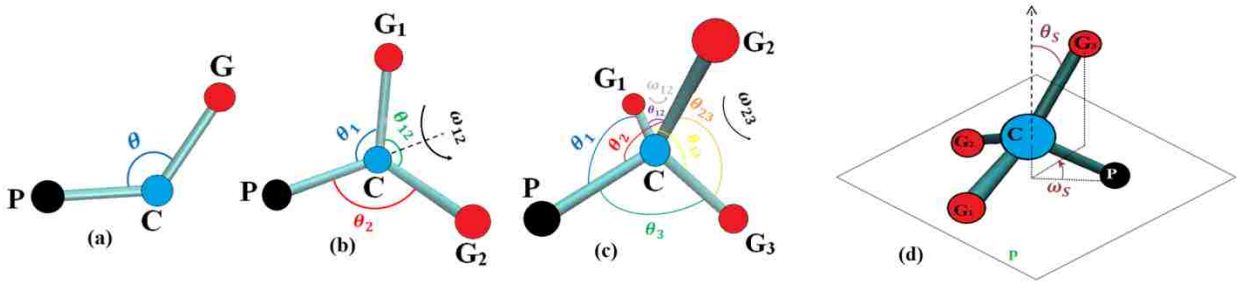

Fig. 2.1. Growth of (a) one-branched, (b) two-branched, and (c) three-branched molecules. (d) New variables, θ_S and ω_S, are introduced to define the orientation of the third branch in case (c).

Then a linear interpolation is done to find and tabulate values of θ at desired P's. These P values are typically evenly spaced with a certain ΔP interval, say 0.001. In order to generate a new conformation, a random number, R, is chosen uniformly between 0 and 1. Based on where this random number is located in the P range, say in the i-th interval, with $P(i) < R < P(i+1)$ and a $\Delta P(i) = P(i+1) - P(i)$, the corresponding θ interval can be determined, which is between $\theta(i)$ and $\theta(i+1)$ and an angle would be generated uniformly in that $\Delta\theta(i)$ interval. The detailed balance condition is

$$\sin\theta(o)\exp\left(-\beta U_{\text{bend}}(\theta(o))\right)\frac{\Delta P(n)}{\Delta\theta(n)}acc(o\to n) =$$

$$\sin\theta(n)\exp\left(-\beta U_{\text{bend}}(\theta(n))\right)\frac{\Delta P(o)}{\Delta\theta(o)}acc(n\to o) \qquad (2.4)$$

where the $\sin\theta$ term is the Jacobian factor for the bending angle.

Although the θ intervals are equal in probability, they are different in sizes, i.e., the sizes of these intervals are exactly inversely proportional to $f(\theta)$ when they are infinitely small. Thus, they are much smaller toward the most probable region. Since these intervals are equally likely to be picked for the angle generation, the sampling becomes denser toward more probable regions. If viewed this way, this method uses the same idea behind a number of other existing techniques, such as the aggregation-volume-bias Monte Carlo[76-78] that has led to recent successes

in simulating nucleation events.[46, 79-80] That is, first divide the phase space that is originally rather heterogeneous into various smaller regions so that within each region the probability densities are more or less uniform, and then sample the phase space region by region. In the work by Macedonia and Maginn,[44] a similar idea was proposed for the linear molecules.

2.2.2. Regrowth of a two-branched molecule.

In this case (see Fig. 2.1b), segments G_1 and G_2 must be grown. θ_1, θ_2, and θ_{12} are the three bending angles and ω_{12} is the dihedral angle between PCG_1 and PCG_2 planes. The geometry of this molecule can be specified by θ_1, θ_2, and ω_{12}, prescribed by the following probability density function:

$$f(\theta_1, \theta_2, \omega_{12}) \propto \sin\theta_1 \sin\theta_2 \exp\{-\beta[U_{\text{bend}}(\theta_1) + U_{\text{bend}}(\theta_2) + U_{\text{bend}}(\theta_{12})]\} \quad (2.5)$$

The following geometrical equation shows how θ_1, θ_2, θ_{12}, and ω_{12} are related:

$$\cos\theta_{12} = \cos\theta_1 \cos\theta_2 + \sin\theta_1 \sin\theta_2 \cos\omega_{12} \quad (2.6)$$

For 2-methyl-propane, all three bending angles are equivalent and should have the same average distribution that can be calculated as follows:

$$f(\theta_1) = \int_0^{\pi} \int_0^{2\pi} f(\theta_1, \theta_2, \omega_{12}) d\theta_2 d\omega_{12} \quad (2.7)$$

Similarly, the average distribution of ω_{12} is

$$f(\omega_{12}) = \int_0^{\pi} \int_0^{\pi} f(\theta_1, \theta_2, \omega_{12}) d\theta_1 d\theta_2 \quad (2.8)$$

It is expected that the distributions of θ_1, θ_2, and ω_{12} angles are interdependent. Thus if one uses the average distributions based on the above equations to generate these angles, the acceptance rate can be still rather poor. In order to consider the interdependencies of these angles while keeping the size of tabulated data reasonable, a more delicate procedure was employed. First, the dependencies of the distribution of θ_2 on θ_1 were analyzed by fixing θ_1 at various values (typically coinciding with the interval positions) and integrating the probability density function *f*

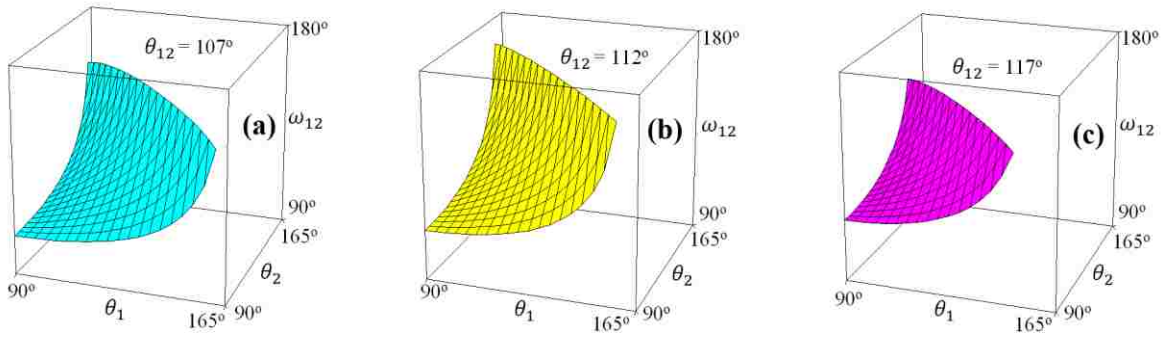

Fig. 2.2. ω_{12} (obtained from Eq. (2.6)) is drawn in the probable range of θ_1 and θ_2 at three different values of θ_{12} (all the angles are in units of degree).

over the ω_{12} space. Then, the dependencies of the distribution of ω_{12} on both θ_1 and θ_2 were considered. In principle, θ_1 and θ_2 must be treated as independent variables but this would entail a large computational task in terms of both computer time used for integration and data generated that needs to be stored. Instead we tried to identify a collective degree of freedom that can be representative of a group of θ_1 and θ_2 values when the ω_{12} distributions are similar so that the dimensionality of this problem can be reduced. Realizing that the distribution of ω_{12} is mostly determined by the Boltzmann weight governed by $U_{\text{bend}}(\theta_{12})$, ω_{12} is drawn as function of θ_1 and θ_2 at three different θ_{12} values when this Boltzmann weight reaches the maximum or half of the maximum (see Fig. 2.2). This can be used as a guide to estimate the peak position and the width of the ω_{12} distribution. As shown in Fig. 2.2, for all three θ_{12} surfaces the change in ω_{12} is very significant along the diagonal direction but negligible in the off-diagonal direction. Based on this observation, θ_{sum} (= $\theta_1 + \theta_2$) was introduced to be such a collective variable and the distributions of ω_{12} were analyzed at constant values of θ_{sum}.

As discussed in section 2.2.1, this sampling scheme involves a careful division of the phase space into smaller regions each of which consists of states with similar probability density to overcome the heterogeneity issue present in the original space. For the two-branched case, this

space is multi-dimensional and there are many ways to divide this space. To further reduce the amount of data that needs to be stored, in our implementation, this space division is done first statically along θ_1 (similarly to the one-branched case) using the intervals obtained from the average θ_1 distribution, then dynamically along θ_2 (depending on the θ_1 value picked), followed by another dynamic division along ω_{12} (depending on θ_{sum}). An interpolated scheme is used to obtain both θ_2 and ω_{12} intervals on-the-fly. Specifically, the θ_2 (or ω_{12}) interval positions obtained from the θ_2 (or ω_{12}) distributions at various θ_1 (or θ_{sum}) values, can be casted in a polynomial function. A third-order polynomial was found to represent a good balance between the accuracy desired and the amount of data/time needed to generate these intervals during the simulation runs. For example, the starting position of the *i*th interval for θ_2 is interpolated using the following formula:

$$\theta_2(i) = a_0^i + a_1^i\theta_1 + a_2^i\theta_1^2 + a_3^i\theta_1^3 \quad (2.9)$$

where a_0, a_1, a_2, and a_3 are the coefficients that would give the best fits to the sets of interval positions considered. In order to make sure that the deviations between the original interval positions/lengths and the interpolated ones are small (e.g., a threshold on the relative error of 0.1% and 1% was used for the interval positions and the interval lengths, respectively), the entire θ_1 (or θ_{sum}) space is divided into several regions and this interpolation is performed for these different ranges of θ_1 (or θ_{sum}), with each region yielding a different set of coefficients. For 2-methylpropane, 4 sets of coefficients were used for interpolating the θ_2 intervals and 11 sets were used for ω_{12} to achieve the desired accuracy.

2.2.3. Regrowth of a three-branched molecule.

In this case (see Fig. 2.1c), segments G_1, G_2, and G_3 must be grown. θ_1, θ_2, θ_3, θ_{12}, θ_{23}, and θ_{13} are the six bending angles. ω_{12} (or ω_{23}) is the dihedral angle between PCG_1 (or PCG_3) and PCG_2 planes. The probability density function is described as a function of θ_1, θ_2, θ_3, ω_{12}, and ω_{23} as:

$$f(\theta_1,\theta_2,\theta_3,\omega_{12},\omega_{23}) \propto \sin\theta_1 \sin\theta_2 \sin\theta_3 \exp\{-\beta[U_{\text{bend}}(\theta_1)+U_{\text{bend}}(\theta_2)+U_{\text{bend}}(\theta_3)+ U_{\text{bend}}(\theta_{12})+U_{\text{bend}}(\theta_{23})+U_{\text{bend}}(\theta_{13})]\} \quad (2.10)$$

For neo-pentane, all the bending angles are expected to have the same average distribution prescribed by the following formula:

$$f(\theta_1)=\int_0^{\pi}\int_0^{\pi}\int_0^{2\pi}\int_0^{2\pi} f(\theta_1,\theta_2,\theta_3,\omega_{12},\omega_{23})d\theta_2 d\theta_3 d\omega_{12} d\omega_{23} \quad (2.11)$$

Also all the dihedral angles would have the same average distribution as follows:

$$f(\omega_{12})=\int_0^{\pi}\int_0^{\pi}\int_0^{\pi}\int_0^{2\pi} f(\theta_1,\theta_2,\theta_3,\omega_{12},\omega_{23})d\theta_1 d\theta_2 d\theta_3 d\omega_{23} \quad (2.12)$$

A sparse-grid integration method was used to efficiently compute these high dimension integrals.[81]

The regrowth of the first two branches follows the same procedure as described for the two-branched case. For the last branch, instead of using θ_3 and ω_{23}, two new angles, θ_S and ω_S, are introduced to define its orientation as the dependencies of these two angles on the existing three-branched geometry can be more easily determined (see below). As shown in Fig. 2.1d, θ_S is defined as the polar angle between this branch and the normal of the plane made by the other three ending atoms (called PG_1G_2 Plane), whereas ω_S can be defined as the azimuth angle between the projection of the last branch and the projection of any of the other three existing branches onto this plane. Using this new set of angles, the probability density can be described as follows:

$$f(\theta_1, \theta_2, \omega_{12}, \theta_S, \omega_S) \propto \sin\theta_1 \sin\theta_2 \sin\theta_S \exp\{-\beta[U_{\text{bend}}(\theta_1) + U_{\text{bend}}(\theta_2) + U_{\text{bend}}(\theta_3) + U_{\text{bend}}(\theta_{12}) + U_{\text{bend}}(\theta_{13}) + U_{\text{bend}}(\theta_{23})]\} \quad (2.13)$$

Since the last branch determines θ_3, θ_{13}, and θ_{23}, as expected from Eq. (2.13) the most probable orientation of this vector would be decided by when all these three angles are optimized, i.e., to be close to the equilibrium tetrahedral angle. This can be achieved by placing the last branch around the normal vector of the PG_1G_2 Plane so that it is about equally far away from the existing three branches, like a perfect tetrahedral geometry. Thus it is natural to describe the orientation of the last branch relative to this normal vector using the set of θ_S and ω_S coordinates. In addition, the interdependencies between these two variables and those that define the existing 3-branched geometry can be more conveniently considered. For ω_S, the distribution was found rather uniform irrespective the molecular geometry, eliminating the need to include this coordinate for special treatment as it can be simply generated by the conventional uniform sampling scheme. To examine how the θ_S distribution is dependent on the molecular geometry, a collective coordinate is introduced, called the solid angle Ω (defined by the tetrahedron shaped by those three existing branches),[82] with

$$\cos\frac{\Omega}{2} = \frac{1+\cos\theta_1+\cos\theta_2+\cos\theta_{12}}{4\cos\frac{\theta_1}{2}\cos\frac{\theta_2}{2}\cos\frac{\theta_{12}}{2}} \quad (2.14)$$

This variable is a good measure of how closely (or sparsely) distributed the three existing branches are. For example, geometries with large Ω values correspond to a scenario when these branches are far from each other, which, in turn, would limit greatly the amount of space accessible for the last branch. That is, the θ_S distribution would be narrower and shift to smaller values. Thus, the θ_S distribution was treated as dependent upon only one coordinate Ω, instead of originally three coordinates, θ_1, θ_2, and ω_{12}. This greatly reduces the dimensionality of this

problem and correspondingly the computational expense. The division of the space along the θ_S coordinate follows the same procedure as developed for θ_2 and ω_{12}, which includes an interpolation of those θ_S intervals, pre-calculated at different Ω values, into a third-order polynomial function of Ω so that these intervals can be generated on-the-fly later during the production run for any molecular geometry with any Ω value. It was found that 4 sets of coefficients are sufficient for interpolating the θ_S intervals with the desired accuracy.

2.3. Results of density-guided method

Described in the following are the results obtained for the three different cases included in this study, i.e., propane, 2-methylpropane, and 2,2-dimethylpropane. For each case, we show that the new method proposed in Section 2.2 samples the correct probability density distributions for the various geometrical parameters specific to that particular molecule by comparing to the solutions obtained from the numerical integration over the analytical formula presented in Section 2.2. The results generated from the CD-CBMC method are included in this comparison as well. For each case, both the acceptance rate and the computer time are compared between the new method and CD-CBMC. Finally, the advantages of this new method over other methods such as Boltzmann rejection and arbitrary trial distribution CBMC are discussed.

2.3.1. Regrowth of a one-branched (linear) molecule.

For this molecule, the bending angle θ is the only variable required to define its geometry and the normalized distribution of this angle is prescribed by Eq. (2.2). Simulation runs using both the new method and the CD-CBMC method were carried out on a single propane molecule. Each run consists of 10^8 conformational moves to obtain the distribution of θ values. In Fig. 2.3,

the distributions produced from the simulation runs using both methods were compared to that predicted by Eq. (2.2). As shown in this figure, both methods sample the correct θ distribution.

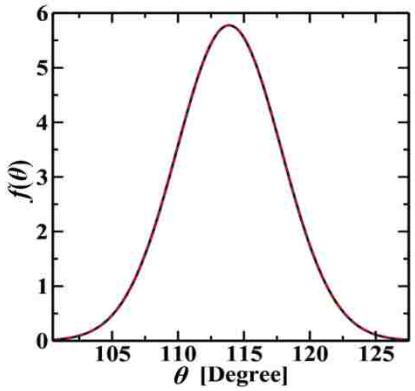

Fig. 2.3. Bending angle distribution of propane obtained from numerical integration according to Eq. (2.2) (solid line) and from the simulation using CD-CBMC (blue dotted line) or the density-guided method (red dashed line).

However, the amount of time needed by these two methods can differ significantly from each other. Table 2.2 contains the time required by CD-CBMC using different numbers of trials and the density-guided method using different ΔP values, as well as the acceptance rate obtained for each case. In general, the use of smaller ΔP improves the acceptance rate for the new method at a cost of only a minor increase on the CPU time. On the other hand, the acceptance rate of CD-CBMC improves by about 10 times from a use of a single trial to a use of 10^4 trials but this improvement can be barely balanced by the increase on the computational requirement since it is directly proportional to the number of trials (note that here only a single molecule is considered, for large systems this increase on the computational expense becomes slightly less noticeable due to the significant computational overhead on the nonbonded interactions). In addition, even with a rather coarse division of the space at $\Delta P = 0.1$, this method yields an acceptance rate of 73.22%. Clearly this method has a significantly better performance than CD-CBMC in terms of both acceptance rate and computer time.

This method can be further optimized with a flexible choice of ΔP at different regions of θ. In particular, toward the two end (either when θ or P is small or large), $f(\theta)$ changes very rapidly and smaller ΔP intervals would be desirable to keep each region being uniform in terms

Table 2.2. Acceptance rate and time of simulation for 10^8 CBMC steps for propane.

Method	Number of trials	%Acceptance	Time(s)
CD-CBMC	1	10.06	13
	10	51.11	129
	100	84.11	1203
	1000	94.97	12431
	10000	98.41	125130
Density-guided (one trial)	ΔP	%Acceptance	Time(s)
	0.1	73.22	11
	0.01	96.20	11
	0.001	99.52	11
	0.0001	99.94	11

of the probability density, which is essential for achieving a high acceptance rate. Indeed, for the case with $\Delta P = 0.001$, by dividing the two ending regions with $P \leq 0.001$ or $P \geq 0.999$ further in a logarithmic way until the last segment has a length of 10^{-8} in terms of P, the acceptance rate was found to increase to 99.68%.

2.3.2. Regrowth of a two-branched molecule.

For this molecule, the geometry is specified by three variables, two bending angles (θ_1 and θ_2, with the same distribution as prescribed by Eq. (2.7)) and one dihedral angle (ω_{12}, with a distribution defined by Eq. (2.8)). In Fig. 2.4, the distributions produced from the simulation runs using both methods with 10^8 Monte Carlo moves were compared to those predicted by Eqs. (2.7) and (2.8). As shown in this figure, both methods sample the correct distributions for both bending and dihedral angles. Whereas in the CD-CBMC simulation run 10^3 trials were used for each angle with a yielded acceptance rate of 89.05%, in the run with the density-guided method one single trial was used for each angle (about 2-3 orders of magnitude more efficient than CD-CBMC) with an even better acceptance rate of 98.26%. When using the average distributions prescribed by Eqs. (2.7) and (2.8) to generate θ_1, θ_2, and ω_{12} independently without taking into

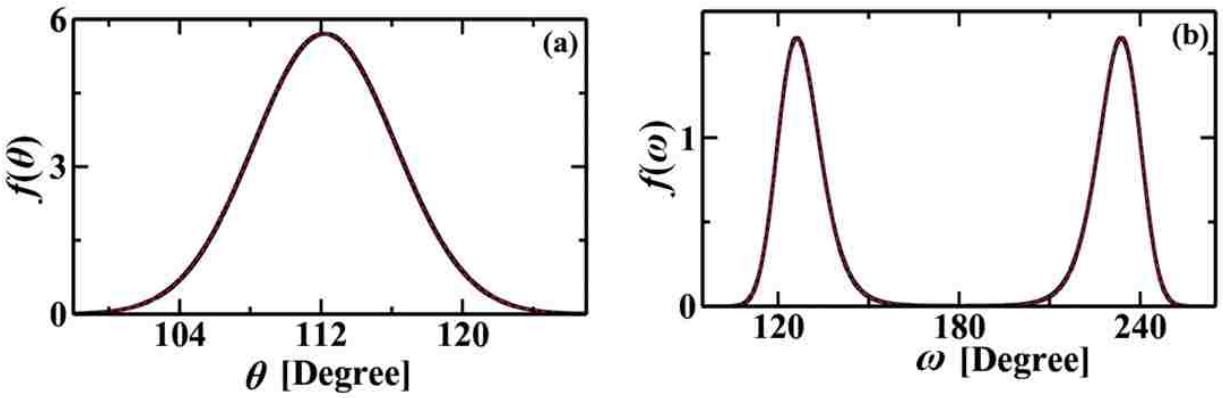

Fig. 2.4. Distributions of (a) the bending angles and (b) dihedral angle for 2-methylpropane obtained from numerical integration (solid lines) (i.e., according to Eqs. (2.7) and (2.8), respectively) and from the simulation using CD-CBMC (blue dotted line) and the density-guided method (red dashed line).

their interdependencies, the acceptance rate lowers significantly to 60.48%. This indicates that these variables are coupled closely to each other and it is important to analyze their relationship.

As shown in Fig. 2.5a, initially increasing θ_1 leads to a shift of the θ_2 distribution to larger values until θ_1 reaches 130° (which is above the equilibrium angle of 112°), further increase in θ_1 leads to an opposite shift and a narrower θ_2 distribution. The existence of this turning point can be explained from the need to have all three bending angles (including θ_{12}, the bending angle between the two branches, see Fig. 2.1b) close to the equilibrium value. It should be noted that from Eq. (2.7), the probability becomes already quite low at that turning point (i.e., the integrated probability to have θ_1 above 130° is only 3×10^{-4}). Thus, for the most important part of the phase space, θ_2 is only weakly dependent on θ_1. If one ignores this part of dependencies and uses the average distribution prescribed by Eq. (2.7) to generate θ_1 and θ_2, an acceptance rate of 97.77% is obtained.

In contrast, ω_{12} is much more strongly coupled with θ_1 and θ_2. As shown in Fig. 2.5b, the ω_{12} distribution is strongly dependent on the bending angles. In analyzing this part of

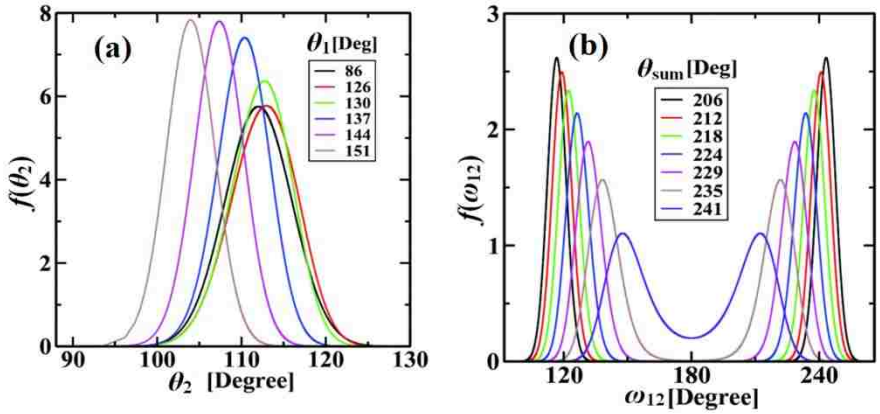

Fig. 2.5. (a) Distributions of the bending angle θ_2 obtained at different θ_1 values using numerical integration for 2-methylpropane. (b) Distributions of the dihedral angle ω_{12} obtained at different θ_{sum} (= $\theta_1 + \theta_2$) using numerical integration for 2-methylpropane.

dependencies, a collective coordinate, $\theta_{sum} = (\theta_1 + \theta_2)$, is introduced for the reasons presented in the section 2.2.2 (i.e., mainly to lower the dimensionality/complexity of this problem) and the

ω_{12} distribution is plotted as function of θ_{sum}. It is clear that with increasing θ_{sum}, the ω_{12} distribution changes significantly, becoming broader and more closely centered toward the value of π (e.g., from a bimodal distribution at low θ_{sum} values to a single-peaked distribution when θ_{sum} is above 250°). Again the change on the ω_{12} distribution can be explained by the need to have all three bending angles including θ_{12} close to the equilibrium value, θ_0. For example, at large values of θ_{sum} (or large values of θ_1 and θ_2 so that both are close to θ_0 to minimize the bending energies due to these two angles), in order to keep θ_{12} close to θ_0, the two branches must be as far as possible with ω_{12} approaching π, as expected from Eq. (2.6). Also expected from this equation, the ω_{12} distribution is symmetrical at π (i.e., the same θ_{12} is obtained at ω_{12} or at $2\pi-\omega_{12}$), which is another important feature of Fig. 2.5b. Thus only half of this distribution (or ω_{12} intervals), either for the range [0, π] or [π, 2π] , need to be included. This leads to further saving of the amount of data required to be stored by this method. Listed in Table 2.3 are the

Table 2.3. Acceptance rate and time of simulation for 10^8 CBMC steps for 2-methylpropane.

Method	Number of trials	%Acceptance	Time(s)
CD-CBMC	1	0.18	59
	10	21.34	448
	100	71.93	4214
	1000	89.05	41240
	10000	92.99	406800
Density-guided	1	98.26	95

time and acceptance rate of CD-CBMC using different numbers of trials and the density-guided method using a single trial. For this case, it is clear that the density-guided method easily outperforms the CD-CBMC on both CPU time required and the acceptance rate obtained.

2.3.3. Regrowth of a three-branched molecule.

For this molecule, the distribution for the bending angles (θ_1, θ_2, θ_3, θ_{12}, θ_{13}, and θ_{23}) can be all described by Eq. (2.11), and the distribution for the dihedral angles (ω_{12}, ω_{23}, and ω_{13}) can be described by Eq. (2.12). As shown in Fig. 2.6, these distributions can be sampled correctly by both CD-CBMC and the density-guided method. Whereas in the CD-CBMC simulation run a use of 10^3 trials for each angle yielded an acceptance rate of only 62.36%, in the run with the density-guided method where only one single trial was used for each angle a nearly perfect acceptance rate of 95.98% can be still achieved. Again for the density-guided method, it is important to consider the interdependencies between the various variables that govern the geometry. For example, when using the averaging distributions to generate the bending and dihedral angles (in this case, θ_1, θ_2, θ_3, ω_{12}, and ω_{23}), an acceptance rate of 51.62% was obtained.

Although for the first two branches, the same procedure developed for the two-branched molecule was applied to this three-branched molecule (namely, θ_1 was picked from the average bending angle distribution, then θ_2 was picked from a θ_1-dependent distribution, whereas ω_{12} was picked from a distribution depending on the value of θ_{sum}), the interdependencies between θ_1, θ_2, and ω_{12} exhibit significant differences between these two cases. For example, the increase in θ_1 only pushes the θ_2 distribution to smaller value (see Fig. 2.7a). The peak positions for the ω_{12} distribution shift closer to π at larger θ_{sum} values but at a much slower pace and within the part of

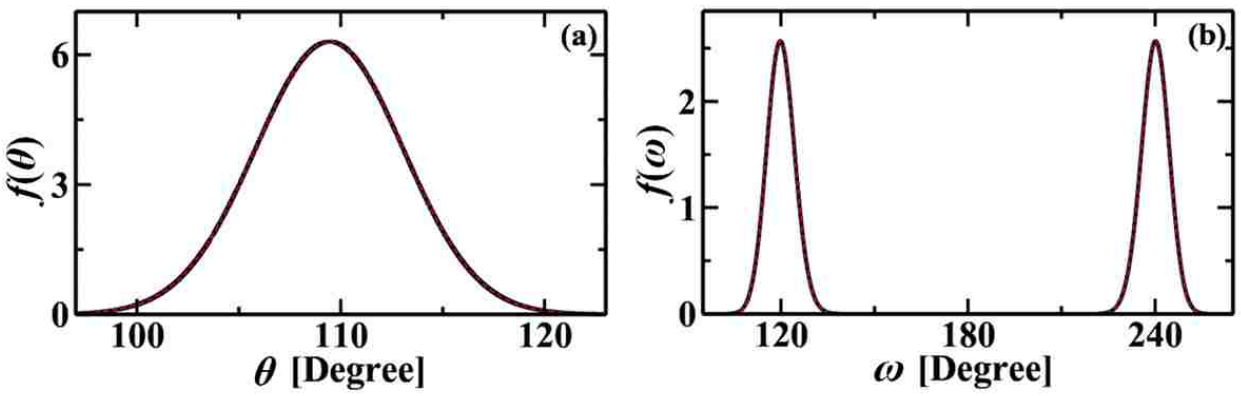

Fig. 2.6. Distributions of (a) bending angles and (b) dihedral angles for 2,2-dimethylpropane obtained from numerical integration (solid lines) (i.e., according to Eqs. (2.11) and (2.12), respectively) and from the simulation using CD-CBMC (blue dotted line) and the density-guided method (red dashed line).

space accessible by this molecule they never reach π to form a single-peaked distribution (see Fig. 2.7b). These differences are caused by the presence of the third branch (and addition of three bending angles due to this branch), which limits both bending and dihedral angles to much smaller range.

Instead of using θ_3 and ω_{23}, the orientation of the third branch is specified by θ_S and ω_S since the dependencies of these two variables on the current geometry can be conveniently casted in terms of the solid angle, Ω (defined by the tetrahedron shaped by those three existing branches). In addition, the ω_S distribution for the range of Ω values accessible by this molecule was found to be nearly flat. Thus generation of ω_S follows the conventional uniform-sampling scheme (i.e., generated randomly/uniformly within the range of 0 and 2π). The θ_S distribution was found to depend slightly on Ω. As shown in Fig. 2.7c, for larger Ω (when the three existing branches are far apart, which would leave less space for the last branch), the θ_S distribution

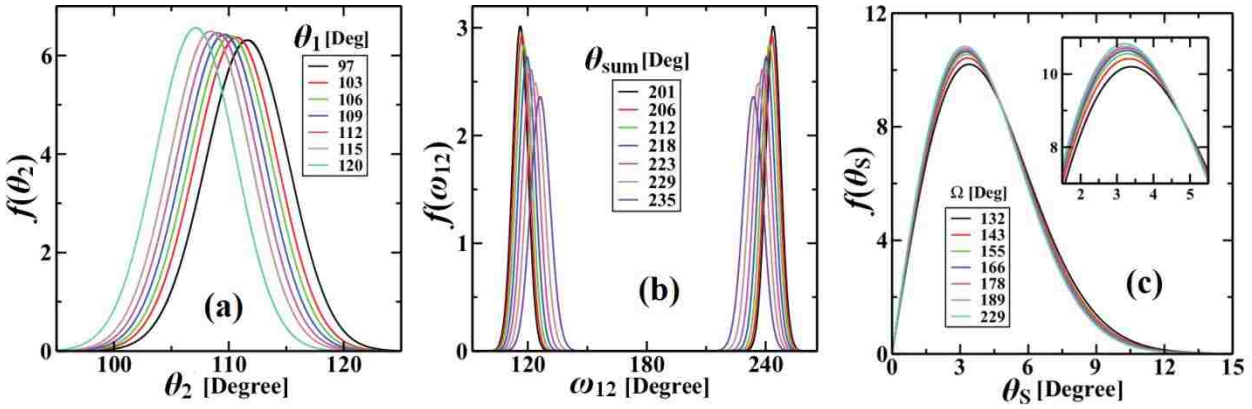

Fig. 2.7. (a) Distributions of the bending angle θ_2 obtained at different θ_1 values using numerical integration for 2,2-dimethylpropane. (b) Distributions of the dihedral angle ω_{12} obtained at different θ_{sum} (= θ_1 + θ_2) using numerical integration for 2,2-dimethylpropane. (c) Distributions of the polar angle θ_S obtained at different solid angle Ω using numerical integration for 2,2-dimethylpropane.

becomes narrower and shifts closer to zero as expected. However, compared to θ_2 and ω_{12}, the change of the θ_S distribution is significantly smaller. If one neglects this part of dependency entirely by using the θ_S distribution averaged over all Ω values for the new method, an acceptance rate of 95.52% is obtained, which is still far better than CD-CBMC.

Listed in Table 2.4 are the time and acceptance rate of CD-CBMC using different numbers of trials and the new method using a single trial. For this case, the new method performs significantly better than CD-CBMC in terms of both CPU time and the acceptance rate.

2.3.4. Comparison with the other methods

It is necessary to discuss how the density-guided method compares to the other methods, such as CBMC using the Boltzmann rejection scheme,[37] CD-CBMC,[45] and CBMC using the arbitrary trial distribution,[54] for the bending angle sampling for both linear and branched molecules.

For linear molecules, all the methods above, can sample the bending angle correctly, but the acceptance rate is high for the Boltzmann rejection scheme and CD-CBMC only when many trials are generated. This is because trials are not generated according to the true distribution (Eq. (2.2)). For instance, in CD-CBMC trials are generated from the sine distribution. Thus, only few generated angles have significant chance to be accepted, which is why a lot of trials are required.

Table 2.4. Acceptance rate and time of simulation for 10^8 CBMC steps for 2,2-dimethylpropane

Method	Number of trials	%Acceptance	Time(s)
CD-CBMC	1	6×10^{-4}	112
	10	3.16	841
	100	44.10	8104
	1000	62.36	80110
	10000	65.61	775000
Density-guided	1	95.98	138

For branched molecules, the Boltzmann rejection scheme cannot sample bending angles correctly when branches are regrown sequentially without considering the coupling between these branches. Since in CD-CBMC branches are regrown simultaneously, it is able to sample bending angles correctly, but since the selection of each bending angle is based on the Boltzmann weight and phase space ($\sin\theta$) governed by this angle alone and this selection is carried out sequentially without considering the interdependencies between these angles, it cannot reach high acceptance rate even at high trial numbers (see, e.g., Table 2.4). In arbitrary trial distribution CBMC, a "Coupled to Pre-Nonbond (CPN) CD-CBMC" formula is used to

reach high acceptance rates but this still requires a high number of trials (e.g., 100 to 1000 in order to obtain an acceptance rate above 90%).

In contrast, the density-guided method generates all the required geometrical variables using tabulated distributions obtained originally from the true distributions by taking into account the interdependencies between these variables, so just one trial is required to achieve high acceptance rates for both linear and branched molecules.

2.4. Jacobian-Gaussian method

The Jacobian-Gaussian method is a robust and general approach for generating angle trials for both linear and branched molecules. It is also very straightforward to be implemented for systems using harmonic bending potential (Eq. (2.1)) which is a popular potential in many force fields including consistent force field (CFF),[83] TraPPE,[45, 75, 84-90] Amber,[91-92] OPLS-AA,[93] and CHARMM.[94-95] In addition, it can be conveniently extended to nonharmonic bending potentials. This approach does not require curve fitting or memory storage needed for preparing conformation libraries or tables.

As it is explained in section 1.4, the probability density function for the intramolecular interactions is proportional to $J\exp(-\beta U^{\text{intra}})$. Unlike CD-CBMC, in which trial generation is based on purely the Jacobian, the Jacobian-Gaussian method generates K_{Trial} trials according to $\exp(-\beta U^{\text{intra}})$ and one of them (say, the ith-trial) is selected according to its Jacobian factor, J_i, as follows:

$$P_{\text{select}}(i) = \frac{J_i}{W_J} \tag{2.15}$$

with

$$W_J = \sum_{i=1}^{K_{\text{Trial}}} J_i \tag{2.16}$$

Finally, the new conformation is accepted according to the ratio of the Rosenbluth weights of new and old conformations, i.e. min[1, $W_J(n)/W_J(o)$], where *o* and *n* represent old and new conformations respectively. This method is tested on propane, 2-methylpropane, 2,2-dimethylpropane, and acetone.

2.4.1. Gaussian random number generator

The probability density function of a Gaussian distribution is described by

$$f(x|\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \tag{2.17}$$

where μ and σ are the mean and the standard deviation, respectively. If μ = 0 and σ = 1, it is called standard Gaussian distribution. There are analytical methods for generating random numbers with Gaussian distribution. In the Box-Muller[96] method, two independent random numbers, R_1 and R_2, are generated uniformly on (0, 1), then two independent random numbers with standard Gaussian distribution are obtained by

$$\begin{cases} Z_1 = \sqrt{-2\ln R_1}\cos(2\pi R_2) \\ Z_2 = \sqrt{-2\ln R_1}\sin(2\pi R_2) \end{cases} \tag{2.18}$$

Then, two independent Gaussian random numbers with μ and σ parameters are attained by

$$\begin{cases} X_1 = Z_1\sigma + \mu \\ X_2 = Z_2\sigma + \mu \end{cases} \tag{2.19}$$

2.4.2. Regrowth of a one-branched (linear) molecule.

In this case (Fig. 2.1a), according to the probability density function (Eq. (2.2)), the Jacobian factor is

$$J(\theta) = \sin\theta \tag{2.20}$$

When a harmonic potential (i.e., Eq. (2.1)) is used for $U_{\text{bend}}(\theta)$, the exponential term in Eq. (2.2) corresponds to a Gaussian distribution with $\mu = \theta_0$ and $\sigma = (\beta k_\theta)^{-0.5}$. Since Gaussian random

numbers are generated on (-∞, +∞), in order to generate a valid trial, bending angle θ is generated according to its corresponding Gaussian distribution until it satisfies the bending angle condition such that $\theta \in (0, \pi)$. For this case, the probability of generating an angle outside this interval is less than 10^{-20}. Thus, each generated angle is very likely to be a valid trial to be used in the next step, i.e., trial selection using Eqs. (2.15) and (2.16).

2.4.3. Regrowth of a two-branched molecule.

In this case (see Fig. 2.1b), when θ_1, θ_2, and ω_{12} are used as the growing variables (Eq. (2.5)), the Jacobian factor is

$$J(\theta_1, \theta_2, \omega_{12}) = \sin\theta_1 \sin\theta_2 \tag{2.21}$$

However, the fact that the energetic term contributes most to the probability distribution and that this term can be conveniently expressed as a function of the three bending angles would naturally lead to the idea of using θ_1, θ_2, and θ_{12} (instead of ω_{12}) as the growing variables. A new Jacobian factor, $J(\theta_1, \theta_2, \theta_{12})$, must be used in conjunction with a differential volume element expressed by this new set of coordinates. Eq. (2.6) can be written as:

$$\cos\omega_{12} = \frac{\cos\theta_{12} - \cos\theta_1 \cos\theta_2}{\sin\theta_1 \sin\theta_2} \tag{2.22}$$

Since $-1 \leq \cos(\omega_{12}) \leq 1$, it can be inferred from Eq. (2.22) that

$$\cos(\theta_1 + \theta_2) \leq \cos\theta_{12} \leq \cos(\theta_1 - \theta_2) \tag{2.23}$$

or

$$|\theta_1 - \theta_2| \leq \theta_{12} \leq \min\{(\theta_1 + \theta_2), 2\pi - (\theta_1 + \theta_2)\} \tag{2.24}$$

The absolute value and min in Eq. (2.24) guarantee that $\theta_{12} \in [0, \pi]$.

This new Jacobian factor, $J(\theta_1, \theta_2, \theta_{12})$, is determined by

$$J(\theta_1, \theta_2, \theta_{12}) = \sin\theta_1 \sin\theta_2 \left|\frac{\partial\omega_{12}}{\partial\theta_{12}}\right| \tag{2.25}$$

From Eq. (2.22), we have

$$\frac{\partial \omega_{12}}{\partial \theta_{12}} = \frac{\sin\theta_{12}}{\sin\theta_1 \sin\theta_2 \sin\omega_{12}} \tag{2.26}$$

So, Eq. (2.25) becomes

$$J(\theta_1, \theta_2, \theta_{12}) = \left|\frac{\sin\theta_{12}}{\sin\omega_{12}}\right| \tag{2.27}$$

Each combination of bending angles θ_1, θ_2, and θ_{12} is valid if geometrical constraints (Eq. (2.24)) are satisfied. According to Eq. (2.22), for each valid set of (θ_1, θ_2, θ_{12}), there are two possible answers for ω_{12} (i.e., ω_{12} ϵ (0, π) or (π, 2π)). In order to verify that θ_1, θ_2, and θ_{12} can span the whole space that is spanned by θ_1, θ_2, and ω_{12}, it should be proved that

$$\int_0^{\pi}\int_0^{\pi}\int_0^{2\pi} \sin\theta_1 \sin\theta_2 \, d\omega_{12} d\theta_2 d\theta_1 = 2\int_0^{\pi}\int_0^{\pi}\int_{\theta_{12}^{\min}}^{\theta_{12}^{\max}} \left|\frac{\sin\theta_{12}}{\sin\omega_{12}}\right| d\theta_{12} d\theta_2 d\theta_1 \tag{2.28}$$

where the factor 2 on the right side appears because of the two possible answers for ω_{12}. $\theta_{12}^{\min}$ and $\theta_{12}^{\max}$ are the lower and upper limits of θ_{12} (Eq. (2.24)). The three dimensional integral on the left side of Eq. (2.28) is equal to 8π. The $\sin\omega_{12}$ term on the right side of the integral can be substituted by $(1-\cos^2\omega_{12})^{0.5}$ and $\cos\omega_{12}$ can be written in terms of θ_1, θ_2, and θ_{12} using Eq. (2.22) and we have

$$2\int_0^{\pi}\int_0^{\pi}\int_{\theta_{12}^{\min}}^{\theta_{12}^{\max}} \frac{\sin\theta_1 \sin\theta_2 \sin\theta_{12}}{\sqrt{\sin^2\theta_1 \sin^2\theta_2 - (\cos\theta_{12} - \cos\theta_1 \cos\theta_2)^2}} d\theta_{12} d\theta_2 d\theta_1 =$$

$$2\int_0^{\pi}\int_0^{\pi}(-\sin\theta_1 \sin\theta_2)\left[\sin^{-1}\left(\frac{\cos\theta_{12} - \cos\theta_1 \cos\theta_2}{\sin\theta_1 \sin\theta_2}\right)\right]_{\theta_{12}^{\min}}^{\theta_{12}^{\max}} d\theta_2 d\theta_1 =$$

$$2\int_0^{\pi}\int_0^{\pi}(-\sin\theta_1 \sin\theta_2)\left[\sin^{-1}\left(\frac{\cos(\theta_1+\theta_2) - \cos\theta_1 \cos\theta_2}{\sin\theta_1 \sin\theta_2}\right) -\right.$$

$$\left.\sin^{-1}\left(\frac{\cos(\theta_1-\theta_2) - \cos\theta_1 \cos\theta_2}{\sin\theta_1 \sin\theta_2}\right)\right] d\theta_2 d\theta_1 =$$

$$2\int_0^{\pi}\int_0^{\pi}(-\sin\theta_1 \sin\theta_2)[\sin^{-1}(-1) - \sin^{-1}(1)] d\theta_2 d\theta_1 =$$

$$2\pi\int_0^{\pi}\int_0^{\pi} \sin\theta_1 \sin\theta_2 \, d\theta_2 d\theta_1 = 8\pi \tag{2.29}$$

In order to generate one valid trial, θ_1, θ_2, and θ_{12} are generated independently and simultaneously according to their corresponding Gaussian distributions until the following conditions are satisfied

1. $\theta_1 \in (0, \pi)$

2. $\theta_2 \in (0, \pi)$

3. $\theta_{12} \in [|\theta_1 - \theta_2|, \min\{(\theta_1 + \theta_2), 2\pi - (\theta_1 + \theta_2)\}]$

For a valid set of (θ_1, θ_2, θ_{12}), one of the two possible answers for ω_{12} (Eq. (12)) is chosen randomly.

2.4.4. Regrowth of a three-branched molecule.

In this case (see Fig. 2.1c), when θ_1, θ_2, θ_3, ω_{12}, and ω_{23} are used as the growing variables (Eq. (2.10)), the Jacobian factor is

$$J(\theta_1, \theta_2, \theta_3, \omega_{12}, \omega_{23}) = \sin\theta_1 \sin\theta_2 \sin\theta_3 \tag{2.30}$$

Again for the same reason as mentioned for the two-branched case, it is more convenient to use completely bending angles as the growing variables, i.e., replacing ω_{12} and ω_{23} by θ_{12} and θ_{23}. A new Jacobian factor needs to be determined for this set of growing variables. In addition to Eq. (2.22), we have

$$\cos\omega_{23} = \frac{\cos\theta_{23} - \cos\theta_2 \cos\theta_3}{\sin\theta_2 \sin\theta_3} \tag{2.31}$$

so

$$\frac{\partial\omega_{23}}{\partial\theta_{23}} = \frac{\sin\theta_{23}}{\sin\theta_2 \sin\theta_3 \sin\omega_{23}} \tag{2.32}$$

Using Eqs. (2.26) and (2.32), growing variables are transformed into (θ_1, θ_2, θ_3, θ_{12}, θ_{23}) and the Jacobian is

$$J(\theta_1, \theta_2, \theta_3, \theta_{12}, \theta_{23}) = \left|\frac{\sin\theta_{12} \sin\theta_{23}}{\sin\theta_2 \sin\omega_{12} \sin\omega_{23}}\right| \tag{2.33}$$

In order to generate one valid trial, θ_1, θ_2, θ_3, θ_{12} and θ_{23} are generated independently and simultaneously according to their corresponding Gaussian distributions until the following conditions are satisfied

1. $\theta_1 \in (0, \pi)$

2. $\theta_2 \in (0, \pi)$

3. $\theta_{12} \in [|\theta_1 - \theta_2|, \min\{(\theta_1 + \theta_2), 2\pi - (\theta_1 + \theta_2)\}]$

* One of the two answers for ω_{12} (Eq. (2.22)) is chosen randomly.

4. $\theta_3 \in (0, \pi)$

5. $\theta_{23} \in [|\theta_2 - \theta_3|, \min\{(\theta_2 + \theta_3), 2\pi - (\theta_2 + \theta_3)\}]$

* One of the two answers for ω_{23} (Eq. (2.31)) is chosen randomly.

6. $e^{-\beta U_{bend}(\theta_{13})} \geq \text{random}(0, 1)$

where random(0, 1) is a random number which is generated uniformly on (0, 1). Condition (6) ensures that the generated trial is also taking into account the bending potential due to θ_{13}.

Using the above procedure, it is straightforward to extend this approach to cases containing even more branches.

2.5. Results of Jacobian-Gaussian method

2.5.1. Methodology verification and efficiency

In this section, it will be demonstrated that the Jacobian-Gaussian (JG) method is both accurate and fast. In order to show that JG reproduces correct results, the angle distributions obtained by 10^9 MC moves using this method with one trial (K_{Trial} = 1 in Eq. (2.16)) are compared with the expected distributions for linear (Eq. (2.2)) and branched (Eqs. (2.7)-(2.8) and (2.11)-(2.12)) molecules. Fig. 2.8 compares these two distributions of the bending angle for

propane. Because of the symmetry in 2-methylpropane and 2,2-dimethylpropane, there is only one type of bending (or dihedral) angles. Fig. 2.9 compares the simulated to the expected distributions for both bending and dihedral angles of 2-methylpropane and 2,2-dimethylpropane. Figs. 2.8 and 2.9 show that in these cases, JG reproduces the correct distributions.

In order to examine the speed and the efficiency of JG, both the time and the acceptance rate of 10^8 single trial MC moves of the density-guided (DG) method and this method are compared. As shown in Table 2.5, in the case of propane, DG is faster because there is only one variable and one table is scanned to generate one trial. However, for 2-methylpropane, several tables must be used for trial generation which makes DG slower. In addition, JG is very fast because the trial generation loop in section 2.4.3 is very likely to produce one acceptable trial (i.e., within

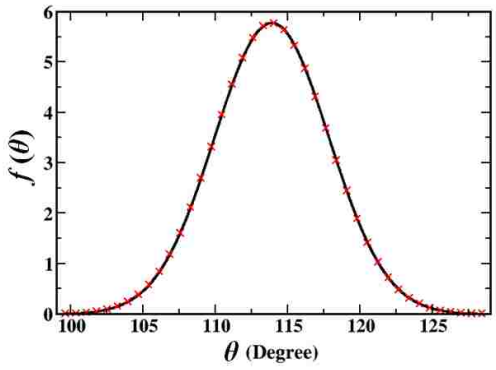

Fig. 2.8. Expected (black line) vs. simulated (red crosses) distributions obtained for the bending angle of propane.

geometrical constraints) only in one run. In 2,2-dimethylpropane, JG is slower because the trial generation loop (section 2.4.4) is often required to be implemented two or three times to satisfy all conditions, in particular, condition (6). These results demonstrate that both the speed and the efficiency of JG method are on the same order of the DG method at least for this set of molecules. In addition, it has several advantages over DG. First, there is no need for time or memory for table preparation or storage. Second, for systems containing branched molecules

with different bending potentials, DG needs different tables for different growth directions, but JG can easily be adapted to different growth directions. Third, while it is simple to extend JG to molecules with higher number of branches, growing more than three branches in DG requires development of new collective variables and enormous amount of time for multidimensional integrations.

It should be noted that DG attempts to generate each trial according to its expected probability (determined by both the Boltzmann and the Jacobian factor) so that every trial generated is always acceptable. However, JG only takes into account the Boltzmann factor when generating each trial and the Jacobian factor is only considered when accepting/rejecting the entire move. This may potentially affect its overall acceptance rate, in particular, when not only the Boltzmann exponential term but also the Jacobian factor contributes significantly to the probability distribution, which happens in planar molecules. Indeed, using JG with θ_1, θ_2, and θ_{12} as growing variables on a planar molecule, acetone, yields substantially lower acceptance rates, i.e., 30% lower than the linear and branched alkane cases examined above when using only one trial (see Table 2.6). This occurs because acetone is a planar molecule and dihedral angles have peaks close to $\omega_{12} = \pi$ (see Fig. 2.10b, d) where there is a singularity in the Jacobian factor (see Eq. (2.27)). This effect can be also observed in Table 2.5 where the acceptance rate for 2,2-dimethylpropane is 2% higher than 2-methylpropane. The dihedral angle distribution for 2,2-dimethylpropane (Fig. 2.9d) is tighter than 2-methylpropane (Fig. 2.9b) due to the presence of the third branch whose bending potentials prevent other branched to be located in one plane.

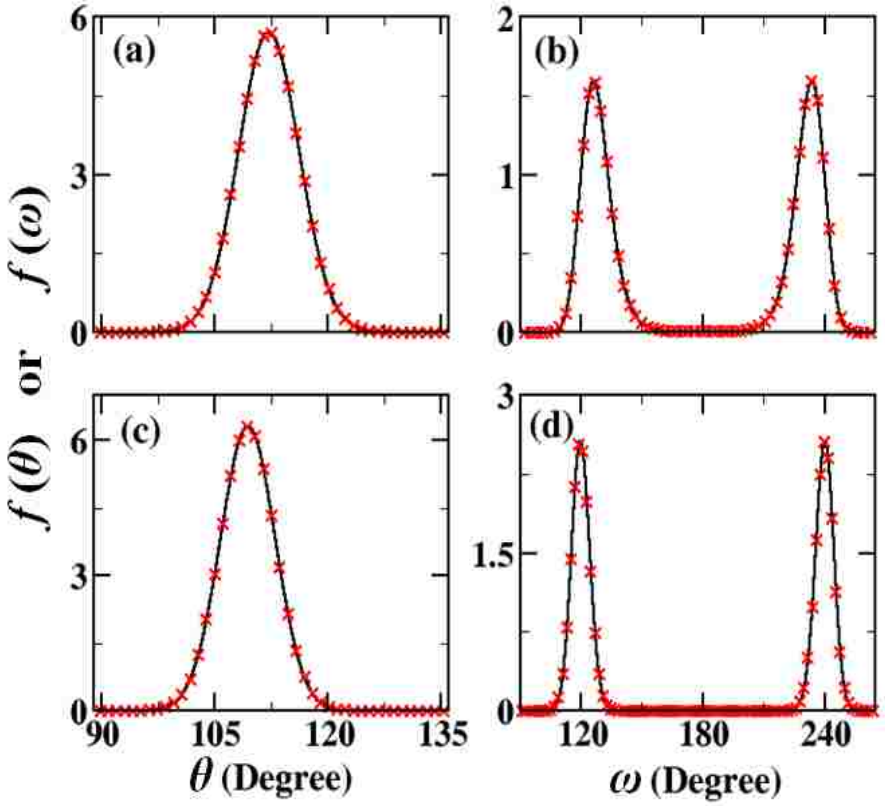

Fig. 2.9. Expected (solid lines) vs. simulated (red crosses) distributions obtained for (a) the bending angle of 2-methylpropane, (b) the dihedral angle of 2-methylpropane, (c) the bending angle of 2,2-dimethylpropane, and (d) the dihedral angle of 2,2-dimethylpropane.

This singularity issue can be avoided by choosing θ_1, θ_2, and ω_{12} as the growing variables for acetone because the Jacobian factor there is simply $\sin\theta_1 \sin\theta_2$ (see Eq. (2.21)). However, one would still need to explicitly take into account the Boltzmann factor due the θ_{12} angle to ensure a good overall acceptance rate. This can be done via an additional Boltzmann rejection step. Thus, the following procedure is developed: θ_1, θ_2, and ω_{12} are generated independently and simultaneously, where θ_1 and θ_2 must be bending angles (i.e. on (0, π)) that are sampled from their corresponding Gaussian distributions and ω_{12} is generated uniformly on (0, 2π), until $\exp[-\beta U_{\text{bend}}(\theta_{12})] \geq \text{random}(0, 1)$. In acetone, there are two types of bending (CH_3-C=O and CH_3-C-CH_3) and two types of dihedral (CH_3-C(=O)-CH_3 and CH_3-C(-CH_3)=O) angles. As shown in

Table 2.5. Time and acceptance rates of 10^8 single trial MC moves using the density-guided or the Jacobian-Gaussian method.

Molecule	Density-guided		Jacobian-Gaussian	
	Time (s)	Acceptance (%)	Time (s)	Acceptance (%)
Propane	11	99.52	16	98.26
2-Methylpropane	95	98.26	38	94.89
2,2-Dimethylpropane	138	95.98	235	96.88

Fig. 2.10, JG can reproduce the expected distributions for all angles with both sets of growing variables. Table 2.6 compares the speed and the efficiency of these two different JG procedures. As shown by this table, a single trial with (θ_1, θ_2, ω_{12}) is three to four times slower than a single trial with (θ_1, θ_2, θ_{12}) which is due to the required attempts for trial generation according to the Boltzmann distribution of θ_{12} via the additional Boltzmann rejection step (similar to the 2,2-dimethylpropane case discussed above). However, the acceptance rate is nearly perfect even with a single trial when using this new procedure. In contrast, with the old procedure even a use of 20 trials, the acceptance rate is still far from being perfect. Thus, it would be more preferable to use JG with (θ_1, θ_2, ω_{12}) than (θ_1, θ_2, θ_{12}) for branched planar molecules. It should be also noted that for all branched molecules, each dihedral distribution has two peaks that are symmetric around π. For non-planar molecules these two peaks are separated far from each other (see Fig. 2.9b, d). In contrast, for planar molecules these two peaks overlap each other (see Fig. 2.10b, d) that causes the overall dihedral distribution to be much broader. This also supports why generating ω_{12} from a uniform distribution works well for planar molecules.

Table 2.6. Time and acceptance rates of 10^8 MC moves for acetone with different number of trials using the Jacobian-Gaussian method with two sets of growing variables.

Growing Variables	Number of trials	Time (s)	Acceptance rate (%)
(θ_1, θ_2, θ_{12})	1	43	65.84
	2	104	71.66
	3	176	74.87
	5	291	78.61
	8	478	81.75
	10	604	83.13
	20	1222	86.88
(θ_1, θ_2, ω_{12})	1	158	97.67

2.5.2. Extension to other potentials

In some force fields, such as the Kirkwood-Buff force field,[97-98] an improper potential is used frequently to force the molecule to be planar. Bending potentials in some force fields contain cubic and quartic terms (e.g., COMPASS,[99-100]) in addition to the quadratic term or a 1-3 nonbonded term such as Urey-Bradley in CHARMM.[95] In order to ensure that the trials are generated also according to these extra terms, similar to the regrowth of a three-branched molecule, Boltzmann rejection steps must be added to the trial generation loop. JG can be also extended to the GROMOS[101] force field where bending energy is proportional to $[\cos(\theta) - \cos(\theta_0)]^2$. The Jacobian factor can be adapted in a way such that $\cos(\theta)$ is generated according to its Gaussian distribution.

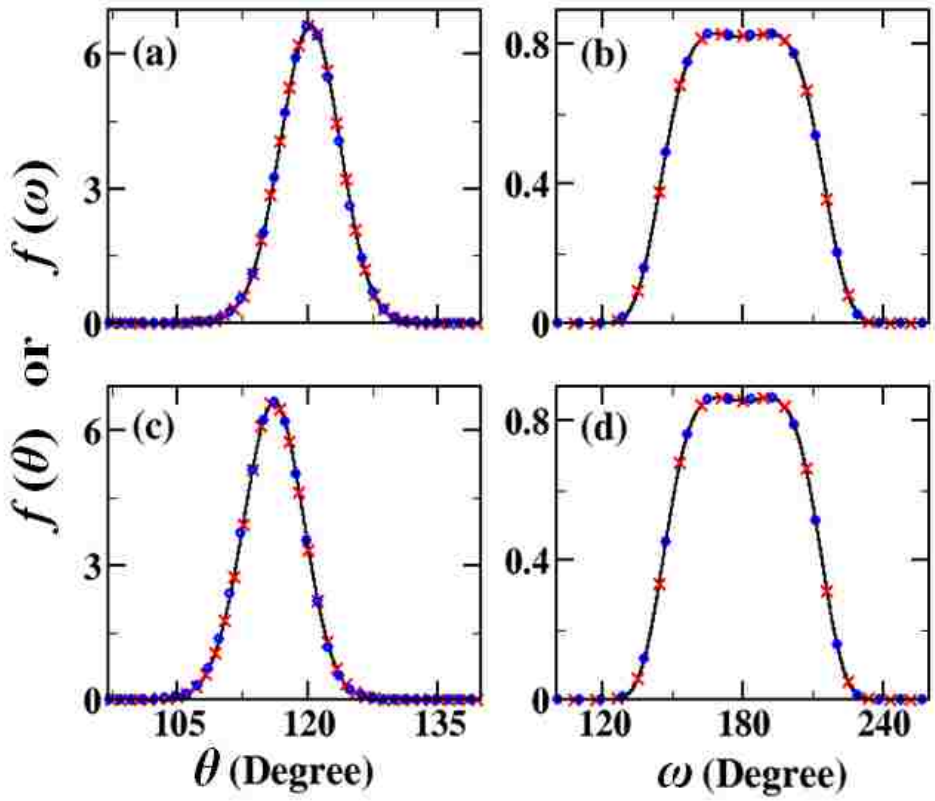

Fig. 2.10. Expected (solid lines) vs. simulated distributions (red × for (θ_1, θ_2, θ_{12}) and blue circles for (θ_1, θ_2, ω_{12}) as the growing variables) obtained for (a) the CH3-C=O bending angle, (b) the CH3-C(=O)-CH3 dihedral angle, (c) the CH3-C-CH3 bending angle, and (d) the CH_3-C(-CH_3)=O dihedral angle of acetone.

CHAPTER 3. SAMPLING INTERNAL SECTIONS OF CYCLIC AND POLYMERIC MOLECULES

3.1. Introduction

In this chapter, we extend the Jacobian-Gaussian (JG) method to improve the efficiency of sampling inner sections of molecules. In previous chapter, this method was developed based on two pillars. First, the conventional growth variables are transformed into those used explicitly in expressing the various intramolecular energies via simple transformations, so that these energetic terms can be considered directly in the trial generation. Second, basic geometrical constraints are applied to ensure that the generated trials are valid, which avoids the need of a biasing probability function. In previous chapter, only bending angle potential presents in intramolecular energies and Gaussian random number generators were used to generate bending angles. In this chapter, it is also required to generate a torsional angle, φ, from its probability density function $\exp[-\beta U_{\text{tor}}(\varphi)]$. For this purpose, φ is generated uniformly on $(0, 2\pi)$ until $\exp[-\beta U_{\text{tor}}(\varphi)] \geq \text{random}(0, 1)$, where random(0, 1) is a random number generated uniformly on (0, 1).

A TraPPE-UA model for linear[84] and cyclic[74] alkanes is used in this work where all atoms in CH_x group (e.g., CH_2 or CH_3) are united in one pseudoatom. The C-C bond length is fixed at 1.54 Å. $l_{i,i+1}$ represents the bond length that connects segment i and $i + 1$. The C-C-C bending angle, θ, has a harmonic potential (Eq. (2.1)) with the force field parameters of propane (Table 2.1). θ_i represents a bending angle made by $i - 1$, i, and $i + 1$ segments. Four sequential segments, C-C-C-C, make a torsional angle, φ, which is the angle between the two planes made by the first three and the last three segments. An OPLS united atom torsional potential[102] is used

$$U_{\text{tor}}(\varphi) = c_0 + c_1[1 + \cos(\varphi)] + c_2[1 - \cos(2\varphi)] + c_3[1 + \cos(3\varphi)] \quad (3.1)$$

where $c_0 = 0$, $c_1 = 2.9518$, $c_2 = -0.5669$, and $c_3 = 6.5793$ (all in kJ mol^{-1}). $\varphi_{i,i+3}$ represents the torsional angle made by segments i, $i + 1$, $i + 2$, and $i + 3$. There is a nonbonded pairwise-

additive Lennard-Jones (LJ) 12-6 potential between segments i and j that are either in two molecules or in the same molecule with more than three bonds between them

$$U_{LJ}(r_{ij}) = 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] \tag{3.2}$$

where r_{ij} is the distance between segments i and j, σ_{ij} and ε_{ij} are the LJ parameters that are computed using the Lorentz-Berthelot combining rules[103]

$$\begin{cases} \sigma_{ij} = (\sigma_i + \sigma_j)/2 \\ \varepsilon_{ij} = \sqrt{\varepsilon_i \varepsilon_j} \end{cases} \tag{3.3}$$

In linear alkanes, for CH_2, ε = 46 k_B and σ = 3.95 Å, and for CH_3, ε = 98 k_B and σ = 3.75 Å. In cyclic alkanes, for CH_2, ε = 51 k_B and σ = 3.89 Å, where k_B is the Boltzmann constant. The temperature for all simulations in this work is T = 300 K unless stated otherwise.

3.2. Method

Fig. 3.1 compares traditional procedure, which is explained in section 1.4, and the new procedure for growing three segments (i.e. segments 1-3) between two segments that are fixed at distance d (i.e., segments 0 and 4). For simplicity, it is assumed that two trials are generated and one of them is selected. In the traditional procedure, growing segments are removed from the old conformation in step 1. In step 2, two position trials are generated for segment 1. Then each trial is weighted with the biasing probability function, g, according to the distance of each trial from the final fixed segment (i.e. r_1 and r_2), and one of them is selected. The position of segment 2 is

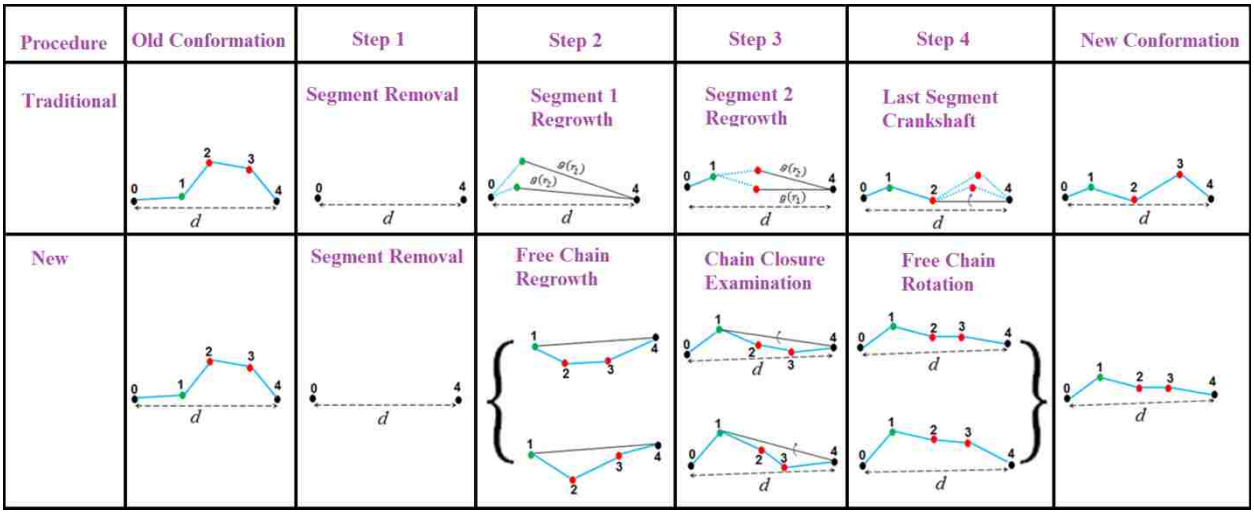

Fig. 3.1. Growing three segments between two fixed segments using the traditional and the new procedure.

determined in step 3 which is similar to step 2. For the last growing segment, two trials are generated using crankshaft moves in step 4, and one of them is selected, which forms the new conformation. In the new procedure, the first step is similar to the traditional procedure. In the next few steps, two trials are generated in parallel. In step 2, a free chain is grown which starts with segment 1 and ends with segment 4. The end-to-end distance of the free chain, the bond length between segment 0 and segment 1, and distance d must form a triangle to ensure chain closure and consequently, the biasing probability function is not required. This is examined in step 3, and if the triangle is not formed, step 2 is repeated. The final step is to rotate the freely grown chain around the line that passes segment 1 and segment 4 to form two new conformations. Then, one of them is selected as the new conformation.

The following sections explain the mathematical derivations of the new procedure for sampling internal parts of molecules where segments between two fixed segments are relocated. In section 3.2.1, it is assumed that there are two fixed segments at a definite distance and other

segments must be regrown between these two segments. Then, section 3.2.2 utilizes the developed method in section 3.2.1 to relocate the internal parts of a molecule.

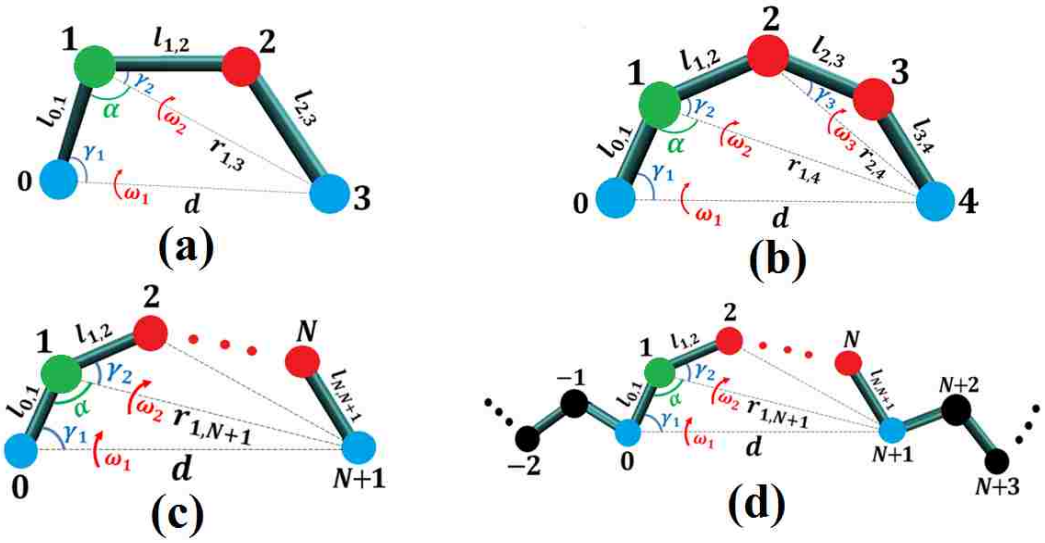

Fig. 3.2. Regrowth of (a) two, (b) three, and (c) N segments between two fixed points. (d) Regrowth of N segments in a chain.

3.2.1. Segment regrowth between two fixed points

This section describes the regrowth of different numbers of segments between two fixed endpoints as those shown in Figs. 3.2a, 3.2b, and 3.2c for regrowth of 2, 3, and N segments respectively. In this case, there are two kinds of interactions, $U^{\text{intra-in}}$ and $U^{\text{inter-in}}$. $U^{\text{intra-in}}$ is the sum of all bending and torsional energies (i.e., all intramolecular energies). $U^{\text{inter-}in}$ is the sum of all nonbonded pair interactions. K_{IN} trials are generated according to $\exp(-\beta U^{\text{intra-in}})$ and one of them (say, the k-th trial) is selected with the following probability

$$P_{\text{Select}}(k) = \frac{J_k \exp[-\beta U^{\text{inter-in}}(k)]}{W_{\text{IN}}} \tag{3.4}$$

where W_{IN} is the Rosenbluth[21] weight

$$W_{\text{IN}} = \sum_{k=1}^{K_{\text{IN}}} J_k \exp\left[-\beta U^{\text{inter-in}}(k)\right] \tag{3.5}$$

Finally, the new conformation is accepted according to the ratio of the Rosenbluth weights of new (n) and old (o) conformations, i.e., min[1, $W_{IN}(n)/W_{IN}(o)$].

A. Regrowth of two segments between two fixed points

This case is shown in Fig. 3.2a where segments 0 and 3 are fixed at distance d and segments 1 and 2 must be regrown. $U^{\text{intra-in}}$ is

$$U^{\text{intra-in}} = U_{\text{bend}}(\theta_1) + U_{\text{bend}}(\theta_2) + U_{\text{tor}}(\varphi_{0,3}) \tag{3.6}$$

Because there is no $U^{\text{inter-in}}$, the probability density function will be

$$f(d, \gamma_1, \omega_2) \propto \frac{\sin\gamma_1 d^2}{l_{1,2} l_{2,3} r_{1,3}} \exp\left[-\beta U^{\text{intra-in}}\right] \tag{3.7}$$

where γ_1 is $\widehat{1,0,3}$ angle, ω_2 is the angle between $\overline{0,1,3}$ and $\overline{1,2,3}$ planes, and $r_{1,3}$ is the distance between segments 1 and 3. The Jacobian factor, $\sin\gamma_1\ d^2\ (l_{1,2}\ l_{2,3}\ r_{1,3})^{-1}$, is the product of polar angle, $\sin\gamma_1$, radial distance, d^2, and rotational move,[60] $(l_{1,2}\ l_{2,3}\ r_{1,3})^{-1}$.

In regular CBMC, this molecule is regrown freely by generating θ_1, θ_2, and $\varphi_{0,3}$. So, the probability density function can also be written in terms of θ_1, θ_2, and $\varphi_{0,3}$, when segments 2 and 3 are grown freely

$$f(\theta_1, \theta_2, \varphi_{0,3}) \propto \sin\theta_1 \sin\theta_2 \exp\left[-\beta U^{\text{intra-in}}\right] \tag{3.8}$$

Comparing Eqs. (3.7) and (3.8), we can write

$$\sin\theta_1 \sin\theta_2 = \frac{\sin\gamma_1 d^2}{l_{1,2} l_{2,3} r_{1,3}} \left|\frac{\partial(d,\gamma_1,\omega_2)}{\partial(\theta_1,\theta_2,\varphi_{0,3})}\right| \tag{3.9}$$

To regrow segments 1 and 2 directly according to the Boltzmann factor specified by those intramolecular angles while at the same time satisfying the given constraint, it is more convenient to transform the growth variables from (d, γ_1, ω_2) into (d, θ_1, θ_2) via the following Jacobian factor

$$J(d,\theta_1,\theta_2) = \frac{\sin\gamma_1 d^2}{l_{1,2}l_{2,3}r_{1,3}}\left|\mathbf{det}\begin{pmatrix}\frac{\partial\gamma_1}{\partial\theta_1} & \frac{\partial\gamma_1}{\partial\theta_2}\\ \frac{\partial\omega_2}{\partial\theta_1} & \frac{\partial\omega_2}{\partial\theta_2}\end{pmatrix}\right| \tag{3.10}$$

The following two geometrical equations are held in this case

$$l_{0,1}^2 + d^2 - 2l_{0,1}d\cos\gamma_1 = l_{1,2}^2 + l_{2,3}^2 - 2l_{1,2}l_{2,3}\cos\theta_2 \tag{3.11}$$

$$\cos\omega_2 = \frac{\cos\theta_1 - \cos\alpha\cos\gamma_2}{\sin\alpha\sin\gamma_2} \tag{3.12}$$

where α and γ_2 are $\widehat{0,1,3}$ and $\widehat{2,1,3}$ angles, respectively. Eqs. (3.11) and (3.12) show that the diagonal components of the determinant in Eq. (3.10) are zero because γ_1 does not depend on θ_1 and ω_2 is independent of θ_2. By differentiating these two equations on both sides, we obtain the two off-diagonal terms

$$\frac{\partial\gamma_1}{\partial\theta_2} = \frac{l_{1,2}l_{2,3}\sin\theta_2}{l_{0,1}d\sin\gamma_1} \tag{3.13}$$

$$\frac{\partial\omega_2}{\partial\theta_1} = \frac{\sin\theta_1}{\sin\alpha\sin\gamma_2\sin\omega_2} \tag{3.14}$$

Substituting Eqs. (3.13) and (3.14) in Eq. (3.10), we have

$$J(d,\theta_1,\theta_2) = \left|\frac{d}{l_{0,1}r_{1,3}}\frac{\sin\theta_1\sin\theta_2}{\sin\alpha\sin\gamma_2\sin\omega_2}\right| \tag{3.15}$$

Eq. (3.12) can lead to two conclusions. First, it can be proved (Eq. (2.24)) that

$$|\alpha - \gamma_2| \le \theta_1 \le \min[(\alpha+\gamma_2), 2\pi - (\alpha+\gamma_2)] \tag{3.16}$$

Second, for each valid set of θ_1, α, and γ_2 (i.e. satisfying Eq. (3.16)), there are two possible solutions for ω_2 (i.e. $\omega_2 \in (0, \pi)$ and $\omega_2 \in (\pi, 2\pi)$).

In order to generate one valid trial at a definite d distance, θ_1 and θ_2 are generated independently and simultaneously according to their corresponding Gaussian distributions until these conditions are satisfied

1. $\theta_2 \in (0, \pi)$

2. $\overline{0,1,3}$ is a triangle (i.e., $|r_{1,3} - l_{0,1}| < d < r_{1,3} + l_{0,1}$)

3. $\theta_1 \in [|\alpha - \gamma_2|, \min\{(\alpha + \gamma_2), 2\pi - (\alpha + \gamma_2)\}]$

* One of the two answers for ω_2 (Eq. (3.12)) is chosen randomly.

4. $e^{-\beta U_{\text{tor}}(\varphi_{0,3})} \geq \text{random}(0, 1)$

ω_1, the rotational angle of segment 1 around d line, is generated uniformly on (0, 2π) because the probability density function (Eq. (3.7)) is independent of ω_1.

B. Regrowth of three segments between two fixed points

In this case (see Fig. 3.2b), segments 0 and 4 are fixed at distance d and segments 1, 2, and 3 must be regrown. $U^{\text{intra-in}}$ and $U^{\text{inter-in}}$ can be written as

$$U^{\text{intra-in}} = U_{\text{bend}}(\theta_1) + U_{\text{bend}}(\theta_2) + U_{\text{bend}}(\theta_3) + U_{\text{tor}}(\varphi_{0,3}) + U_{\text{tor}}(\varphi_{1,4}) \quad (3.17)$$

$$U^{\text{inter-in}} = U_{\text{LJ}}(r_{0,4}) \quad (3.18)$$

where $r_{0,4} = d$. The probability density function is

$$f(d, \gamma_1, \gamma_2, \omega_2, \omega_3) \propto \frac{\sin\gamma_1 \sin\gamma_2 d^2}{l_{2,3} l_{3,4} r_{2,4}} \exp\left[-\beta\left(U^{\text{intra-in}} + U^{\text{inter-in}}\right)\right] \quad (3.19)$$

where γ_1 is $\widehat{1,0,4}$, γ_2 is $\widehat{2,1,4}$, ω_2 is the angle between $\overline{0,1,4}$ and $\overline{1,2,4}$ planes, ω_3 is the angle between $\overline{1,2,4}$ and $\overline{2,3,4}$ planes, and $r_{2,4}$ is the distance between segments 2 and 4. In order to generate trials according to exp(-$\beta U^{\text{intra-in}}$), the growth variables are transformed from (d, γ_1, γ_2, ω_2, ω_3) into (d, θ_1, θ_2, θ_3, $\varphi_{1,4}$) by the following Jacobian factor

$$J(d, \theta_1, \theta_2, \theta_3, \varphi_{1,4}) = \frac{\sin\gamma_1 \sin\gamma_2 d^2}{l_{2,3} l_{3,4} r_{2,4}} \left|\frac{\partial\omega_2}{\partial\theta_1}\right| \left|\frac{\partial\gamma_1}{\partial r_{1,4}}\right| \left|\frac{\partial(r_{1,4}, \gamma_2, \omega_3)}{\partial(\theta_2, \theta_3, \varphi_{1,4})}\right| \quad (3.20)$$

According to the law of cosines, we have

$$r_{1,4}^2 = l_{0,1}^2 + d^2 - 2 l_{0,1} d \cos\gamma_1 \quad (3.21)$$

Thus, we can write

$$\left|\frac{\partial\gamma_1}{\partial r_{1,4}}\right| = \frac{r_{1,4}}{l_{0,1} d \sin\gamma_1} \quad (3.22)$$

According to Eq. (3.9), we have

$$\left|\frac{\partial(r_{1,4},\gamma_2,\omega_3)}{\partial(\theta_2,\theta_3,\varphi_{1,4})}\right| = \frac{l_{2,3}l_{3,4}r_{2,4}}{r_{1,4}^2}\frac{\sin\theta_2\sin\theta_3}{\sin\gamma_2} \tag{3.23}$$

Substituting Eqs. (3.14), (3.22), and (3.23) in Eq. (3.20), we have

$$J(d,\theta_1,\theta_2,\theta_3,\varphi_{1,4}) = \left|\frac{d}{l_{0,1}r_{1,4}}\frac{\sin\theta_1\sin\theta_2\sin\theta_3}{\sin\alpha\sin\gamma_2\sin\omega_2}\right| \tag{3.24}$$

As it can be seen in Fig. 3.2b, the regrowth process should produce a subsection that starts with segment 1 and ends with segment 4. Eq. (3.23), which was derived from Eq. (3.9), implies that this subsection can be produced either restrictively, where growing variables are $r_{1,4}$, γ_2, and ω_3, or freely, where growing variables are θ_2, θ_3, and $\varphi_{1,4}$. In the former case, growing variables are interdependent, which makes it difficult to generate them simultaneously. In the latter case, growing variables are independent, and because they are also energy variables, each of them can be generated directly according to its own Boltzmann distribution. Thus, to generate a valid trial at a definite d distance, a free chain, starting with segment 1 and ending with segment 4, is grown in vacuum by generating θ_2 and θ_3 according to their Gaussian distributions and $\varphi_{1,4}$ according to its torsional distribution, and θ_1 is generated simultaneously according to its Gaussian distribution until these conditions are satisfied

1. $\theta_2, \theta_3 \in (0, \pi)$

2. $\overline{0,1,4}$ is a triangle (i.e. $|r_{1,4} - l_{0,1}| < d < r_{1,4} + l_{0,1}$)

3. $\theta_1 \in [|\alpha - \gamma_2|, \min\{(\alpha + \gamma_2), 2\pi - (\alpha + \gamma_2)\}]$

* One of the two answers for ω_2 (Eq. (3.12)) is chosen randomly. γ_1 is calculated by the cosine law for $\overline{0,1,4}$ triangle and ω_1 is generated uniformly on (0, 2π). Segment 1 is regrown using $l_{0,1}$, γ_1, and ω_1. At this step, because the positions of segments 1 and 4 are determined, the freely grown chain is inserted into the system in such a way that the starting and the ending segments of this chain are located at the positions of segments 1 and 4, respectively. The inserted chain is

rotated around the $r_{1,4}$ line so that the angle between the $\overline{0,1,4}$ plane and the $\overline{1,2,4}$ plane becomes equal to ω_2.

4. $e^{-\beta U_{\text{tor}}(\varphi_{0,3})} \geq \text{random}(0,1)$

C. Regrowth of N segments between two fixed points

In this case (Fig. 3.2c), segments 0 and $N + 1$ are fixed at distance d and segments 1, 2, …, N must be regrown. $U^{\text{intra-in}}$ and $U^{\text{inter-in}}$ can be written as

$$U^{\text{intra-in}} = \sum_{i=1}^{N} U_{\text{bend}}(\theta_i) + \sum_{i=0}^{N-2} U_{\text{tor}}(\varphi_{i,i+3}) \tag{3.25}$$

$$U^{\text{inter-in}} = \sum_{i=0}^{N-3} \sum_{j=i+4}^{N+1} U_{\text{LJ}}(r_{ij}) \tag{3.26}$$

The procedure that is developed in the previous section is generalized in this section. The Jacobian factor is the product of a few factors encountered in growing a free chain (i.e., $\prod_{i=2}^{N} \sin\theta_i$), generating θ_1 instead of ω_2 (i.e., $\sin\theta_1/\sin\alpha \sin\gamma_2 \sin\omega_2$), and the chain closure restriction (i.e., $d/l_{0,1}r_{1,N+1}$)

$$J(d, \theta_1, \theta_2, \dots, \theta_N, \varphi_{1,4}, \varphi_{2,5}, \dots, \varphi_{N-2,N+1}) = \left| \frac{d}{l_{0,1}r_{1,N+1}} \frac{\prod_{i=1}^{N} \sin\theta_i}{\sin\alpha \sin\gamma_2 \sin\omega_2} \right| \tag{3.27}$$

For one valid trial to be generated at a definite d distance, a free chain, starting with segment 1 and ending with segment $N + 1$, is grown in vacuum (i.e., generate θ_2, θ_3, …, θ_N according to their corresponding Gaussian distributions and $\varphi_{1,4}$, $\varphi_{2,5}$, …, $\varphi_{N-2,N+1}$ from their torsional distributions), and θ_1 is generated independently and simultaneously according to its Gaussian distribution until these conditions are satisfied

1. $\theta_2, \theta_3, \dots, \theta_N \in (0, \pi)$

2. $\overline{0,1,N+1}$ is a triangle (i.e., $|r_{1,N+1} - l_{0,1}| < d < r_{1,N+1} + l_{0,1}$)

3. $\theta_1 \in [|\alpha - \gamma_2|, \min\{(\alpha + \gamma_2), 2\pi - (\alpha + \gamma_2)\}]$

* One of the two answers for ω_2 from Eq. (3.12) is chosen randomly. γ_1 is calculated by the cosine law for the $\overline{0,1,N+1}$ triangle and ω_1 is generated uniformly on (0, 2π). Segment 1 is regrown using $l_{0,1}$, γ_1, and ω_1. At this step, because the positions of segments 1 and N + 1 are determined, the freely grown chain is inserted into the system in such a way that the starting and the ending segments of this chain are located at the positions of segments 1 and N + 1, respectively. The inserted chain is rotated around the $r_{1,N+1}$ line so that the angle between the $\overline{0,1,N+1}$ plane and the $\overline{1,2,N+1}$ plane becomes equal to ω_2.

4. $e^{-\beta U_{\text{tor}}(\varphi_{0,3})} \geq \text{random}(0,1)$

3.2.2. Regrowth of *N* sequential segments in a molecule

In order to locally sample the internal sections of a molecule (see Fig. 3.2d), N sequential segments are selected randomly as growing segments (segments 1, 2, …, N), the two segments before and after growing ones are considered as fixed endpoints (segments 0 and N+1), other segments of the molecule are colored in black as shown in Fig. 3.2d. In addition to $U^{\text{intra}-\text{in}}$ and $U^{\text{inter}-\text{in}}$ (Eqs. (3.25) and (3.26)), $U^{\text{intra}-\text{out}}$ and $U^{\text{inter}-\text{out}}$ can also be present in this case because of the presence of other segments. $U^{\text{intra}-\text{out}}$ is the sum of all bending and torsional energies where at least one member used to define these angles is a growing segment and at least one member comes from the other segments. For instance, for the molecule shown in Fig. 3.2d, we have

$$U^{\text{intra}-\text{out}} = U_{\text{bend}}(\theta_0) + U_{\text{bend}}(\theta_{N+1}) + U_{\text{tor}}(\varphi_{-2,1}) + U_{\text{tor}}(\varphi_{-1,2}) + U_{\text{tor}}(\varphi_{N-1,N+2}) + U_{\text{tor}}(\varphi_{N,N+3}) \quad (3.28)$$

$U^{\text{inter}-\text{out}}$ is the sum of all nonbonded energies between growing segments and other segments or segments of other molecules. These N segments are regrown according to the procedure explained in previous section. There is one degree of freedom left in this case, which

can be used to rotate all regrown segments simultaneously around the line that passes the two fixed segments (i.e., segments 0 and N + 1). This degree of freedom means that, because ω_1 is generated randomly on (0, 2π), it is possible to generate different values of ω_1 without altering the other variables. Although $U^{\text{intra}-\text{in}}$ and $U^{\text{inter}-\text{in}}$ are independent of ω_1, $U^{\text{intra}-\text{out}}$ and $U^{\text{inter}-\text{out}}$ depend on ω_1. In this case, similar to previous cases, trials are generated according to exp(-$\beta U^{\text{intra}-\text{in}}$), and the Jacobian factor and other energetic terms must be included in the Rosenbluth weight. Because calculating intermolecular interactions requires computing distances, they are more computationally expensive than intramolecular energies. To reduce the computational cost, $U^{\text{intra}-\text{out}}$ is coupled[45] to $U^{\text{inter}-\text{in}}$ and $U^{\text{inter}-\text{out}}$. In addition, because nonbonded segments at shorter distances, which are stored in a neighbor list, have a higher impact on $U^{\text{inter}-\text{out}}$ and a greater effect on accepting or rejecting a conformation, $U^{\text{inter}-\text{out}}$ of each growing segment can be split[36] into two parts

$$U^{\text{inter}-\text{out}} = U_{r<r_{\text{CBMC}}}^{\text{inter}-\text{out}} + U_{r\geq r_{\text{CBMC}}}^{\text{inter}-\text{out}} \tag{3.29}$$

where r_{CBMC} is the split-energy cutoff. Because $U_{r<r_{\text{CBMC}}}^{\text{inter}-\text{out}}$ is calculated within a short distance, it is less computationally expensive and appears in the Rosenbluth weight. However, $U_{r\geq r_{\text{CBMC}}}^{\text{inter}-\text{out}}$, which is more computationally expensive, is computed only in the end when determining the overall acceptance probability. The whole procedure can be summarized in the following 9 steps

1. Select N sequential segments randomly and identify fixed endpoints.

2. Generate one trial for growing segments according to exp(-$\beta U^{\text{intra}-\text{in}}$) as explained in section 3.2.1.

3. Generate K_{Rot} trials of ω_1 uniformly on (0, 2π) and calculate $U^{\text{intra}-\text{out}}$ for each of them.

4. Select one of the trials of ω_1 (say, the k-th trial) with this probability

$$P_{\text{Select}}(k) = \frac{\exp\left[-\beta U^{\text{intra-out}}(k)\right]}{W_{\text{Rot}}} \tag{3.30}$$

with

$$W_{\text{Rot}} = \sum_{k=1}^{K_{\text{Rot}}} \exp\left[-\beta U^{\text{intra-out}}(k)\right] \tag{3.31}$$

5. Repeat steps 2-4 for K_{Trial} times to obtain K_{Trial} trials.

6. Calculate $U^{\text{inter-in}}$ and $U_{r<r_{\text{CBMC}}}^{\text{inter-out}}$ for K_{Trial} trials.

7. Select one trial (say, the i-th trial) with this probability

$$P_{\text{Select}}(i) = \frac{W_{\text{Rot}}(i) J_i \exp\left[-\beta\left(U^{\text{inter-in}}(i) + U_{r<r_{\text{CBMC}}}^{\text{inter-out}}(i)\right)\right]}{W_{\text{Trial}}} \tag{3.32}$$

with

$$W_{\text{Trial}} = \sum_{i=1}^{K_{\text{Trial}}} W_{\text{Rot}}(i) J_i \exp\left[-\beta\left(U^{\text{inter-in}}(i) + U_{r<r_{\text{CBMC}}}^{\text{inter-out}}(i)\right)\right] \tag{3.33}$$

8. Calculate $U_{r\geq r_{\text{CBMC}}}^{\text{inter-out}}$ for the selected trial in step 7.

9. The new conformation is accepted with this probability

$$P_{\text{Accept}}(o \to n) = \min\left\{1, \frac{W_{\text{Trial}}(n)}{W_{\text{Trial}}(o)} \exp\left[-\beta\left(U_{r\geq r_{\text{CBMC}}}^{\text{inter-out}}(n) - U_{r\geq r_{\text{CBMC}}}^{\text{inter-out}}(o)\right)\right]\right\} \tag{3.34}$$

3.3. Results and discussion

3.3.1. Segment regrowth between fixed points

For the new methodology to be verified, torsional distributions produced by this method are compared to those yielded from an MC simulation using regular CBMC to grow a free chain starting with segment 0 and ending with segment $N + 1$. It is assumed that all segments are CH_2 for linear alkanes and all bond lengths are 1.54 Å. The simulation run with regular CBMC produces the expected distribution of each torsional angle (i.e., $\varphi_{0,3}$, $\varphi_{1,4}$, …, $\varphi_{N-2,N+1}$) as well as the end-to-end distance (i.e., the distance between segments 0 and segments $N + 1$) distribution. Then, an MC simulation is run using fixed endpoints CBMC in which segments 0 and $N + 1$ are

fixed at distance d and segments 1, 2, …, N are regrown between them using the JG method to find the distribution of each torsional angle. This simulation is repeated at different values of d, which are generated from the end-to-end distribution obtained from the regular CBMC simulation, to compute the ensemble average distribution of each torsional angle.

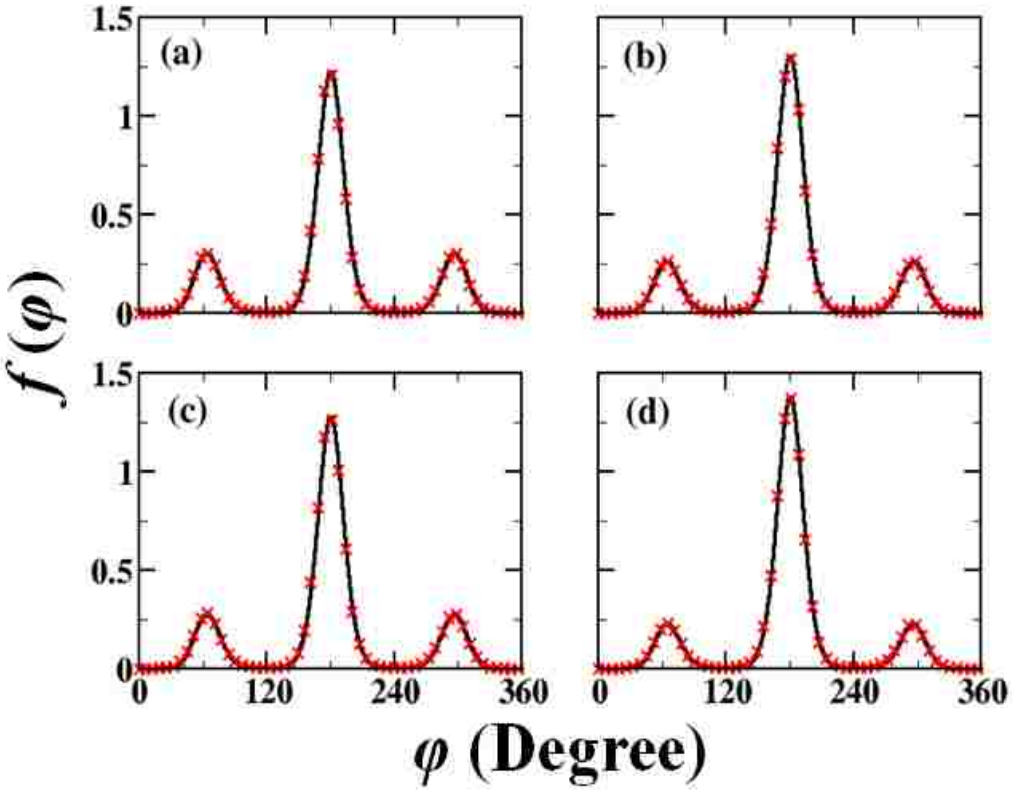

Fig. 3.3. Torsional distributions obtained from regular CBMC (solid black lines) vs. fixed-endpoints CBMC (red ×) for (a) $\varphi_{0,3}$ of a two-segments regrowth, (b) $\varphi_{0,3}$ (or $\varphi_{1,4}$) of a three-segments regrowth, and (c) $\varphi_{0,3}$ (or $\varphi_{2,5}$) and (d) $\varphi_{1,4}$ of a four-segment regrowth.

In this section, torsional distributions are presented for the growth of two, three, and four segments between two fixed points. In the case of a two-segment regrowth (see Fig. 3.2a), there is one torsional angle, $\varphi_{0,3}$. For a three-segment regrowth (see Fig. 3.2b), there are two torsional angles, $\varphi_{0,3}$ and $\varphi_{1,4,}$ whose distributions are equal because of the symmetry. A four-segment

regrowth involves three torsional angles, $\varphi_{0,3}$, $\varphi_{1,4}$, and $\varphi_{2,5}$. Again because of symmetry, $\varphi_{0,3}$ and $\varphi_{2,5}$ have the same distributions. Fig. 3.3 compares the two torsional distributions obtained by regular CBMC and fixed endpoints CBMC, which proves that JG indeed produces the correct results. Each torsional angle distribution has a global maximum which occurs at the trans conformation ($\varphi = \pi$) and two local symmetric maximums at the gauche conformations.

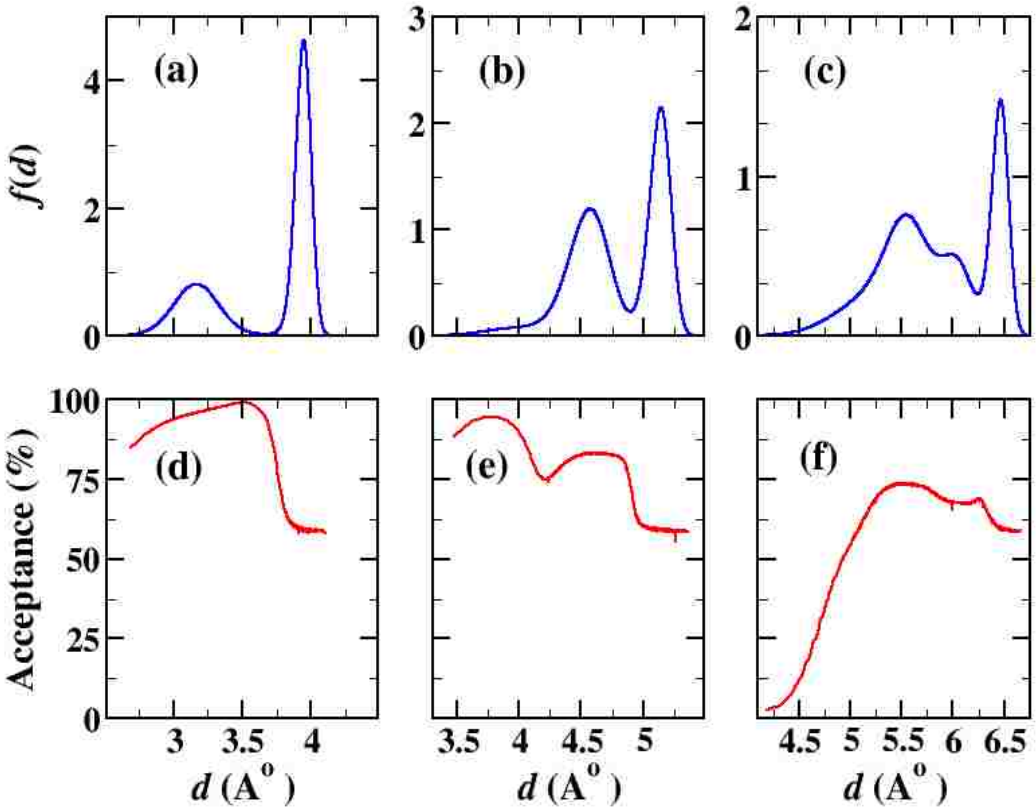

Fig. 3.4. Distance distribution between two fixed points separated by (a) two, (b) three, and (c) four segments. Fixed endpoints growth acceptance rates for growing (d) two, (e) three, and (f) four segments as function of this end-to-end distance.

Panels a, b, and c in Fig. 3.4 show the end-to-end distributions for fixed endpoints regrowth of two, three, and four segments, respectively. The peaks of each distance distribution occur when bending angles are at their equilibrium values and each torsional angle is located at one of the three maximums. In the case of the two-segment regrowth (see Fig. 3.2a), the distance

distribution (see Fig. 3.4a) has two peaks at $d \approx 3.15$ and 3.96 Å, which correspond to the gauche and trans conformations, respectively. The three-segment regrowth distribution (see Fig. 3.4b) shows one peak at $d \approx 5.15$ Å, when both torsional angles, $\varphi_{0,3}$ and $\varphi_{1,4}$, are at the trans conformations, and one peak at $d \approx 4.57$ Å, when one torsional angle is at the trans conformation and the other one is at the gauche conformation. There are other peaks at shorter distances when both torsional angles are at the gauche conformations, but these peaks are diminished due to the Lennard-Jones repulsion between segments 0 and 4. In all cases, the global maximum of the end-to-end distance distribution happens when all torsional angles are at their trans conformations.

The acceptance rates obtained for the fixed endpoints regrowth with one trial (i.e., $K_{\mathrm{IN}} = 1$ in Eq. (3.5)) as a function of the end-to-end distance for two, three, and four segments are displayed in Figs. 3.4d, 3.4e, and 3.4f, respectively. The acceptance rate can be affected by two factors: $U^{\text{inter-in}}$ and singularity in Jacobian. The acceptance rate for growing four segments (see Fig. 3.4f) is significantly lower at short end-to-end distances due to the $U^{\text{inter-in}}$ factor, i.e., Lennard-Jones repulsions between segments 0 and 4 and/or between segments 1 and 5. This $U^{\text{inter-in}}$ term cannot affect either two- or three-segment regrowths. Although it is absent in the former, in the latter it is equal for the old and new conformations, which will be counterbalanced in the detailed balance condition (see Eq. (3.18)). Thus, for these two cases, the Jacobian factor would be the only source to affect the acceptance rates. For these two (and also the four-segment regrowth), noticeably lower acceptance rates (~ 58%) are observed at longer end-to-end distances. This could be explained by the fact that the denominator of the Jacobian (see Eqs. (3.15), (3.24), and (3.27)) includes a $\sin\omega_2$ term which approaches zero when ω_2 approaches 0 or π. This causes a singularity issue, which affects directly the trans conformation (which occurs at longer end-to-end distances) because $\omega_2 \rightarrow 0$ there, but not so much for the other stable gauche conformations

where ω_2 is not close to either 0 or π. This singularity issue was also observed for the regrowth of a two-branched planar molecule in chapter 2, which led to the proposal of using a different set of variables for the regrowth procedure. Here it is also possible to change the growth variables to obtain higher acceptance rates. For instance, in the two-segment regrowth, the singularity issue can be avoided if (d, ω_2, θ_2) are used as the growth variables since the Jacobian factor becomes

$$J(d, \omega_2, \theta_2) = \frac{d \sin\theta_2}{l_{0,1} r_{1,3}} \tag{3.35}$$

Table 3.1. Ensemble averages of the acceptance rates (%) using JG for growing N segments between two fixed points with K_{IN} trials.

N	K_{IN}				
	1	2	5	10	20
2	71.94	76.98	82.89	86.67	89.71
3	71.73	77.19	83.18	86.92	89.93
4	65.28	71.46	78.70	83.18	86.91
5	62.62	69.24	77.02	81.94	86.01
6	58.77	66.27	74.97	80.44	84.86
7	56.26	64.28	73.65	79.43	84.16
8	52.97	61.77	71.77	77.99	83.10
9	50.33	59.61	70.24	76.93	82.31
10	47.45	57.24	68.55	75.59	81.26

In this case, for each trial generation at a definite value of d, θ_2 is generated from its corresponding Gaussian distribution, and ω_2 is generated uniformly on (0, 2π) until $\overline{0,1,3}$ becomes a triangle and $\exp[-\beta(U_{bend}(\theta_1) + U_{tor}(\varphi_{0,3}))] \geq$ random(0, 1). The acceptance rate at

every end-to-end distance was found to be nearly 97%. However, trial generation is 4-5 times slower due to the requirement of a generated trial to be according to both bending (θ_1) and torsional ($\varphi_{0,3}$) distributions. Although changing the growth variables can improve the acceptance rates by a few percent for this simple two-segment regrowth, it has no noticeable effect on other more complicated cases. As explained later, the acceptance rate of relocating internal segments of molecules is mainly determined by $U^{\text{intra-out}}$ and $U^{\text{inter-out}}$.

Listed in Table 3.1 are the ensemble averages of the acceptance rates obtained using the JG method for growing N segments ($2 \le N \le 10$) with different number of trials ($1 \le K_{\text{IN}} \le 20$). It is clear from these data that high acceptance rates are attainable with relatively low number of trials. In addition, increasing the number of trials, K_{IN}, has a higher effect in a larger number of growing segments, N, because the use of just a few choices can quickly allow the molecule to find a more suitable conformation in terms of $U^{\text{inter-in}}$ by avoiding bad contacts. Typically, the number of trials used in the regular CBMC to explore this relatively soft, nonbonded

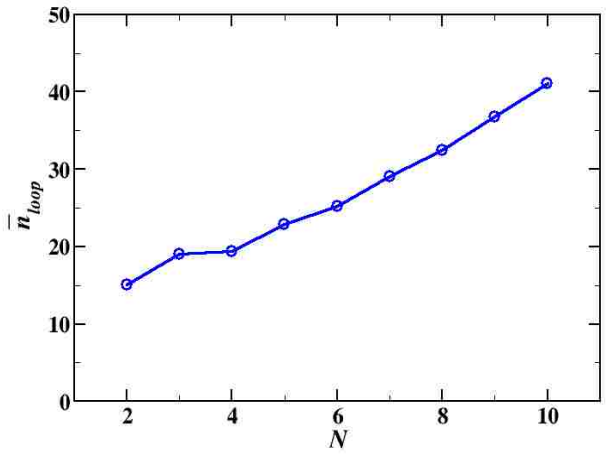

Fig. 3.5. Average number of generation loops needed to generate a valid trial for growing N segments between two fixed points.

configurational space is around 10 for each growing segment. In contrast, the presence of singularity and, correspondingly a rather sharp distribution of the Jacobian factor at certain

geometry (such as trans) would require a significantly larger number of trials, similar to the use of a large number of trials in the regular CBMC (up to 1000) to sample a rather stiff bending angle. Thus, using a K_{IN} up to 20 only a small improvement is observed with regard to this singularity issue. However, because this problem only occurs at certain geometry, for all cases the overall acceptance rate obtained from all possible geometries is above 80% when using 20 trials.

$\bar{n}_{\mathrm{loop}}$ is defined as the average number of times that the trial generation loop is implemented until a valid trial, which satisfies all conditions, is produced. Fig. 3.5 shows that $\bar{n}_{\mathrm{loop}}$ increases linearly with the number of growing segments. Energetic condition (i.e., $\varphi_{0,3}$ must be sampled according to the torsional distribution via the Boltzmann rejection scheme) is the main reason why several loops are needed at low number of growing segments. However, as the number of growing segments increases, the distance distributions become broader. Then, geometrical conditions (i.e., conditions 2 and 3) also increase the average number of loops required. In general, $\bar{n}_{\mathrm{loop}}$ is computationally reasonable because a free chain can be grown rapidly according to its intramolecular interactions.

3.3.2. Regrowth of *N* sequential segments in a chain molecule

The new method is examined on *n*-C_{20} and *n*-C_{100} alkane molecules. The procedure of growing *N* internal segments is explained in section 3.2.2.

The conformational space of this molecule can also be sampled using regular CBMC moves where one random segment is chosen and all segments are removed toward one random end and then regrown segment by segment. In growing each segment, 20 trials are generated where for each trial, one bending and one torsion angles are generated according to their probability density functions and the Boltzmann factor of intermolecular energy terms is calculated for each

trial. One of the trials is selected according to this Boltzmann factor. Our computations show that increasing the number of trials does not remarkably improve acceptance rates particularly for high number of segments.

For this N-segment regrowth, regular CBMC is expected to yield higher acceptance rates than fixed endpoints CBMC for three reasons. First, regular CBMC is not restricted by the endpoint, so it has more freedom to avoid unfavorable high-energy conformations. Second, in regular CBMC, intramolecular energies (i.e., bending and torsional energies) are independent of each other, whereas several intramolecular energies (see Eq. (3.28)) are coupled to each other in fixed endpoints CBMC. Third, because regular CBMC is implemented segment by segment and the intermolecular energy of each growing segment is considered at its growth steps, the intermolecular energies of sequential growing segments are independent to a certain extent. In contrast, the JG method for sampling internal segments regrows all selected segments for each trial generation and then calculates the intermolecular energies (i.e., $U^{\text{inter-in}}$ and $U^{\text{inter-out}}$) for all growing segments simultaneously, so the trial needs to be energetically favorable for all growing segments, which is more difficult. Thus, the acceptance rates of regular CBMC moves for different number of growing segments can be considered as the upper limit for the acceptance rates of fixed endpoints CBMC moves.

In our simulation, K_{Rot} (see Eqs. (3.30) and (3.31)) is set to 100 as further increase of this parameter does not lead to any appreciable improvement in the acceptance rate. Following previous work,[73] r_{CBMC} (see Eqs. (3.32)-(3.34)) is set to 5 Å. Fig. 3.6 shows acceptance rates and representative snapshots obtained for the n-C_{20} and n-C_{100} chains. It is clear that, with the increase of the chain length, nonbonded interactions become more important in forming conformations, i.e., they make the chain fold on itself such that each segment is surrounded

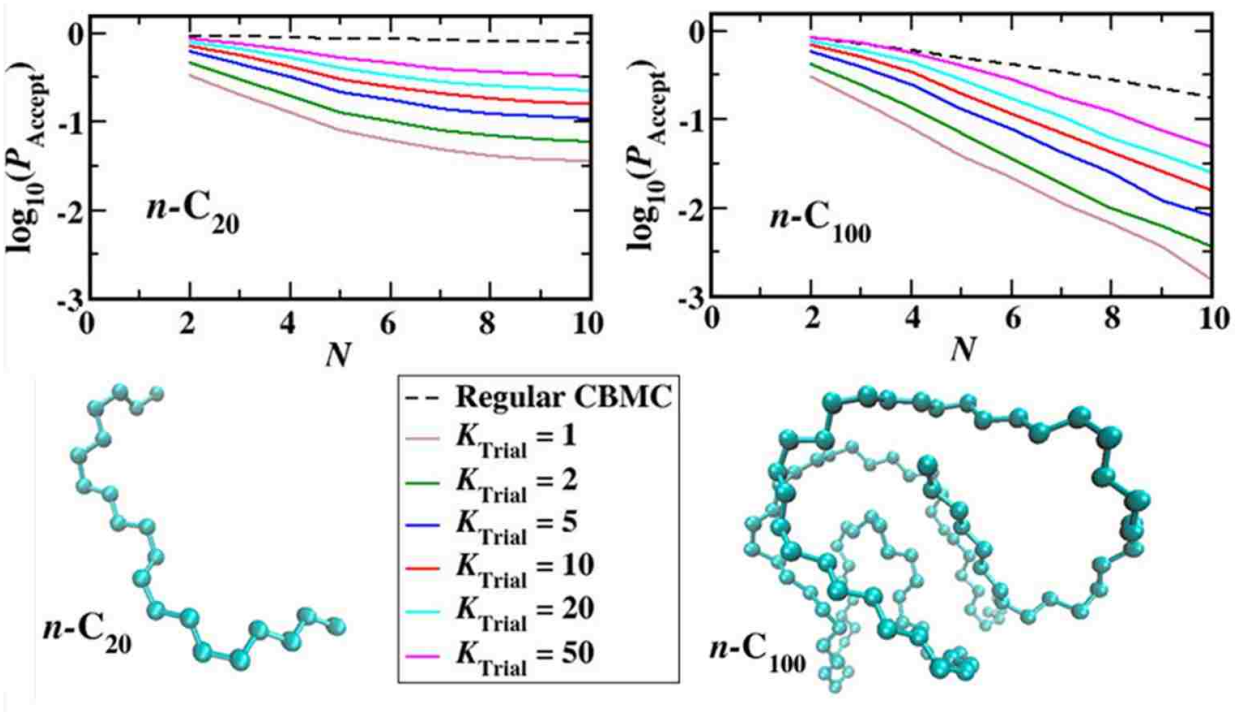

Fig. 3.6. Acceptance rates of regular CBMC (dashed line) with 20 trials and fixed endpoints CBMC (solid lines) with different number of trials for growing N segments in n-C_{20} and n-C_{100} alkane chains. Representative snapshots of n-C_{20} and n-C_{100} alkane chains.

by more nonbonded segments. Thus, the available space for growing segments using regular and fixed-endpoints CBMC moves becomes even more restricted. As a result, these moves are more likely to be rejected. This can be observed in Fig. 3.6 where the acceptance rates obtained for n-C_{20} are substantially higher than those obtained for n-C_{100}. Because this issue is present in both regular and fixed endpoints CBMC, the lower acceptance rates observed for both n-C_{20} and n-C_{100} when using fixed endpoints vs. regular CBMC are mainly due to $U^{\text{intra-out}}$. Comparing these results with those obtained with previous approaches[72-73] proves that JG is much closer to the upper limit (i.e., the acceptance rates obtained by regular CBMC).

The efficiency of the new method in conformation generation is compared with both the crankshaft and the rebridging configurational bias (RCB)[72] method by measuring the decay rate of half-chain end-to-end autocorrelation function.[72] In the JG move, $K_{\text{Rot}} = 30$, $r_{\text{CBMC}} = 5$ Å, and

K_{Trial} = 1, and 10 trials are generated for each crankshaft move. These parameters are chosen to yield an optimal ratio of the acceptance rate to the CPU time. JG and RCB are compared in simulating an isolated *n*-C_{70} alkane chain at 400 K using the NERD[104] force field and a softer bending potential with $k_\theta = 31250K$ (while bond lengths and torsional angle potential remain the same). The autocorrelation function of the crankshaft algorithm for this molecule serves as a reference to compare JG with RCB. JG and the crankshaft algorithm are also compared in relaxing an isolated *n*-C_{100} alkane chain at 300 K using the TraPPE-UA force field that has a stronger bending potential (Table 2.1). In all simulations, after each move, a one-site regular CBMC is performed to vary the position of one of the two end segments randomly. The results of all simulations are shown in Fig. 3.7. According to Chen and Escobedo,[72] for an isolated *n*-C_{70} chain, the autocorrelation function in the RCB method reaches zero when the autocorrelation function of crankshaft move is about 0.82 that takes more time than JG (see Fig. 3.7). Thus, the new method can produce new conformations very efficiently without the requirement for biasing

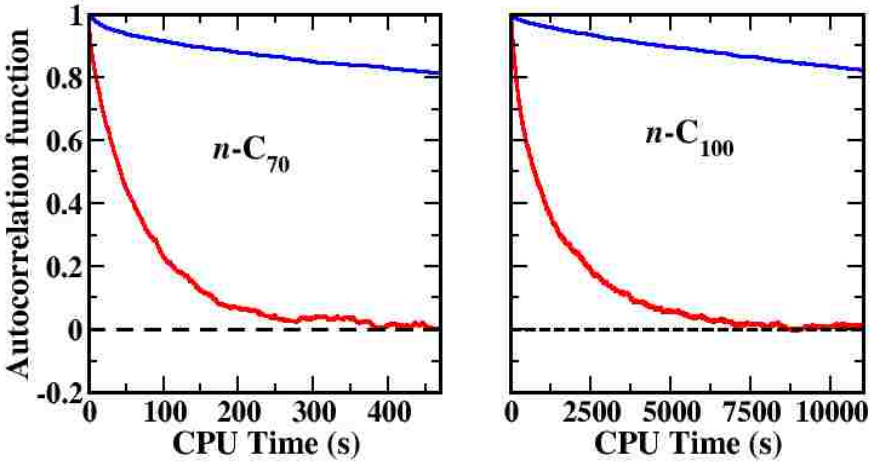

Fig. 3.7. Half-chain end-to-end autocorrelation function of JG (red) and crankshaft (blue) moves for an isolated *n*-C_{70} alkane chain at 400*K* and an isolated *n*-C_{100} alkane chain at 300*K*.

probability functions. For an isolated n-C_{100} chain, Fig. 3.7 shows that JG is also very efficient for models with strong intramolecular interactions at low temperatures when the crankshaft algorithm is very time consuming in producing new conformations.

3.3.3. Regrowth of *N* sequential segments in a cyclic molecule

JG is also examined on growing the internal segments of cyclic molecules with $K_{Rot} = 100$ and $r_{CBMC} = 5$Å. Simulations were run for cyclododecane and cyclohexane as examples of large and small cyclic molecules, respectively.

Fig. 3.8a shows the acceptance rates for growing different number of segments in cyclododecane. In comparison with growing the internal segments of a linear chain, the acceptance rate is lower for this case, because it is less probable, particularly at higher number of growing segments, to regrow segments between two fixed points at shorter distances. As explained in section 3.3.1 (see Fig. 3.4f), the acceptance rate is lower due to nonbonded repulsions in $U^{inter\text{-}in}$.

The acceptance rates for growing two and three segments in cyclohexane are shown in Fig. 3.8b. For this molecule, a different torsional potential model[105] is used (see Fig. 3.8c). The simulated average distribution of all torsional angles in cyclohexane has a peak close to 55^o (Fig. 3.8c) due to the rigid structure of the ring which has been observed in the previous experimental[106] and simulation[107] works. Because each pair of segments inside this molecule is separated by fewer than four bonds, there is no nonbonded interaction inside the molecule. Thus, $U^{intra\text{-}out}$ is the only factor which affects the acceptance rate. Because this energetic term is similar for growing different number of segments, the acceptance rates for growing two and three segments in cyclohexane are close to each other. In addition, the high acceptance rates for growing three segments prove that our method can be used to regrow the whole molecule where the first three

segments are regrown using the regular procedure described in chapter 2 and the last three segments are regrown using the procedure described here. Thus, the new method is applicable to transfer cyclic molecules between phases in grand canonical[108] and Gibbs[109] ensembles.

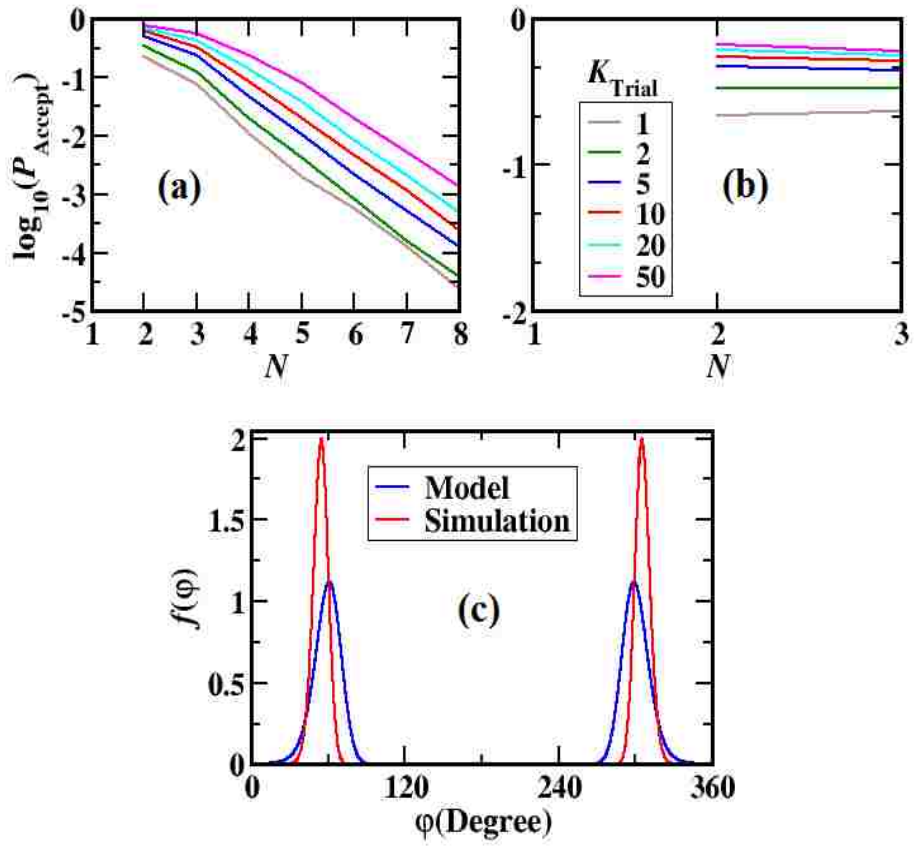

Fig. 3.8. Acceptance rates of growing *N* segments of (a) cyclododecane and (b) cyclohexane with different number of trials. (c) Torsional potential model (blue) and average torsional angle distribution (red) for cyclohexane.

3.3.4. Extensions to other cases

As it was explained in section 2.5.2, this method can be extended to nonharmonic bending potentials using extra Boltzmann rejection steps. In some force fields, such as TraPPE-UA for acrylates,[110] there are 1-4 potentials in addition to torsional interactions. These extra energetic terms can be included in the Rosenbluth weight of $U^{\text{inter-in}}$ or $U^{\text{intra-out}}$.

It is also possible to extend this method to fully flexible molecules where bond lengths are also generated according to their probability density functions using the Boltzmann rejection scheme in a decoupled[111] style.

Because a free chain of segments 1, 2, …, $N + 1$ is grown in this method, it can be extended to a molecule where segments 2 or 3 or … or N are branched points. Furthermore, if segment 1 is a branched point, other bending and dihedral angles can be generated independently and simultaneously as described in previous chapter. In this case, each rotational angle generation must consider all associated torsional energies in the Boltzmann rejection step.

CHAPTER 4. WATER-AMMONIA/AMINE NUCLEATION

4.1. Introduction

Nucleation is a common event that occurs in many biological,[112] industrial,[113] and atmospheric[114] phenomena. So, nucleation affects cloud formation,[115] weather and climate change,[116] solar radiation,[117] and public health.[118] Nucleation happens when the system is not at equilibrium.[119] For instance, when the vapor pressure is greater than the saturation pressure, few molecules of the vapor phase aggregate and form a liquid cluster to reduce the free energy. However, since the ratio of the surface to volume is high for small clusters, the free energy of the cluster surface increases the total free energy. Thus, the free energy profile passes a maximum which is called nucleation barrier. Nucleation can occur homogeneously, where there is only one molecular type, or heterogeneously, i.e., in presence of other agents such as surface of a solid or other molecular types. In classical nucleation theory for homogeneous nucleation,[120] it is assumed that the cluster is spherical and the properties of the cluster, such as liquid density ρ_l and surface tension σ, are equal to those of the bulk phase. Thus, the free energy difference can be written as a function of the cluster size n as follows

$$\Delta G(n) = -n\Delta\mu + A\sigma \tag{4.1}$$

where $\Delta\mu$ is the chemical potential difference and A is the area of the cluster surface. The chemical potential difference for vapor liquid nucleation can be written as

$$\Delta\mu = k_B T \ln\left(\frac{\rho}{\rho^{\text{sat}}}\right) \tag{4.2}$$

Where ρ is the vapor density and ρ^{sat} is the saturation vapor density. The area of a spherical cluster of size n is

$$A = \left(\frac{36\pi}{\rho_l^2}\right)^{\frac{1}{3}} n^{\frac{2}{3}} \tag{4.3}$$

In the atmosphere, precursor gases, such as sulfuric acid, ammonia, etc., can act as nucleating agents[121] where condensation of water molecules happens and ultrafine aerosols form which can grow to larger particles. The annual emissions of ammonia, methylamine (MA), dimethylamine (DMA), and trimethylamine (TMA) are 58000, 96.2, 38.2, and 196 Gg/yr respectively.[122] Thus, experimental[123-125] and computational[126-128] works have been done to study the effect of these species in atmospheric nucleation. Computational studies usually use density functional theory (DFT) where only a few molecules present in the system and it would be very expensive for higher number of molecules. On the other hand, since nucleation occurs at molecular levels, it is very difficult to be observed experimentally. In this chapter, MC is applied to cover a wide range of cluster sizes, from a few to tens of molecules, to study the effect of ammonia/amines on water nucleation. The details of simulations are explained in section 4.2 and results are discussed in section 4.3.

4.2. Simulation details

In these simulations, water and ammonia are assumed to be rigid molecules where bond lengths (e.g., O-H and N-H) and bending angles (e.g., $\widehat{HOH}$ and $\widehat{HNH}$) are fixed. A four-site[129] and a five-site[130] potentials are used for water and ammonia respectively. Lennard-Jones parameters are zero for hydrogens and nonzero for oxygen (or nitrogen). Partial positive charges are located on hydrogen sites whereas partial negative charge is not located at the position of

Table 4.1. Force field parameters for amines.

Bond	length (Å)	Angle	k_θ/k_B (K)	θ_0 (Deg)	Amine	Site	σ (Å)	ε (K)	q (e)
N-C	1.448	H-N-H	43910	106.4	MA	N	3.34	111	-0.892
						H	0	0	0.356
						CH_3	3.75	98	0.18
N-H	1.01	H-N-C	62500	112.9	DMA	N	3.52	58	-0.745
						H	0	0	0.385
						CH_3	3.75	98	0.18
		C-N-C	50356	109.5	TMA	N	3.78	12	-0.54
						CH_3	3.75	98	0.18

oxygen (or nitrogen), but on the symmetry axis with a displacement from oxygen (or nitrogen). A transferable potential for phase equilibria-explicit hydrogen (TraPPE-EH) has been proposed[89] for amines where all hydrogen atoms are treated explicitly. In order to reduce computational costs, a TraPPE-UA is used in these simulations where CH_3 group is considered as one pseudoatom. Bond lengths are rigid and bending angles have harmonic potential (Eq. (2.1)). Intermolecular interactions include Lennard-Jones and electrostatic components. Force field parameters are presented in Table 4.1. In order to ensure that the united atom model is accurate enough, we run GEMC simulations using Towhee package[47, 131-135] to obtain vapor-liquid phase coexistence curve and compare it with experimental results.

In our MC nucleation simulation, in addition to conventional translation and rotation moves, we use aggregation volume bias Monte Carlo (AVBMC)[76-77, 108] to swap molecules between the gas phase and the cluster. For flexible amine molecules, CBMC is used to sample molecular

conformations and to regrow a molecule in a swap move. We also examine the Jacobian-Gaussian method for these molecules. A self-adaptive umbrella sampling (US)[53, 78, 136] is used to calculate the two-dimensional nucleation free energy (NFE) plot that is a function of number of

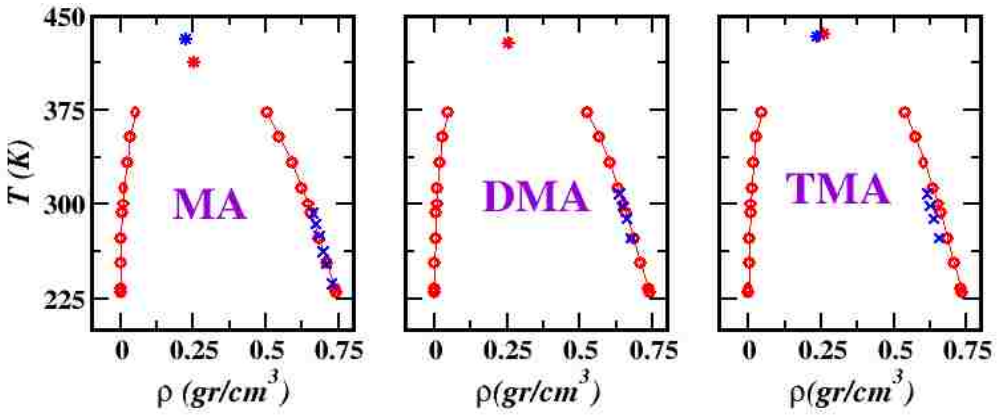

Fig. 4.1. GEMC simulation results (red) vs. experimental data (blue) of vapor-liquid coexistence curve for MA, DMA, and TMA.

molecules of water n_W and ammonia/amine n_A. After obtaining NFE plot at two arbitrary gas phase densities for water ρ_W and ammonia/amine ρ_A, NFE can be calculated at other gas phase densities, ρ'_W and ρ'_A, as following[137]

$$\Delta G_{\rho'_W,\rho'_A}(n_W,n_A) = \Delta G_{\rho_W,\rho_A}(n_W,n_A) - n_W k_B T \ln\left(\frac{\rho'_W}{\rho_W}\right) - n_A k_B T \ln\left(\frac{\rho'_A}{\rho_A}\right) \quad (4.4)$$

As a reference point for NFE, the free energy is set to be zero for a concentration of 1 droplet/Å^3. The concentration of a cluster of size n can be written as the sum of cluster concentrations of sizes n with different combinations of water and ammonia/amine

$$P^{\mathrm{tot}}(n) = \sum_{n_W=0}^{n} P(n_W, n-n_W) \quad (4.5)$$

or

$$\exp[-\Delta G^{\mathrm{tot}}(n)/k_B T] = \sum_{n_W=0}^{n} \exp[-\Delta G(n_W, n-n_W)/k_B T] \quad (4.6)$$

This one-dimensional free energy is used to examine the effect of the second molecule type on water nucleation. An arbitrary free energy barrier is chosen as ΔG^{onset}. The activity of each molecule type is defined as

$$a = \frac{\rho}{\rho^o} \tag{4.7}$$

Table 4.2. Simulation and experimental properties for MA, DMA, and TMA.

Amine	Force Field	T_C (K)	ρ_C (gr/cm^3)	T_B (K)
MA	TraPPE-UA	412.9	0.253	254.9
	Experiment	431	0.224	267
DMA	TraPPE-UA	428.3	0.254	266.5
	Experiment	438	-	281
TMA	TraPPE-UA	435.3	0.261	265.8
	Experiment	433	0.234	275

where ρ^o is the gas phase density of the molecule type which results in ΔG^{onset} barrier for homogeneous nucleation. Using Eq (4.6) for $\Delta G^{tot} = \Delta G^{onset}$ and Eq. (4.4), it is possible to calculate a_A vs. a_W plot (or onset plot) where A and W subscripts stand for ammonia/amine and water respectively. If the onset plot is below the diagonal line, the presence of the second molecule type enhances the nucleation of water because a lower gas phase density is required for nucleation and if the plot is above the diagonal line, the two molecule types are reluctant to nucleate with each other. Simulations of binary nucleation were run at 230K and 300K as low and high temperatures.

4.3. Results and discussions

Fig. 4.1 compares GEMC simulation results and experimental data[89, 138-139] of vapor-liquid coexistence curve for MA, DMA, and TMA. The critical temperature T_C and the critical density ρ_C are calculated using equilibrium densities of liquid ρ_{liq} and vapor ρ_{vap} phases according to the scaling law[110]

$$\rho_{liq} - \rho_{vap} = B(T - T_C)^{0.325} \tag{4.8}$$

$$\frac{1}{2}\left(\rho_{liq} + \rho_{vap}\right) = \rho_C + A(T - T_C) \tag{4.9}$$

where A and B are constants. The normal boiling point T_B is calculated according to Clausius-Clapeyron equation.[140] Table 4.2 present simulation and experimental[89] values for critical properties and normal boiling points for MA, DMA, and TMA. Fig. 4.1 and Table 4.2 show that TraPPE-UA is an accurate force field to be used in nucleation simulations.

Fig. 4.2 shows two-dimensional NFE contours for binary nucleation of water with ammonia/amine in 230 and 300K at given gas phase densities. The nucleation path can be determined according to the saddle point which can move to water-rich domain or ammonia/amine-rich domain or vanish by varying gas phase densities. So, the nucleation mechanism depends on gas phase densities.

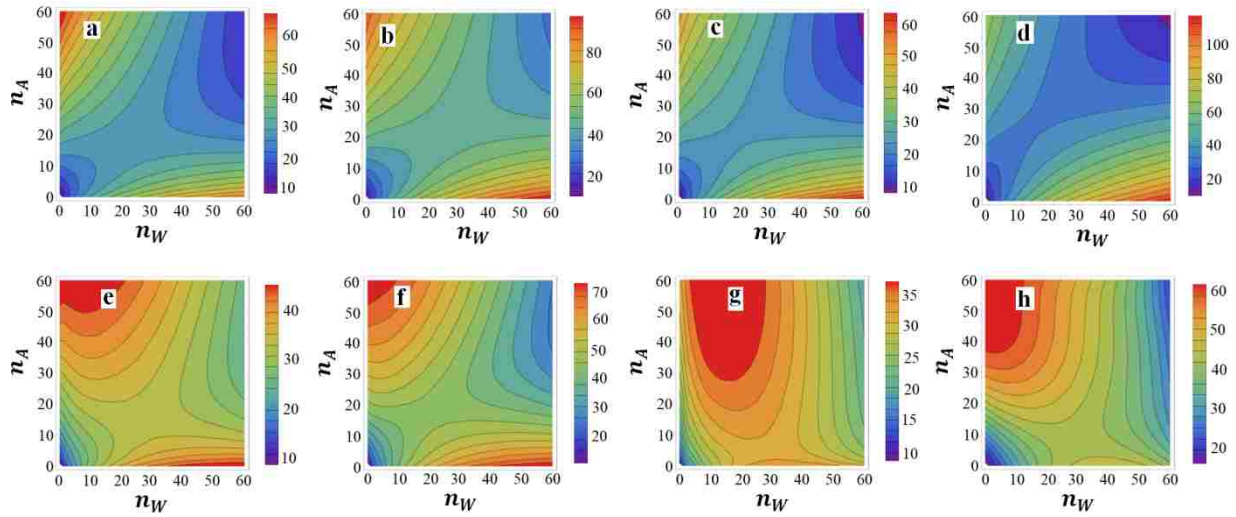

Fig. 4.2. Contour of NFEs (in units of k_BT) for (a) water-ammonia at 300K with $\rho_W = 3 \times 10^{-6}$ Å^{-3} and $\rho_A = 2 \times 10^{-4}$ Å^{-3}, (b) water-ammonia at 230K with $\rho_W = 2 \times 10^{-8}$ Å^{-3} and 3×10^{-5} Å^{-3}, (c) water-MA at 300K with $\rho_W = 2.75 \times 10^{-6}$ Å^{-3} and $\rho_A = 1.75 \times 10^{-4}$ Å^{-3}, (d) water-MA at 230K with $\rho_W = 1.45 \times 10^{-8}$ Å^{-3} and $\rho_A = 2 \times 10^{-5}$ Å^{-3}, (e) water-DMA at 300K with $\rho_W = 3.75 \times 10^{-6}$ Å^{-3} and $\rho_A = 1.4 \times 10^{-4}$ Å^{-3}, (f) water-DMA at 230K with $\rho_W = 3 \times 10^{-8}$ Å^{-3}, $\rho_A = 1.75 \times 10^{-5}$ Å^{-3}, (g) water-TMA at 300K with $\rho_W = 5 \times 10^{-6}$ Å^{-3} and $\rho_A = 1.4 \times 10^{-4}$ Å^{-3}, and (h) water-TMA at 230K with $\rho_W = 5 \times 10^{-8}$ Å^{-3} and $\rho_A = 3 \times 10^{-5}$ Å^{-3}.

Fig. 4.3 presents the onset plots at 230K, where ΔG^{onset} = 50.64 k_BT, and 300K, where ΔG^{onset} = 32.24 k_BT. These results show that as temperature increases, water becomes more reluctant to co-nucleate with ammonia/amine. DFT calculations also show[141] that while the free energy for MA-water system is positive at 298.15K, it is negative at 216.65K. It can be seen that while MA enhances water nucleation more than ammonia at 230, MA is more reluctant to nucleate with water than ammonia at 300K. These onset plots are affected by two factors: the

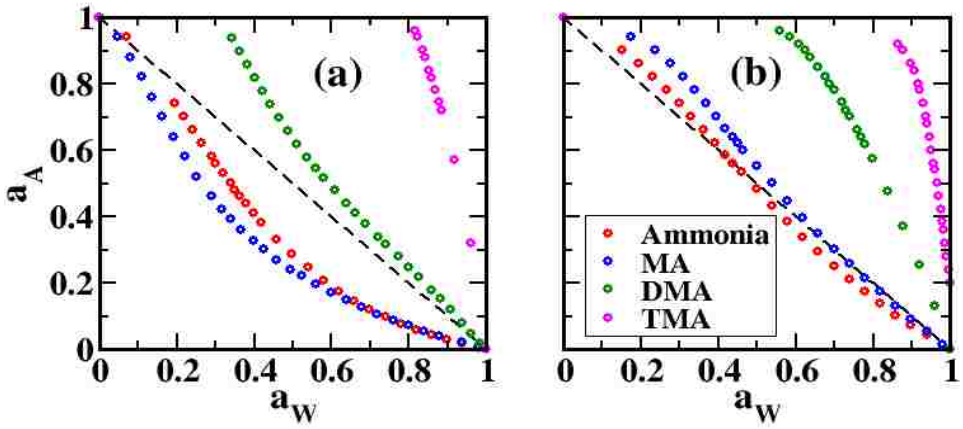

Fig. 4.3. Onset plots at (a) 230K and (b) 300K.

interactions between water and ammonia/amine and the stability of the cluster. As the number of methyl groups increases, the second molecule type becomes more hydrophobic and consequently co-nucleation with water is more unfavorable. In addition, clusters with lower surface free energies are more stable. These two factors are assessed quantitatively as follows.

Fig. 4.4 shows a few snapshots of binary clusters at 230K and 300K. It can be seen that in both temperatures, water molecules are more likely to locate in the center of the cluster and ammonia/amine molecules are more probable to be at the surface. This can also be observed in radial number density plots (Fig. 4.5) for oxygen and nitrogen for clusters of 40 water and 40 ammonia/amine molecules. These results indicate that the second molecule type with less methyl groups is more likely to penetrate the cluster and co-nucleate with water.

According to Figs. 4.4 and 4.5, it is reasonable to analyze the stability and surface free energy of the cluster by calculating surface tensions of pure ammonia/amine. Using Eqs. (4.1) and (4.3) for classical nucleation theory, we can define $\delta\Delta G$ as

$$\delta\Delta G(n) = \Delta G(n) - \Delta G(n-1) = \left(\frac{36\pi}{\rho_l^2}\right)^{\frac{1}{3}} \sigma \left(n^{\frac{2}{3}} - (n-1)^{\frac{2}{3}}\right) - \Delta\mu \qquad (4.10)$$

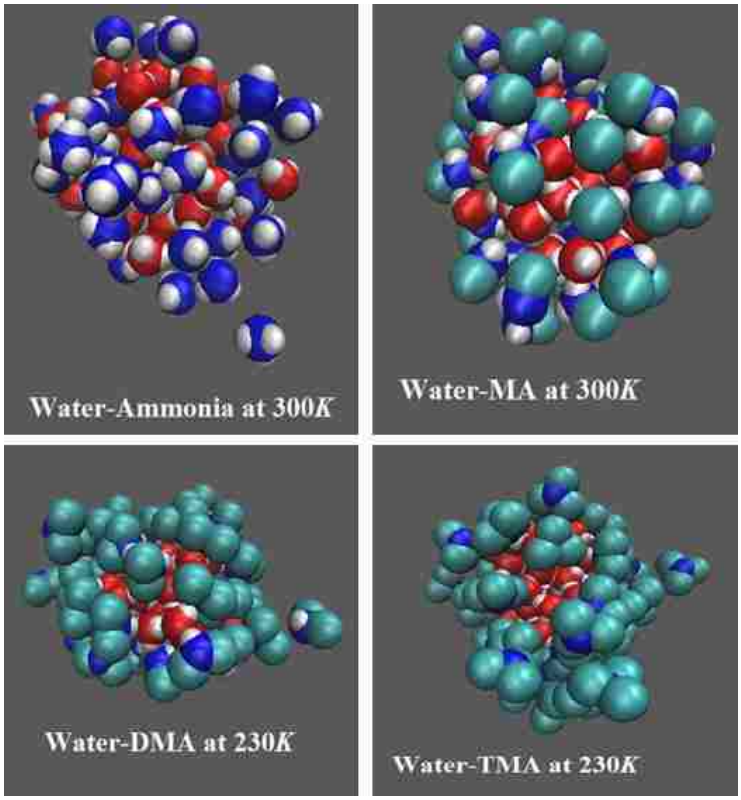

Fig. 4.4. Sample snapshots of binary clusters.

So, the surface tension can be calculated from the slope of $\delta\Delta G$ vs. $\left(n^{\frac{2}{3}} - (n-1)^{\frac{2}{3}}\right)$ plot. Fig. 4.6 presents $\delta\Delta G$ plots of homogeneous nucleation for different molecule types. At small cluster sizes (i.e., large $\left(n^{\frac{2}{3}} - (n-1)^{\frac{2}{3}}\right)$), there is a negative deviation from the CNT prediction due to entropic effects[120] which prevent small clusters from constructing spherical shapes. However, at large cluster sizes (i.e., small $\left(n^{\frac{2}{3}} - (n-1)^{\frac{2}{3}}\right)$), there is a linear behavior as predicted by CNT where the slope is used to calculate surface tensions. Table 4.3 presents surface tensions calculated from MC simulation of homogeneous nucleation. At 300K, the surface tension of water is greater than other component which justifies the presence of water inside binary

clusters. Comparing ammonia and MA, the difference of their surface tensions is greater at 230K which causes MA to form a more stable cluster.

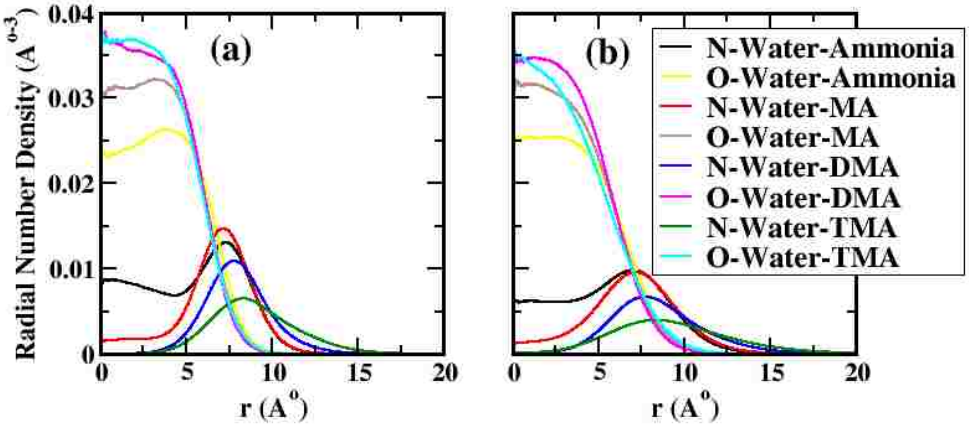

Fig. 4.5. Radial number density for oxygen and nitrogen for clusters of 40 water and 40 ammonia/amine molecules at (a) 230K and (b) 300K.

Finally, using the Jacobian-Gaussian for amines makes MC simulation 4-5 times faster compare to uniform trial generation.

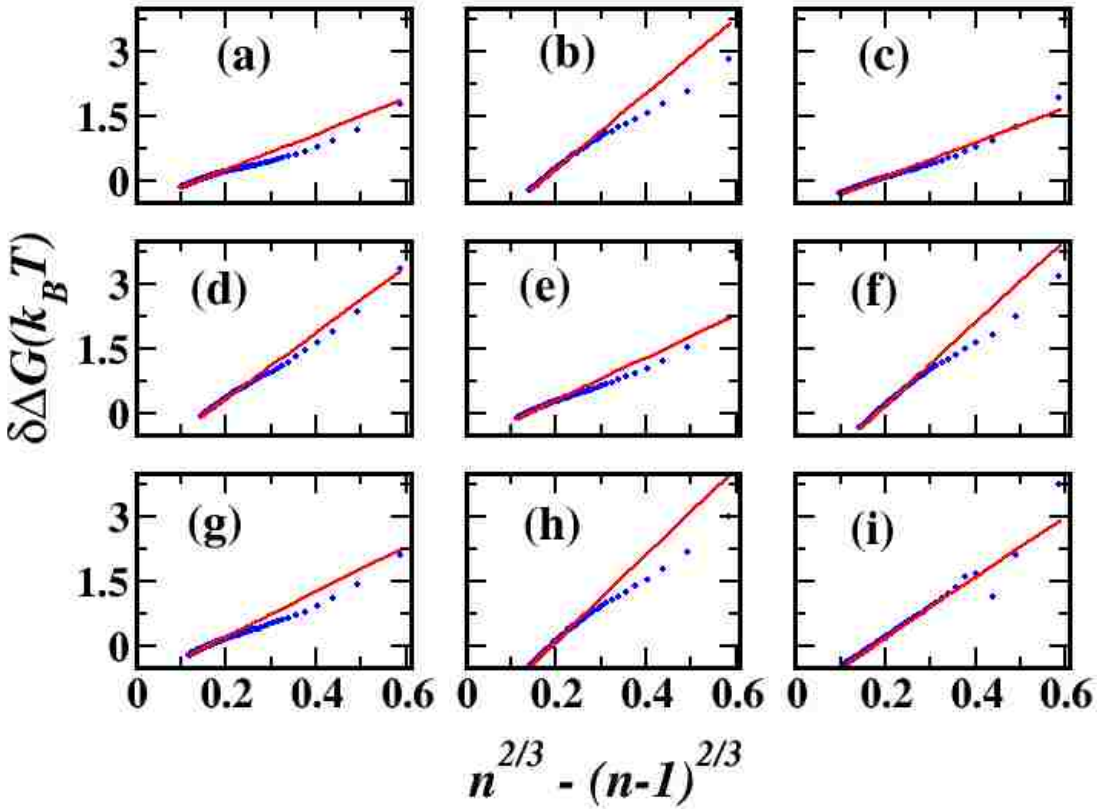

Fig. 4.6. Comparison of simulation results (blue dots) and CNT (red lines) for $\delta\Delta G$ plots of homogeneous nucleation for (a) ammonia at 300K and $\rho = 4 \times 10^{-4}$ Å^{-3}, (b) ammonia at 230K and $\rho = 7 \times 10^{-5}$ Å^{-3}, (c) MA at 300K and $\rho = 3 \times 10^{-4}$ Å^{-3}, (d) MA at 230K and $\rho = 3 \times 10^{-5}$ Å^{-3}, (e) DMA at 300K and $\rho = 1.7 \times 10^{-4}$ Å^{-3}, (f) DMA at 230K and $\rho = 3 \times 10^{-5}$ Å^{-3}, (g) TMA at 300K and $\rho = 1.5 \times 10^{-4}$ Å^{-3}, (h) TMA at 230K and $\rho = 4 \times 10^{-5}$ Å^{-3}, and (i) water at 300K and $\rho = 4 \times 10^{-6}$ Å^{-3}.

Table 4.3. Surface tensions of different compounds at high and low temperatures.

Compound	$T(K)$	σ (dyne/cm)
Ammonia	230	48.7
	300	27.0
MA	230	29.7
	300	18.5
DMA	230	29.0
	300	17.7
TMA	230	25.2
	300	15.5
Water	300	61.1

CHAPTER 5. CONCLUSIONS

In this dissertation, we have developed new methods to improve the efficiency of configurational-bias Monte Carlo for sampling molecular conformations. These methods showed to be superior to previous approaches in sampling complicated molecules such as branched, polymeric, and cyclic molecules. The Jacobian-Gaussian method has been examined in amine nucleation simulation to increase simulation speed. These methods are hoped to be used in simulating complex molecules, such as polypeptides, polypeptoids, polynucleotides, etc., to sample their conformational spaces to study their physical and mechanical properties.

REFERENCES

1. Rivetti, C.; Walker, C.; Bustamante, C., Polymer chain statistics and conformational analysis of DNA molecules with bends or sections of different flexibility11Edited by D. Draper. *J. Mol. Biol.* **1998,** *280* (1), 41-59.

2. Tamai, Y.; Konishi, T.; Einaga, Y.; Fujii, M.; Yamakawa, H., Mean-square radius of gyration of oligo- and poly(methyl methacrylate)s in dilute solutions. *Macromolecules* **1990,** *23* (18), 4067-4075.

3. Mark, J. E.; Flory, P. J., Dipole Moments of Chain Molecules. I. Oligomers and Polymers of Oxyethylene. *J. Am. Chem. Soc.* **1966,** *88* (16), 3702-3707.

4. Lim, K. C.; Heeger, A. J., Spectroscopic and light scattering studies of the conformational (rod-to-coil) transition of poly(diacetylene) in solution. *J. Chem. Phys.* **1985,** *82* (1), 522-530.

5. Fujiwara, Y.; Flory, P. J., Small-Angle X-Ray Scattering by Polymer Chains in the Submolecular Range. *Macromolecules* **1970,** *3* (3), 288-293.

6. Spiess, H. W., Molecular dynamics of solid polymers as revealed by deuteron NMR. *Colloid Polym. Sci.* **1983,** *261* (3), 193-209.

7. Edberg, R.; Morriss, G. P.; Evans, D. J., Rheology of n-alkanes by nonequilibrium molecular dynamics. *J. Chem. Phys.* **1987,** *86* (8), 4555-4570.

8. Flory, P. J.; Hoeve, C. A. J.; Ciferri, A., Influence of bond angle restrictions on polymer elasticity. *Journal of Polymer Science* **1959,** *34* (127), 337-347.

9. Maier, B.; Rädler, J. O., Conformation and Self-Diffusion of Single DNA Molecules Confined to Two Dimensions. *Phys. Rev. Lett.* **1999,** *82* (9), 1911-1914.

10. Georgescu, R. E.; Alexov, E. G.; Gunner, M. R., Combining Conformational Flexibility and Continuum Electrostatics for Calculating pKas in Proteins. *Biophys. J.* **2002,** *83* (4), 1731-1748.

11. Yamakawa, H., On the theory of the second virial coefficient for polymer chains. *Macromolecules* **1992,** *25* (7), 1912-1916.

12. Glabe, C. G., Conformation-dependent antibodies target diseases of protein misfolding. *Trends Biochem. Sci.* **2004,** *29* (10), 542-547.

13. Frenkel, D.; Smit, B., *Understanding Molecular Simulation*. Academic Press, Inc.: 2001.

14. Jorgensen, W. L.; Tirado-Rives, J., Monte Carlo vs Molecular Dynamics for Conformational Sampling. *The Journal of Physical Chemistry* **1996,** *100* (34), 14508-14513.

15. Grinstead, C. M.; Snell, J. L., *Introduction to Probability*. American Mathematical Society: 1997.

16. Casella, G.; Robert, C. P.; Wells, M. T., Generalized Accept-Reject sampling schemes. In *A Festschrift for Herman Rubin*, DasGupta, A., Ed. Institute of Mathematical Statistics: Beachwood, Ohio, USA, 2004; Vol. Volume 45, pp 342-347.

17. Jaynes, E. T.; Bretthorst, G. L., *Probability Theory: The Logic of Science*. Cambridge University Press: 2003.

18. Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E., Equation of State Calculations by Fast Computing Machines. *J. Chem. Phys.* **1953,** *21* (6), 1087-1092.

19. Montroll, E. W., Markoff Chains and Excluded Volume Effect in Polymer Chains. *J. Chem. Phys.* **1950,** *18* (5), 734-743.

20. Orr, W. J. C., Statistical treatment of polymer solutions at infinite dilution. *Transactions of the Faraday Society* **1947,** *43* (0), 12-27.

21. Rosenbluth, M. N.; Rosenbluth, A. W., Monte Carlo Calculation of the Average Extension of Molecular Chains. *J. Chem. Phys.* **1955,** *23* (2), 356-359.

22. Wall, F. T.; Erpenbeck, J. J., New Method for the Statistical Computation of Polymer Dimensions. *J. Chem. Phys.* **1959,** *30* (3), 634-637.

23. Wall, F. T.; Erpenbeck, J. J., Statistical Computation of Radii of Gyration and Mean Internal Dimensions of Polymer Molecules. *J. Chem. Phys.* **1959,** *30* (3), 637-640.

24. Grassberger, P., Pruned-enriched Rosenbluth method: Simulations of θ polymers of chain length up to 1000000. *Physical Review E* **1997,** *56* (3), 3682-3693.

25. Verdier, P. H.; Stockmayer, W. H., Monte Carlo Calculations on the Dynamics of Polymers in Dilute Solution. *J. Chem. Phys.* **1962,** *36* (1), 227-235.

26. Hilhorst, H. J.; Deutch, J. M., Analysis of Monte Carlo results on the kinetics of lattice polymer chains with excluded volume. *J. Chem. Phys.* **1975,** *63* (12), 5153-5161.

27. Wall, F. T.; Mandel, F., Macromolecular dimensions obtained by an efficient Monte Carlo method without sample attrition. *J. Chem. Phys.* **1975,** *63* (11), 4592-4595.

28. MacDonald, B.; Jan, N.; Hunter, D. L.; Steinitz, M. O., Polymer conformations through 'wiggling'. *J. Phys. A: Math. Gen.* **1985,** *18* (13), 2627-2631.

29. Siepmann, J. I., A method for the direct calculation of chemical potentials for dense chain systems. *Mol. Phys.* **1990,** *70* (6), 1145-1158.

30. Widom, B., Some Topics in the Theory of Fluids. *J. Chem. Phys.* **1963,** *39* (11), 2808-2812.

31. Siepmann, J. I.; Frenkel, D., Configurational bias Monte Carlo: a new sampling scheme for flexible chains. *Mol. Phys.* **1992,** *75* (1), 59-70.

32. Frenkel, D.; Mooij, G. C. A. M.; Smit, B., Novel scheme to study structural and thermal properties of continuously deformable molecules. *J. Phys.: Condens. Matter* **1992,** *4* (12), 3053-3076.

33. Siepmann, J. I.; McDonald, I. R., Monte Carlo simulations of mixed monolayers. *Mol. Phys.* **1992,** *75* (2), 255-259.

34. Siepmann, J. I.; McDonald, I. R., Monte Carlo study of the properties of self-assembled monolayers formed by adsorption of CH3(CH2)15SH on the (111) surface of gold. *Mol. Phys.* **1993,** *79* (3), 457-473.

35. de Pablo, J. J.; Laso, M.; Suter, U. W., Simulation of polyethylene above and below the melting point. *J. Chem. Phys.* **1992,** *96* (3), 2395-2403.

36. Vlugt, T. J. H.; Martin, M. G.; Smit, B.; Siepmann, J. I.; Krishna, R., Improving the efficiency of the configurational-bias Monte Carlo algorithm. *Mol. Phys.* **1998,** *94* (4), 727-733.

37. Siepmann, J. I.; Karaborni, S.; Smit, B., Simulating the critical behaviour of complex fluids. *Nature* **1993,** *365* (6444), 330-332.

38. Smit, B.; Karaborni, S.; Siepmann, J. I., Computer simulations of vapor–liquid phase equilibria of n-alkanes. *J. Chem. Phys.* **1995,** *102* (5), 2126-2140.

39. Dijkstra, M., Confined thin films of linear and branched alkanes. *J. Chem. Phys.* **1997,** *107* (8), 3277-3288.

40. Zhuravlev, N. D.; Ilja Siepmann, J., Exploration of the vapour-liquid phase equilibria and critical points of triacontane isomers. *Fluid Phase Equilib.* **1997,** *134* (1–2), 55-61.

41. Cui, S. T.; Cummings, P. T.; Cochran, H. D., Configurational bias Gibbs ensemble Monte Carlo simulation of vapor-liquid equilibria of linear and short-branched alkanes. *Fluid Phase Equilib.* **1997,** *141* (1–2), 45-61.

42. Ilja Siepmann, B. J.; Martin, M. G., Intermolecular potentials for branched alkanes and the vapour-liquid phase equilibria of n-heptane, 2-methylhexane, and 3-ethylpentane. *Mol. Phys.* **1997,** *90* (5), 687-694.

43. Vlugt, T. J. H.; Krishna, R.; Smit, B., Molecular Simulations of Adsorption Isotherms for Linear and Branched Alkanes and Their Mixtures in Silicalite. *The Journal of Physical Chemistry B* **1999,** *103* (7), 1102-1118.

44. Macedonia, M. D.; Maginn, E. J., A biased grand canonical Monte Carlo method for simulating adsorption using all-atom and branched united atom models. *Mol. Phys.* **1999,** *96* (9), 1375-1390.

45. Martin, M. G.; Siepmann, J. I., Novel Configurational-Bias Monte Carlo Method for Branched Molecules. Transferable Potentials for Phase Equilibria. 2. United-Atom Description of Branched Alkanes. *The Journal of Physical Chemistry B* **1999,** *103* (21), 4508-4517.

46. Kathmann, S. M.; Schenter, G. K.; Garrett, B. C.; Chen, B.; Siepmann, J. I., Thermodynamics and Kinetics of Nanoclusters Controlling Gas-to-Particle Nucleation. *The Journal of Physical Chemistry C* **2009,** *113* (24), 10354-10370.

47. Panagiotopoulos, A. Z.; Quirke, N.; Stapleton, M.; Tildesley, D. J., Phase equilibria by simulation in the Gibbs ensemble. *Mol. Phys.* **1988,** *63* (4), 527-545.

48. Smit, B.; De Smedt, P.; Frenkel, D., Computer simulations in the Gibbs ensemble. *Mol. Phys.* **1989,** *68* (4), 931-950.

49. Panagiotopoulos, A. Z., Direct determination of phase coexistence properties of fluids by Monte Carlo simulation in a new ensemble. *Mol. Phys.* **2002,** *100* (1), 237-246.

50. Laso, M.; de Pablo, J. J.; Suter, U. W., Simulation of phase equilibria for chain molecules. *J. Chem. Phys.* **1992,** *97* (4), 2817-2819.

51. Siepmann, J. I.; Karaborni, S.; Smit, B., Vapor-liquid equilibria of model alkanes. *J. Am. Chem. Soc.* **1993,** *115* (14), 6454-6455.

52. Escobedo, F. A.; De Pablo, J. J., Simulation and prediction of vapour-liquid equilibria for chain molecules. *Mol. Phys.* **1996,** *87* (2), 347-366.

53. Chen, B.; Siepmann, J. I.; Oh, K. J.; Klein, M. L., Simulating vapor–liquid nucleation of n-alkanes. *J. Chem. Phys.* **2002,** *116* (10), 4317-4329.

54. Martin, M. G.; Frischknecht, A. L., Using arbitrary trial distributions to improve intramolecular sampling in configurational-bias Monte Carlo. *Mol. Phys.* **2006,** *104* (15), 2439-2456.

55. Snurr, R. Q.; Bell, A. T.; Theodorou, D. N., Prediction of adsorption of aromatic hydrocarbons in silicalite from grand canonical Monte Carlo simulations with biased insertions. *The Journal of Physical Chemistry* **1993,** *97* (51), 13742-13752.

56. Martin, M. G.; Biddy, M. J., Monte Carlo molecular simulation predictions for the heat of vaporization of acetone and butyramide. *Fluid Phase Equilib.* **2005,** *236* (1), 53-57.

57. Baumgärtner, A.; Binder, K., Monte Carlo studies on the freely jointed polymer chain with excluded volume interaction. *J. Chem. Phys.* **1979,** *71* (6), 2541-2545.

58. Pertsin, A. J.; Hahn, J.; Grossmann, H. P., Incorporation of bond-length constraints in Monte Carlo simulations of cyclic and linear molecules: Conformational sampling for cyclic alkanes as test systems. *J. Comput. Chem.* **1994,** *15* (10), 1121-1126.

59. Li, X. J.; Chiew, Y. C., Monte Carlo simulation of Lennard-Jones chains. *J. Chem. Phys.* **1994,** *101* (3), 2522-2531.

60. Escobedo, F. A.; Pablo, J. J. d., Extended continuum configurational bias Monte Carlo methods for simulation of flexible molecules. *J. Chem. Phys.* **1995,** *102* (6), 2636-2652.

61. Escobedo, F. A.; Pablo, J. J. d., Monte Carlo simulation of branched and crosslinked polymers. *J. Chem. Phys.* **1996,** *104* (12), 4788-4801.

62. Dijkstra, M.; Frenkel, D.; Hansen, J. P., Phase separation in binary hard-core mixtures. *J. Chem. Phys.* **1994,** *101* (4), 3179-3189.

63. Vendruscolo, M., Modified configurational bias Monte Carlo method for simulation of polymer systems. *J. Chem. Phys.* **1997,** *106* (7), 2970-2976.

64. Sklenar, H.; Wüstner, D.; Rohs, R., Using internal and collective variables in Monte Carlo simulations of nucleic acid structures: Chain breakage/closure algorithm and associated Jacobians. *J. Comput. Chem.* **2006,** *27* (3), 309-315.

65. Dodd, L. R.; Boone, T. D.; Theodorou, D. N., A concerted rotation algorithm for atomistic Monte Carlo simulation of polymer melts and glasses. *Mol. Phys.* **1993,** *78* (4), 961-996.

66. Wu, M. G.; Deem, M. W., Analytical rebridging Monte Carlo: Application to cis/trans isomerization in proline-containing, cyclic peptides. *J. Chem. Phys.* **1999,** *111* (14), 6625-6632.

67. Deem, M. W.; Bader, J. S., A configurational bias Monte Carlo method for linear and cyclic peptides. *Mol. Phys.* **1996,** *87* (6), 1245-1260.

68. Wu, M. G.; Deem, M. W., Efficient Monte Carlo methods for cyclic peptides. *Mol. Phys.* **1999,** *97* (4), 559-580.

69. Uhlherr, A., Monte Carlo Conformational Sampling of the Internal Degrees of Freedom of Chain Molecules. *Macromolecules* **2000,** *33* (4), 1351-1360.

70. Shah, J. K.; Maginn, E. J., A general and efficient Monte Carlo method for sampling intramolecular degrees of freedom of branched and cyclic molecules. *J. Chem. Phys.* **2011,** *135* (13), 134121.

71. Ulmschneider, J. P.; Jorgensen, W. L., Monte Carlo backbone sampling for polypeptides with variable bond angles and dihedral angles using concerted rotations and a Gaussian bias. *J. Chem. Phys.* **2003,** *118* (9), 4261-4271.

72. Chen, Z.; Escobedo, F. A., A configurational-bias approach for the simulation of inner sections of linear and cyclic molecules. *J. Chem. Phys.* **2000,** *113* (24), 11382-11392.

73. Wick, C. D.; Siepmann, J. I., Self-Adapting Fixed-End-Point Configurational-Bias Monte Carlo Method for the Regrowth of Interior Segments of Chain Molecules with Strong Intramolecular Interactions. *Macromolecules* **2000,** *33* (19), 7207-7218.

74. Lee, J.-S.; Wick, C. D.; Stubbs, J. M.; Siepmann *, J. I., Simulating the vapour–liquid equilibria of large cyclic alkanes. *Mol. Phys.* **2005,** *103* (1), 99-104.

75. Stubbs, J. M.; Potoff, J. J.; Siepmann, J. I., Transferable Potentials for Phase Equilibria. 6. United-Atom Description for Ethers, Glycols, Ketones, and Aldehydes. *The Journal of Physical Chemistry B* **2004,** *108* (45), 17596-17605.

76. Chen, B.; Siepmann, J. I., A Novel Monte Carlo Algorithm for Simulating Strongly Associating Fluids: Applications to Water, Hydrogen Fluoride, and Acetic Acid. *The Journal of Physical Chemistry B* **2000,** *104* (36), 8725-8734.

77. Chen, B.; Siepmann, J. I., Improving the Efficiency of the Aggregation−Volume−Bias Monte Carlo Algorithm. *The Journal of Physical Chemistry B* **2001,** *105* (45), 11275-11282.

78. Chen, B.; Siepmann, J. I.; Oh, K. J.; Klein, M. L., Aggregation-volume-bias Monte Carlo simulations of vapor-liquid nucleation barriers for Lennard-Jonesium. *J. Chem. Phys.* **2001,** *115* (23), 10903-10913.

79. Loeffler, T. D.; Chen, B., Surface induced nucleation of a Lennard-Jones system on an implicit surface at sub-freezing temperatures: A comparison with the classical nucleation theory. *J. Chem. Phys.* **2013,** *139* (23), 234707.

80. Kim, H.; Keasler, S. J.; Chen, B., A Nucleation-Based Method to Study Hydrophobic Interactions under Confinement: Enhanced Hydrophobic Association Driven by Energetic Contributions. *The Journal of Physical Chemistry B* **2014,** *118* (24), 6875-6884.

81. Gerstner, T.; Griebel, M., Numerical integration using sparse grids. *Numerical Algorithms* **1998,** *18* (3), 209.

82. Van Oosterom, A.; Strackee, J., The Solid Angle of a Plane Triangle. *Biomedical Engineering, IEEE Transactions on* **1983,** *BME-30* (2), 125-126.

83. Lifson, S.; Warshel, A., Consistent Force Field for Calculations of Conformations, Vibrational Spectra, and Enthalpies of Cycloalkane and n-Alkane Molecules. *J. Chem. Phys.* **1968,** *49* (11), 5116-5129.

84. Martin, M. G.; Siepmann, J. I., Transferable Potentials for Phase Equilibria. 1. United-Atom Description of n-Alkanes. *The Journal of Physical Chemistry B* **1998,** *102* (14), 2569-2577.

85. Chen, B.; Siepmann, J. I., Transferable Potentials for Phase Equilibria. 3. Explicit-Hydrogen Description of Normal Alkanes. *The Journal of Physical Chemistry B* **1999,** *103* (25), 5370-5379.

86. Wick, C. D.; Martin, M. G.; Siepmann, J. I., Transferable Potentials for Phase Equilibria. 4. United-Atom Description of Linear and Branched Alkenes and Alkylbenzenes. *The Journal of Physical Chemistry B* **2000,** *104* (33), 8008-8016.

87. Chen, B.; Potoff, J. J.; Siepmann, J. I., Monte Carlo Calculations for Alcohols and Their Mixtures with Alkanes. Transferable Potentials for Phase Equilibria. 5. United-Atom Description of Primary, Secondary, and Tertiary Alcohols. *The Journal of Physical Chemistry B* **2001,** *105* (15), 3093-3104.

88. Kamath, G.; Cao, F.; Potoff, J. J., An Improved Force Field for the Prediction of the Vapor−Liquid Equilibria for Carboxylic Acids. *The Journal of Physical Chemistry B* **2004,** *108* (37), 14130-14136.

89. Wick, C. D.; Stubbs, J. M.; Rai, N.; Siepmann, J. I., Transferable Potentials for Phase Equilibria. 7. Primary, Secondary, and Tertiary Amines, Nitroalkanes and Nitrobenzene, Nitriles, Amides, Pyridine, and Pyrimidine. *The Journal of Physical Chemistry B* **2005,** *109* (40), 18974-18982.

90. Lubna, N.; Kamath, G.; Potoff, J. J.; Rai, N.; Siepmann, J. I., Transferable Potentials for Phase Equilibria. 8. United-Atom Description for Thiols, Sulfides, Disulfides, and Thiophene. *The Journal of Physical Chemistry B* **2005,** *109* (50), 24100-24107.

91. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A., Development and testing of a general amber force field. *J. Comput. Chem.* **2004,** *25* (9), 1157-1174.

92. Yang, L.; Tan, C.-h.; Hsieh, M.-J.; Wang, J.; Duan, Y.; Cieplak, P.; Caldwell, J.; Kollman, P. A.; Luo, R., New-Generation Amber United-Atom Force Field. *The Journal of Physical Chemistry B* **2006,** *110* (26), 13166-13176.

93. Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J., Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996,** *118* (45), 11225-11236.

94. Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M., CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **1983,** *4* (2), 187-217.

95. MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-

Kuczera, J.; Yin, D.; Karplus, M., All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *The Journal of Physical Chemistry B* **1998,** *102* (18), 3586-3616.

96. Box, G. E. P.; Muller, M. E., A Note on the Generation of Random Normal Deviates. *The Annals of Mathematical Statistics* **1958,** *29* (2), 610-611.

97. Weerasinghe, S.; Smith, P. E., Kirkwood–Buff derived force field for mixtures of acetone and water. *J. Chem. Phys.* **2003,** *118* (23), 10663-10670.

98. Kang, M.; Smith, P. E., A Kirkwood-Buff derived force field for amides. *J. Comput. Chem.* **2006,** *27* (13), 1477-1485.

99. Sun, H., COMPASS: An ab Initio Force-Field Optimized for Condensed-Phase ApplicationsOverview with Details on Alkane and Benzene Compounds. *The Journal of Physical Chemistry B* **1998,** *102* (38), 7338-7364.

100. Bunte, S. W.; Sun, H., Molecular Modeling of Energetic Materials: The Parameterization and Validation of Nitrate Esters in the COMPASS Force Field. *The Journal of Physical Chemistry B* **2000,** *104* (11), 2477-2489.

101. Scott, W. R. P.; Hünenberger, P. H.; Tironi, I. G.; Mark, A. E.; Billeter, S. R.; Fennen, J.; Torda, A. E.; Huber, T.; Krüger, P.; van Gunsteren, W. F., The GROMOS Biomolecular Simulation Program Package. *The Journal of Physical Chemistry A* **1999,** *103* (19), 3596-3607.

102. Jorgensen, W. L.; Madura, J. D.; Swenson, C. J., Optimized intermolecular potential functions for liquid hydrocarbons. *J. Am. Chem. Soc.* **1984,** *106* (22), 6638-6646.

103. Allen, M. P.; Tildesley, D. J., *Computer simulation of liquids*. Clarendon Press: 1989.

104. Nath, S. K.; Escobedo, F. A.; Pablo, J. J. d., On the simulation of vapor–liquid equilibria for alkanes. *J. Chem. Phys.* **1998,** *108* (23), 9905-9911.

105. Keasler, S. J.; Charan, S. M.; Wick, C. D.; Economou, I. G.; Siepmann, J. I., Transferable Potentials for Phase Equilibria–United Atom Description of Five- and Six-Membered Cyclic Alkanes and Ethers. *The Journal of Physical Chemistry B* **2012,** *116* (36), 11234-11246.

106. Bastiansen, O.; Fernholt, L.; Seip, H. M.; Kambara, H.; Kuchitsu, K., Structure of cyclohexane determined by two independent gas electron-diffraction investigations. *J. Mol. Struct.* **1973,** *18* (2), 163-168.

107. Errington, J. R.; Panagiotopoulos, A. Z., New intermolecular potential models for benzene and cyclohexane. *J. Chem. Phys.* **1999,** *111* (21), 9731-9738.

108. Loeffler, T. D.; Sepehri, A.; Chen, B., Improved Monte Carlo Scheme for Efficient Particle Transfer in Heterogeneous Systems in the Grand Canonical Ensemble: Application to Vapor–Liquid Nucleation. *J. Chem. Theory Comput.* **2015,** *11* (9), 4023-4032.

109. Poursaeidesfahani, A.; Torres-Knoop, A.; Dubbeldam, D.; Vlugt, T. J. H., Direct Free Energy Calculation in the Continuous Fractional Component Gibbs Ensemble. *J. Chem. Theory Comput.* **2016,** *12* (4), 1481-1490.

110. Maerzke, K. A.; Schultz, N. E.; Ross, R. B.; Siepmann, J. I., TraPPE-UA Force Field for Acrylates and Monte Carlo Simulations for Their Mixtures with Alkanes and Alcohols. *The Journal of Physical Chemistry B* **2009,** *113* (18), 6415-6425.

111. Martin, M. G.; Thompson, A. P., Industrial property prediction using Towhee and LAMMPS. *Fluid Phase Equilib.* **2004,** *217* (1), 105-110.

112. Blow, D. M.; Chayen, N. E.; Lloyd, L. F.; Saridakis, E., Control of nucleation of protein crystals. *Protein Science : A Publication of the Protein Society* **1994,** *3* (10), 1638-1643.

113. Li, J.; Lee, T.-C., Bacterial ice nucleation and its potential application in the food industry. *Trends Food Sci. Technol.* **1995,** *6* (8), 259-265.

114. Kulmala, M.; Riipinen, I.; Sipilä, M.; Manninen, H. E.; Petäjä, T.; Junninen, H.; Maso, M. D.; Mordas, G.; Mirme, A.; Vana, M.; Hirsikko, A.; Laakso, L.; Harrison, R. M.; Hanson, I.; Leung, C.; Lehtinen, K. E. J.; Kerminen, V.-M., Toward Direct Measurement of Atmospheric Nucleation. *Science* **2007,** *318* (5847), 89-92.

115. Jensen, E. J.; Pfister, L.; Bui, T. P.; Lawson, P.; Baumgardner, D., Ice nucleation and cloud microphysical properties in tropical tropopause layer cirrus. *Atmos. Chem. Phys.* **2010,** *10* (3), 1369-1384.

116. Malhi, Y.; Aragão, L. E. O. C.; Galbraith, D.; Huntingford, C.; Fisher, R.; Zelazowski, P.; Sitch, S.; McSweeney, C.; Meir, P., Exploring the likelihood and mechanism of a climate-change-induced dieback of the Amazon rainforest. *Proceedings of the National Academy of Sciences* **2009,** *106* (49), 20610-20615.

117. Kim, D.; Ramanathan, V., Solar radiation budget and radiative forcing due to aerosols and clouds. *Journal of Geophysical Research: Atmospheres* **2008,** *113* (D2), n/a-n/a.

118. Pöschl, U., Atmospheric Aerosols: Composition, Transformation, Climate and Health Effects. *Angewandte Chemie International Edition* **2005,** *44* (46), 7520-7540.

119. Kaščiev, D., *Nucleation [electronic resource]: basic theory with applications*. Butterworth-Heinemann Limited: 2000.

120. Loeffler, T. D.; Henderson, D. E.; Chen, B., Vapor–liquid nucleation in two dimensions: On the intriguing sign switch of the errors of the classical nucleation theory. *J. Chem. Phys.* **2012,** *137* (19), 194304.

121. Jacob, D. J., *Introduction to Atmospheric Chemistry*. Princeton University Press: 1999.

122. Yu, F.; Luo, G., Modeling of gaseous methylamines in the global atmosphere: impacts of oxidation and aerosol uptake. *Atmos. Chem. Phys.* **2014,** *14* (22), 12455-12464.

123. Kerminen, V. M.; Petäjä, T.; Manninen, H. E.; Paasonen, P.; Nieminen, T.; Sipilä, M.; Junninen, H.; Ehn, M.; Gagné, S.; Laakso, L.; Riipinen, I.; Vehkamäki, H.; Kurten, T.; Ortega, I. K.; Dal Maso, M.; Brus, D.; Hyvärinen, A.; Lihavainen, H.; Leppä, J.; Lehtinen, K. E. J.; Mirme, A.; Mirme, S.; Hõrrak, U.; Berndt, T.; Stratmann, F.; Birmili, W.; Wiedensohler, A.; Metzger, A.; Dommen, J.; Baltensperger, U.; Kiendler-Scharr, A.; Mentel, T. F.; Wildt, J.; Winkler, P. M.; Wagner, P. E.; Petzold, A.; Minikin, A.; Plass-Dülmer, C.; Pöschl, U.; Laaksonen, A.; Kulmala, M., Atmospheric nucleation: highlights of the EUCAARI project and future directions. *Atmos. Chem. Phys.* **2010,** *10* (22), 10829-10848.

124. Qiu, C.; Zhang, R., Multiphase chemistry of atmospheric amines. *Phys. Chem. Chem. Phys.* **2013,** *15* (16), 5738-5752.

125. Kirkby, J.; Curtius, J.; Almeida, J.; Dunne, E.; Duplissy, J.; Ehrhart, S.; Franchin, A.; Gagne, S.; Ickes, L.; Kurten, A.; Kupc, A.; Metzger, A.; Riccobono, F.; Rondo, L.; Schobesberger, S.; Tsagkogeorgas, G.; Wimmer, D.; Amorim, A.; Bianchi, F.; Breitenlechner,

M.; David, A.; Dommen, J.; Downard, A.; Ehn, M.; Flagan, R. C.; Haider, S.; Hansel, A.; Hauser, D.; Jud, W.; Junninen, H.; Kreissl, F.; Kvashin, A.; Laaksonen, A.; Lehtipalo, K.; Lima, J.; Lovejoy, E. R.; Makhmutov, V.; Mathot, S.; Mikkila, J.; Minginette, P.; Mogo, S.; Nieminen, T.; Onnela, A.; Pereira, P.; Petaja, T.; Schnitzhofer, R.; Seinfeld, J. H.; Sipila, M.; Stozhkov, Y.; Stratmann, F.; Tome, A.; Vanhanen, J.; Viisanen, Y.; Vrtala, A.; Wagner, P. E.; Walther, H.; Weingartner, E.; Wex, H.; Winkler, P. M.; Carslaw, K. S.; Worsnop, D. R.; Baltensperger, U.; Kulmala, M., Role of sulphuric acid, ammonia and galactic cosmic rays in atmospheric aerosol nucleation. *Nature* **2011,** *476* (7361), 429-433.

126. Kurtén, T.; Loukonen, V.; Vehkamäki, H.; Kulmala, M., Amines are likely to enhance neutral and ion-induced sulfuric acid-water nucleation in the atmosphere more effectively than ammonia. *Atmos. Chem. Phys.* **2008,** *8* (14), 4095-4103.

127. Nadykto, A.; Yu, F.; Jakovleva, M.; Herb, J.; Xu, Y., Amines in the Earth's Atmosphere: A Density Functional Theory Study of the Thermochemistry of Pre-Nucleation Clusters. *Entropy* **2011,** *13* (2), 554-569.

128. Li, S.; Qu, K.; Zhao, H.; Ding, L.; Du, L., Clustering of amines and hydrazines in atmospheric nucleation. *Chem. Phys.* **2016,** *472* (Supplement C), 198-207.

129. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983,** *79* (2), 926-935.

130. Zhang, L.; Siepmann, J. I., Development of the trappe force field for ammonia. *Collect. Czech. Chem. Commun.* **2010,** *75* (5), 577-591.

131. Martin, M. G., MCCCS Towhee: a tool for Monte Carlo molecular simulation. *Mol. Simul.* **2013,** *39* (14-15), 1212-1222.

132. http://towhee.sourceforge.net.

133. Deng, L.-Y., Efficient and portable multiple recursive generators of large order. *ACM Trans. Model. Comput. Simul.* **2005,** *15* (1), 1-13.

134. Panagiotopoulos, A. Z., Direct determination of phase coexistence properties of fluids by Monte Carlo simulation in a new ensemble. *Mol. Phys.* **1987,** *61* (4), 813-826.

135. Mooij, G. C. A. M.; Frenkel, D.; Smit, B., Direct simulation of phase equilibria of chain molecules. *J. Phys.: Condens. Matter* **1992,** *4* (16), L255.

136. Torrie, G. M.; Valleau, J. P., Monte Carlo free energy estimates using non-Boltzmann sampling: Application to the sub-critical Lennard-Jones fluid. *Chem. Phys. Lett.* **1974,** *28* (4), 578-581.

137. Nellas, R. B.; Chen, B., Towards understanding the nucleation mechanism for multi-component systems: an atomistic simulation of the ternary nucleation of water/n-nonane/1-butanol. *Phys. Chem. Chem. Phys.* **2008,** *10* (4), 506-514.

138. Felsing, W. A.; Thomas, A. R., Vapor Pressures and Other Physical Constants of Methylamine and Methylamine Solutions. *Industrial & Engineering Chemistry* **1929,** *21* (12), 1269-1272.

139. Swift, E., The Densities of Some Aliphatic Amines. *J. Am. Chem. Soc.* **1942,** *64* (1), 115-116.

140. McQuarrie, D. A.; Simon, J. D., *Physical Chemistry: A Molecular Approach*. University Science Books: 1997.

141. Lv, S.-S.; Miao, S.-K.; Ma, Y.; Zhang, M.-M.; Wen, Y.; Wang, C.-Y.; Zhu, Y.-P.; Huang, W., Properties and Atmospheric Implication of Methylamine–Sulfuric Acid–Water Clusters. *The Journal of Physical Chemistry A* **2015,** *119* (32), 8657-8666.

APPENDIX: COPYRIGHT INFORMATION

The following three publications have been reused in chapter 1 (section 1.4), chapter 2 and chapter 3.

1. This Agreement between Mr. Aliasghar Sepehri ("You") and AIP Publishing LLC ("AIP Publishing LLC") consists of your license details and the terms and conditions provided by AIP Publishing LLC and Copyright Clearance Center.

- License Number: 4214950133217
- License date: Oct 23, 2017
- Licensed Content Publisher: AIP Publishing LLC
- Licensed Content Publication: Journal of Chemical Physics
- Licensed Content Title: Improving the efficiency of configurational-bias Monte Carlo: A density-guided method for generating bending angle trials for linear and branched molecules
- Licensed Content Author: Aliasghar Sepehri, Troy D. Loeffler, Bin Chen
- Licensed Content Date: Aug 21, 2014
- Licensed Content Volume: 141
- Licensed Content Issue: 7
- Type of Use: Thesis/Dissertation
- Requestor type: Author (original article)
- Format: Print and electronic
- Portion: Excerpt (> 800 words)
- Will you be translating? No

- Title of your thesis / dissertation: Innovative Monte Carlo Methods for Sampling Molecular Conformations

2. Title: Improving the Efficiency of Configurational-Bias Monte Carlo: A Jacobian–Gaussian Scheme for Generating Bending Angle Trials for Linear and Branched Molecules

- Author: Aliasghar Sepehri, Troy D. Loeffler, Bin Chen
- Publication: Journal of Chemical Theory and Computation
- Publisher: American Chemical Society
- Date: Apr 1, 2017

- One-time permission is granted only for the use specified in your request. No additional uses are granted (such as derivative works or other editions). For any other uses, please submit a new request.

3. Title: Improving the Efficiency of Configurational-Bias Monte Carlo: Extension of the Jacobian–Gaussian Scheme to Interior Sections of Cyclic and Polymeric Molecules

- Author: Aliasghar Sepehri, Troy D. Loeffler, Bin Chen
- Publication: Journal of Chemical Theory and Computation
- Publisher: American Chemical Society
- Date: Sep 1, 2017

PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE This type of permission/license, instead of the standard Terms & Conditions, is sent to you because no fee is being charged for your order. Please note the following:

- Permission is granted for your request in both print and electronic formats, and translations.
- If figures and/or tables were requested, they may be adapted or used in part.
- Please print this page for your records and send a copy of it to your publisher/graduate school.
- Appropriate credit for the requested material should be given as follows: "Reprinted (adapted) with permission from (COMPLETE REFERENCE CITATION). Copyright (YEAR) American Chemical Society." Insert appropriate information in place of the capitalized words.

- One-time permission is granted only for the use specified in your request. No additional uses are granted (such as derivative works or other editions). For any other uses, please submit a new request.

VITA

Aliasghar Sepehri graduated from Amirkabir University of Technology (Tehran Polytechnic) with a bachelor's degree, in 2007, and a master's degree, in 2010, in Chemical Engineering. He entered PhD program in August 2012 at Department of Chemistry at Louisiana State University and plans to graduate his doctoral degree with a specialization in computational chemistry in May 2018.