SANDIA REPORT

SAND2005-7991
Unlimited Release
Printed Jan 2006

Final Report: LDRD Project 84269 Supramolecular Structures of Peptide-Wrapped Carbon Nanotubes

Amalie L. Frischknecht, Marcus G. Martin, and Susan L. Rempe

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550

Sandia is a multiprogram laboratory operated by Sandia Corporation,
a Lockheed Martin Company, for the United States Department of Energy's
National Nuclear Security Administration under Contract DE-AC04-94-AL85000.

Approved for public release; further dissemination unlimited.

Sandia National Laboratories

Issued by Sandia National Laboratories, operated for the United States Department of Energy by Sandia Corporation.

NOTICE: This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government, nor any agency thereof, nor any of their employees, nor any of their contractors, subcontractors, or their employees, make any warranty, express or implied, or assume any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represent that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government, any agency thereof, or any of their contractors or subcontractors. The views and opinions expressed herein do not necessarily state or reflect those of the United States Government, any agency thereof, or any of their contractors.

Printed in the United States of America. This report has been reproduced directly from the best available copy.

Available to DOE and DOE contractors from
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone: (865) 576-8401
Facsimile: (865) 576-5728
E-Mail: reports@adonis.osti.gov
Online ordering: http://www.doe.gov/bridge

Available to the public from
U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Rd
Springfield, VA 22161

Telephone: (800) 553-6847
Facsimile: (703) 605-6900
E-Mail: orders@ntis.fedworld.gov
Online ordering: http://www.ntis.gov/ordering.htm

SAND2005-7991
Unlimited Release
Printed Jan 2006

Final Report: LDRD Project 84269 Supramolecular Structures of Peptide-Wrapped Carbon Nanotubes

Amalie L. Frischknecht
Computational Materials Science & Engineering

Marcus G. Martin
Multiscale Computational Materials Methods

Susan L. Rempe
Computational Bioscience

Sandia National Laboratories
P.O. Box 5800
Albuquerque, NM 87185

Abstract

Carbon nanotubes (CNT) are unique nanoscale building blocks for a variety of materials and applications, from nanocomposites, sensors and molecular electronics to drug and vaccine delivery. An important step towards realizing these applications is the ability to controllably self-assemble the nanotubes into larger structures. Recently, amphiphilic peptide helices have been shown to bind to carbon nanotubes and thus solubilize them in water. Furthermore, the peptides then facilitate the assembly of the peptide-wrapped nanotubes into supramolecular, well-aligned fibers. We investigate the role that molecular modeling can play in elucidating the interactions between the peptides and the carbon nanotubes in aqueous solution. Using ab initio methods, we have studied the interactions between water and CNTs. Classical simulations can be used on larger length scales. However, it is difficult to sample in atomistic detail large biomolecules such as the amphiphilic peptide of interest here. Thus, we have explored both new sampling methods using configurational-bias Monte Carlo simulations, and also coarse-grained models for peptides described in the literature. An improved capability to model these inorganic/biopolymer interfaces could be used to generate improved understanding of peptide-nanotube self-assembly, eventually leading to the engineering of new peptides for specific self-assembly goals.

Contents

1 Introduction **8**

1.1 The original goal 8

1.2 Modeling peptide/nanotube interactions 9

1.2.1 Force fields 9

1.2.2 *Ab initio* calculations 10

1.2.3 Previous modeling work 11

1.2.4 Improved atomistic modeling 12

1.2.5 Coarse-grained models 12

2 Improved intramolecular sampling in configurational-bias Monte Carlo **14**

2.1 Introduction 14

2.2 Algorithms 16

2.3 Results 24

3 Conclusion **29**

References **32**

List of Figures

1.1 Geometry optimized ab initio structure of water binding to the outside of a (6,6) CNT fragment terminated wtih hydrogen atoms. 11

2.1 The nano-1 peptide (helix) next to an (8,8) CNT, with two sodium counterions shown in dark blue. The helix is colored according to residue type: yellow is nonpolar, blue basic, red acidic, and green polar. 27

2.2 A close up of the nano-1 peptide next to the CNT. The side chains are now shown, colored by residue. One of the Phe residues (dark purple) is closest to the tube, with the aromatic ring of the Phe nearly flat to the tube. 28

List of Tables

2.1 Part B bending angle region boundaries and means used with the global Gaussian option as a function of the hybridization of the central atom (i_{from}) in a bending angle. Hybrid patterns not listed here utilize a uniform distribution instead of the multiple Gaussian distribution. 20

2.2 Dihedral angle region boundaries and means used with the global Gaussian option as a function of the hybridization of the two central atoms (i_{from} and i_{prev}) in a dihedral angle. The ordering of the two central atoms is not important. Hybrid patterns not listed here utilize a uniform distribution instead of the multiple Gaussian distribution. 21

2.3 Summary of the two coupled-decoupled configurational-bias methods employed in this work. Each step is either decoupled from all subsequent steps, or is coupled to a single subsequent selection step. 25

Chapter 1

Introduction

In this introductory section, we briefly describe the background and original goals for this LDRD project. We then describe the difficulties involved in achieving these goals, and outline the areas we decided to focus on. The results are given in the following chapter.

1.1 The original goal

Since their discovery in 1991, there has been great interest in using carbon nanotubes for a wide variety of applications. Single-wall carbon nanotubes (SWNTs) are unique nanoscale building blocks for a variety of materials and applications, from nanocomposites, sensors and molecular electronics to drug and vaccine delivery. Carbon nanotubes exhibit environment-sensitive electrical properties that range from metallic to semiconducting, depending on tube diameter and helicity. SWNTs also have enormous Youngs moduli, which ideally could lead to light, very strong nanocomposites. However, because they are extremely hydrophobic, nanotubes typically aggregate strongly. They thus form insoluble aggregates that are difficult to assemble. The key challenge for being able to use carbon nanotubes in practical materials is to be able to solubilize them and subsequently to assemble them in a controlled manner.

Up until very recently, carbon nanotubes were solubilized in water by covalently bonding various functional groups to the nanotubes, a difficult process with low yields and which changes the nanotube electrical properties, or by using surfactants, which limits the accessible concentrations. In 2001 O'Connell and coworkers [1] demonstrated that water soluble polymers could wrap around SWNTs and solubilize them. A possibly better solution to the problem of solubilization is to utilize the nanoscale self-assembly properties of biological molecules. More recently, Zheng and coworkers discovered that single-stranded DNA (ss-DNA) binds strongly to SWNTs [2, 3]. To date, these systems have not been shown to lead to further self-assembly of the polymer/SWNT hybrids.

Another set of biological molecules that could conceivably solubilize nanotubes are proteins or peptides. Amphiphilic α-helices are known to promote self-assembly. Dieckmann and coworkers [5] recently designed and synthesized an amphiphilic helical peptide called nano-1 that binds to SWNTs, solubilizes them in water, and furthermore facilitates assembly of the peptide/SWNT complex into fibers. The diameter of the fibers can be controlled by the salt concentration in solution. The group performed some preliminary molecular modeling using energy minimization, which showed that six nano-1 helices were sufficient to surround an individual (8,8) SWNT. No modeling or theory was done to describe the self-assembly of the peptide-wrapped SWNTs.

We had proposed to investigate the mechanism for the peptide-nanotube assembly using molecular modeling. First we had planned to compare results between molecular dynamics simulations and ab initio calculations to verify the accuracy of the force fields used. We were then going to use molecular simulations to determine the optimum geometry for the bound peptides and the charge state of the system. The various contributions to the energy determined from the simulations were to be used as input to a theory of the thermodynamics of self-assembly of the peptide-wrapped nanotubes.

1.2 Modeling peptide/nanotube interactions

The original goals for this project, as outlined above, turned out to be overly ambitious for a one year project. This became apparent after some initial research into the various modeling possibilities. Ideally, one would like to explore the interactions of the peptide with the CNT in atomistic detail. One might anticipate that the peptide would deform in some way by being next to the CNT, and it would be very useful to be able to sample the various possible conformations of the peptide near the CNT and evaluate its lowest energy conformations. Below we outline what would be necessary to perform these kinds of calculations. Our efforts along these lines form the rest of the report.

1.2.1 Force fields

To perform classical atomistic simulations, the first requirement is an adequate force field. Because of the great interest in biological molecules, there are adequate force fields in the literature to describe peptides in water, along with any necessary counterions or salt. However, there are no force fields specifically designed for the interactions of these species with carbon nanotubes. We have thus adapted a force field developed by Werder et al. [4] for the interaction of water with graphite. Lennard-Jones interaction parameters between the graphite carbon atoms and the water oxygens were determined in order to reproduce the experimentally observed value of the contact angle of a water drop on graphite. From Werder et al.'s data we were able to extract appropriate values to fit with the CHARMM water model, finding values for the carbon-carbon LJ parameters of ε_{CC} = 0.07437 kcal/mol and

σ_{CC} = 3.4414 Å. The CNT/water interactions given by these values are somewhat weaker than those used in previous work which found water "wires" inside carbon nanotubes. We have used these parameters in all simulations to date, although we may modify them in the future based on results of quantum calculations (see below).

The next step was to implement the force fields and create initial configurations of the systems. This has been done in the Monte Carlo code Towhee. The code can create starting coordinates for any single-walled carbon nanotube given the chiral vectors (m,n) and the number of unit cells desired. It can also generate starting configurations for biological molecules such as peptides and nucleic acids.

1.2.2 *Ab initio* calculations

Recently, there has been some controversy about the interaction of water with carbon nanotubes. We have calculated the interaction strength for the binding of carbon nanotubes with water in vacuum using ab initio techniques. As mentioned above, there has been considerable excitement over the theoretical prediction that water enters pristine CNTs and forms an ordered chain. Water has been found to either enter or escape the tubes, depending on small changes in 1 force-field parameter describing tube-water interactions. Since the tubes used in experiments are generally contaminated with hydrophilic material, it is unclear how attractive the water-tube interaction actually should be to represent a pristine system. Therefore we investigated the binding energy of water with the interior of a CNT using the best ab initio techniques available to us.

Accounting mainly for the electrostatic interactions only between a (6,6) CNT and a water molecule, we find a modest attractive interaction energy of -3.4 kcal/mol. With such a weakly attractive interaction, water would most likely not enter a carbon nanotube. Since CNTs are hydrophobic, it is important to estimate the effect of weakly attractive dispersion interactions due to long-range electron correlation between water and the tube. The best way to do this using ab initio methods is to apply an expensive perturbation correction to the uncorrelated Hartree-Fock result. This required purchasing and installing a parallel version of the quantum code onto a machine with more than 3 terabytes of disk space. This calculation produced a stronger interaction energy of -7.8 kcal/mol between the tube and water. These calculated interaction strengths are known to be upper bounds. An estimate of the error due to overlap of basis functions (another very large calculation) indicates that our best predicted interaction energy is slightly stronger than the strength of the average hydrogen-bond interaction, namely -6.6 kcal/mol. In future work, we will adapt the classical force fields to give this interaction energy and then examine the behavior of water interior to the carbon nanotubes.

Following the same procedure as above, we have calculated the interaction of water with the exterior of the CNT. The geometry-optimized structure is shown in Fig. 1.1. We obtain almost the identical interaction energy when just electrostatic effects are taken into account, indicating that confinement in a large tube has little effect on interaction energy.

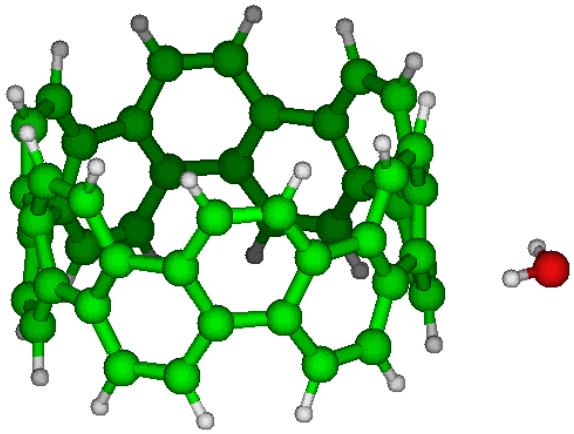

Figure 1.1. Geometry optimized ab initio structure of water binding to the outside of a (6,6) CNT fragment terminated wtih hydrogen atoms.

We expect that dispersion effects significantly strengthen the interaction, just as they did for water inside the tube. This calculation is in progress. In the future, the ab initio binding energies will benchmark classical force field results for the interaction of water with carbon nanotubes.

1.2.3 Previous modeling work

Dieckmann and coworkers [5] did some initial modeling of the interactions of nano-1 with an (8,8) CNT, using energy minimization. Nano-1 forms an amphiphilic α-helix, a secondary protein structure that is stabilized by hydrogen bonds. Amphiphilic helices consist of peptides in which one face of the helix consists of hydrophobic residues and the other face of more polar, hydrophilic residues. These peptides have surfactant-like qualities and spontaneously self-assemble in aqueous solution. The nano-1 sequence is Ac-E(VEAFEKK)(VAAFESK)(VQAFEKK)-(VEAFEHG)-CONH2, where Ac indicates acetylation of the N-terminus, CONH2 indicates amidation of the C-terminus, and the parentheses denote heptad repeats. The peptide was designed to be amphiphilic, with the hydrophobic residues along one side and polar residues along the other side.

Dieckmann *et al.* built up a system of 6 nano-1 helices arranged in a bundle that could surround a CNT. This was done essentially by hand using the commercial modeling programs Insight II and Discover. They put two of these 6-helix bundles together end-to-end, and inserted an (8,8) nanotube into the center. They then added a 5 Å shell of water, and minimized the energy of the system.

From this model building and minimization one obtains some basic geometrical information. It shows that geometrically, 6 nano-1 helices will fit conveniently around the circumference of an (8,8) CNT, and that these 6 helices can be matched up end-to-end with another set of 6 helices further along the axis of the nanotube. This would allow the CNT to be wrapped by peptide along its length. However, this amount of modeling does not prove that this is a stable or preferred arrangement for the peptide around the CNT, only that it is a plausible one. It tells us nothing about the free energy of binding of the peptide to the nanotube, nor the detailed conformations of all the side chains at equilibrium.

1.2.4 Improved atomistic modeling

However, improving on this initial kind of modeling is quite difficult. This is because heteropolymers like a peptide have a large number of possible conformations, which are difficult to sample adequately in a simulation. Furthermore, including the water is of critical importance because the solvent changes all the effective interactions in the system. There are two main simulation techniques one might use on this problem, molecular dynamics (MD) and Monte Carlo (MC). MD simulations are difficult because a macromolecule like nano-1 will diffuse very slowly, especially when surrounded by water. Most of the computation time is used to move the water, rather than the peptide. Also, many conformational changes in peptides or proteins happen on rather slow time scales that are not accessible to MD. MD is thus best suited to asking questions about very local changes in the peptide, starting from an initial set of coordinates obtained from experiment.

MC simulations of large biomolecules are also difficult. In MC one has access to a wider variety of moves that can radically change the conformation of the peptide, such as configurational bias regrowth moves or pivot moves. These work in vacuum but are much less likely to be accepted in a dense solvent such as water. Also, until recently these moves could not adequately sample complex architectures such as the cyclics often found in biomolecules (e.g. in some amino acids such as tryptophan). For this LDRD, we decided to implement some new algorithms for configurational-bias MC (CBMC) that would allow the regrowth of these cyclic pieces of peptides and improve the acceptance rate of CBMC moves. This work is described below in Chap. 2.

1.2.5 Coarse-grained models

Finally, another approach to modeling proteins is to use a coarse-grained model. Many of these have been developed in the literature, and they have been useful in understanding some of the basic physics involved in the problem of protein folding. A recent review of coarse-grained models can be found in Ref 5. As pointed out by Kolinski and Skolnick in this review, there are several choices that must be made in constructing a reduced model of a protein. One must decide on the description of the protein chain itself, whether the model will exist in continuous space or on a lattice, the interactions between the coarse-grained

sites in the protein (the force field), and finally the simulation method. Different decisions are appropriate depending on the question of interest.

The simplest models reduce a peptide to a bead-spring type polymer consisting of two types of beads, representing hydrophobic and polar residues. Originally this model was solved on a lattice, and is termed the HP model [7]. The main weakness of this kind of model is that is cannot produce any secondary structure. Thus these very simple models would not be helpful in modeling nano-1, since this peptide is helical. To model the interactions of nano-1 with a CNT using a reduced model, it would probably be necessary to use an off-lattice simulation with a relatively detailed force field that could mimic the interactions of the side chains with the CNT at some level. Ideally such a model would include the solvent in an implicit way so that efficient sampling of the peptide conformations could be done using e.g. Monte Carlo simulations. A variety of such models have been described in the literature. Fairly coarse models that treat each residue with two united atoms have been studied by Klimov and Thirumalai [8]. A relatively atomistic model that has been used to model the conformations of polypeptides in solution and near surfaces such as semiconductors is a rotational isomeric states (RIS) model of Lustig and coworkers [9, 10]. They kept bond lengths and angles fixed, and sampled continuous torsion angles using a Monte Carlo technique. Future work on the interactions of nano-1 or similar peptides with CNTs would benefit from using these kinds of reduced protein models.

Chapter 2

Improved intramolecular sampling in configurational-bias Monte Carlo

In this chapter, a new formulation of configurational-bias Monte Carlo that uses arbitrary distributions to generate trial bond lengths, angles, and dihedrals is described and shown to provide similar acceptance rates with substantially less computational effort. Several different arbitrary trial distributions are studied and a linear combination of the ideal distribution plus Gaussian distributions automatically fit to the energetic and ideal terms is found to give the best results. The use of these arbitrary trial distributions enables a new formulation of coupled-decoupled configurational bias Monte Carlo that has significantly higher acceptance rates for cyclic molecules. The chemical potential measured via a modified Widom insertion is found to be ill defined in the case of a molecule that has flexible bond lengths due to the unbounded probability distribution that describes the distance between any two atoms. We propose a simple standard state that allows the computation of consistent chemical potentials for molecules with flexible bonds. Finally, we demonstrate that the chemical potential via Widom insertion is not computed properly for molecules with coulombic interactions when the number of trials for any of the nonbonded selection steps is greater than one.

2.1 Introduction

Configurational-bias Monte Carlo (CBMC) was originally developed as a method for computing the chemical potential of polyatomic molecules on a densely populated lattice [11]. It was recast as a Monte Carlo move in continuous space [12, 13] and the chemical potential measurement was generalized to continuous potentials [14]. CBMC was then combined [15] with the Gibbs ensemble [16, 17, 18] triggering a research boom for the computation of polyatomic molecule vapor-liquid coexistence curves. The success of the method is due to its ability to find relatively low intermolecular energy conformations when in-

serting or regrowing a molecule in a dense phase. The early work using this algorithm was performed utilizing linear molecules [19, 20, 21, 22, 23], defined as molecules whose atoms are all bonded to no more than two other atoms. The Boltzmann rejection procedure [24] was a popular and proper method for generating trial bending and dihedral angles in linear molecules and this method was originally adopted for simulation of branched molecules [25, 26, 27, 28, 29], defined as a molecule that contains any atoms that are bonded to three or more other atoms. It was subsequently discovered that the Boltzmann rejection technique does not generate the correct distribution of bending angles and torsions for branched molecules [30]. Two new strategies were developed in response to this revelation. The Macedonia and Maginn method [31] performs a presimulation in order to tabulate sets of molecule fragments that have the correct bending angle distribution and these tables are used to generate trials during the true simulation. The coupled-decoupled method [32] instead generates trials according to the ideal distributions and then incorporates the energy terms into the CBMC acceptance rule.

These two methods both overcome the sampling obstacles for branched molecules, but have different strengths and weaknesses. The Macedonia and Maginn method is computationally expedient because much of the work is performed before the actual simulation and the database of fragments can be reused in subsequent simulations at the same temperatures. However, this method cannot be used in conjunction with other intramolecular sampling schemes (such as single atom translation) because the move is only reversible if the current fragment distribution is present in the table of fragments. The computational advantage is also mitigated somewhat for large, heterogeneous molecules as the amount of memory required to store the fragment tables becomes substantial and eventually requires either a large amount of RAM, or file input and output. In contrast, the coupled-decoupled configurational-bias Monte Carlo (CD-CBMC) algorithm uses an insignificant amount of RAM, but requires a substantial number of bond length, bending angle, and dihedral angle trial attempts (in the range of 100 to 1000) in order to achieve a good acceptance rate. A large number of trials is required because the true distribution (including the phase space terms and the energies) is narrow compared to the ideal distribution, and therefore many of the trials have extremely high energies and a correspondingly low acceptance probability. This is the reason that the bond lengths and bending angles are decoupled from the rest of the terms in the original CD-CBMC formulation as otherwise those selection steps become prohibitively expensive.

This chapter describes a new approach to coupled-decoupled CBMC that uses arbitrary distributions (subject to some mathematical constraints) to generate trial bond lengths, bending angles, and dihedral angles. This new method overcomes much of the computational cost of the original CD-CBMC and enables experimentation with alternative coupled-decoupled formulations. In the course of testing this new method we discovered several problems that occur when CBMC is used to compute the chemical potential for molecules with flexible bonds, coulombic interactions, or fixed-endpoint biasing terms. Solutions to some of these chemical potential CBMC problems are proposed and discussed.

2.2 Algorithms

The concept of arbitrary trial distribution CBMC is based upon the energy bias schemes pioneered by Snurr [33] to improve the sampling of adsorption into porous materials. They presented several specific implementations to bias the grand canonical insertion positions and orientations of molecules in porous materials, but their method is generally applicable to any trial generation and selection step during a CBMC move. A particular trial position ($\mathbf{r}_i$) is generated and selected for further growth during step k of the CBMC procedure according to the selection probability:

$$P_{\text{select}}(\mathbf{r}_i) = \frac{p_{\text{trial}}^{\text{arb}}(\mathbf{r}_i) \times \frac{p_{\text{trial}}^{\text{ideal}}(\mathbf{r}_i)}{p_{\text{trial}}^{\text{arb}}(\mathbf{r}_i)} \exp[-\beta u(\mathbf{r}_i)]}{W_k} \tag{2.1}$$

Here $u(\mathbf{r}_i)$ is the energy of the trial position, $p_{\text{trial}}^{\text{arb}}(\mathbf{r}_i)$ is an arbitrary distribution that is used to generate the trial position, $p_{\text{trial}}^{\text{ideal}}(\mathbf{r}_i)$ is the distribution of trials that would be observed for an ideal system (a system where $u(\mathbf{r}_i) = 0$ for all $\mathbf{r}_i$), and W_k is the Rosenbluth weight of step k. This Rosenbluth weight contains the Boltzmann weight and a ratio of the trial distribution probabilities.

$$W_k = \sum_{j=1}^{n_{\text{trial}}^k} \frac{p_{\text{trial}}^{\text{ideal}}(\mathbf{r}_j)}{p_{\text{trial}}^{\text{arb}}(\mathbf{r}_j)} \exp[-\beta u(\mathbf{r}_j)] \tag{2.2}$$

n_{trial}^k is the number of trials that are generated during step k of the CBMC growth procedure. A CBMC growth consisting of n_{step} growth steps is accepted with the following probability:

$$P_{\text{accept}} = \text{Min}\left[1, \frac{\prod_{\text{k}=1}^{n_{\text{step}}} W_k^{\text{new}}}{\prod_{\text{k}=1}^{n_{\text{step}}} W_k^{\text{old}}}\right] \tag{2.3}$$

The standard CBMC procedure is followed [12, 13] where the new Rosenbluth weight is computed by performing the growth and using random numbers to select a site with the appropriate probability during each step while the old Rosenbluth weight is computed by setting the first trial site of each step to the old position and this trial is always selected so that the old conformation of the molecule is retraced.

The ideal distribution (expressed in cartesian coordinates) for the trial generation of any atom is a uniform distribution with a normalization constant of inverse volume. However, it is often useful to transform this distribution from Cartesian coordinates into a different coordinate system that allows the bond vibration, bending angle, and dihedral angles to be considered separately.

The natural coordinate system to use for bond vibration interactions is the bond length. The ideal trial distribution for bond lengths is

$$p_{\text{trial}}^{\text{ideal}}(l) = C_{\text{bond}}^{\text{ideal}} l^2 \tag{2.4}$$

where l is the bond length and $C_{\text{bond}}^{\text{ideal}}$ is the normalization constant. This normalization constant is not well defined in a molecular simulation as the bond length is not truly bounded for the infinite system represented by periodic boundary conditions and therefore $C_{\text{bond}}^{\text{ideal}}$ goes to zero as the maximum bond length goes to infinity. Even for finite systems it is inconvenient to have a bond length distribution that depends upon the box size as this constant will affect the measurement of the chemical potential. We have responded to this problem by setting $C_{\text{bond}}^{\text{ideal}} = 1.0$ in our simulations.

We consider several arbitrary trial distributions for generating bond lengths. The first is a bounded ideal distribution that has a probability density of

$$p_{\text{trial}}^{\text{b.i.}}(l) = \frac{l^2}{(l_{\text{high}}^3 - l_{\text{low}}^3)/3} \tag{2.5}$$

on the range$[l_{\text{low}}, l_{\text{high}}]$ and zero otherwise. Given the variety in the equilibrium values of bond lengths in a simulation we generally express the high and low bond length bounds as a fraction of the equilibrium bond length ($l_{\text{equil.}}$). There is some danger using this trial generation scheme because simulation ceases to satisfy microscopic reversibility if any "old" bond length is ever outside of these bounds. This can happen when CBMC is used in conjunction with other conformation sampling moves (such as single atom translation moves), or when extremely poor conformations for cyclic molecules are generated in a previous CBMC move. Therefore, you want to set the bounds broadly enough that there is essentially zero probability of accepting a bond length outside of the bounds (due to the high energies from the bond length potential). Conversely, there is pressure to set the bounds as narrowly as possible because $P_{\text{trial}}^{\text{ideal}}$ favors long bond lengths and therefore setting a large range for the bounds means you are generating many bond lengths that are infrequently selected due to their high energies. In previous work [34], we found values of $l_{\text{low}} = 0.85\ l_{\text{equil.}}$ and $l_{\text{high}} = 1.15\ l_{\text{equil}}$ to be a reasonable compromise of these competing desires.

We also consider a Gaussian distribution for bond lengths with an arbitrary trial distribution of

$$p_{\text{trial}}^{\text{Gauss}}(l) = \frac{1}{C_{\text{bond}}^{\text{Gauss}} \sigma^{\text{Gauss}} \sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{l - x^{\text{Gauss}}}{\sigma^{\text{Gauss}}}\right)^2\right] \tag{2.6}$$

where x^{Gauss} and σ^{Gauss} are the mean and standard deviation of the Gaussian distribution, and $C_{\text{bond}}^{\text{Gauss}}$ is an integration constant that accounts for the finite range of the Gaussian

distribution when generating bond trials in a simulation.

$$C_{\text{bond}}^{\text{Gauss}} = 0.5\,[\text{erf}(l_{\text{max}}) - \text{erf}(0.0)] \tag{2.7}$$

Here l_{max} should be set to a relatively large value, but less than half of the minimum box dimension that will occur in the simulation. We typically set l_{max} to the value of the non-bonded cutoff. Two variations of the Gaussian distribution are considered in this work. We originally implemented a 'global Gaussian' version [35] where the mean is set to the equilibrium bond length, and a global standard deviation is utilized regardless of the mean. We have subsequently developed an 'autofit Gaussian' version where the mean and standard deviation are both numerically fit to a discretized form of the true distribution for bond lengths, independent of all other interactions.

$$p_{\text{bond}}^{\text{true}}(l) = C_{\text{bond}}^{\text{true}} l^2 \exp\left[-\beta u_{\text{bond}}(l)\right] \tag{2.8}$$

The value of $C_{\text{bond}}^{\text{true}}$ is irrelevant because we are only fitting the mean and standard deviation of this distribution and then accounting for the normalization using Equation 2.7.

As described previously [32], we break the bending angle selection into two steps. Part A generates bending angles relative to the atom we are currently growing from (i_{from}) and another atom that has previously been grown that is bonded to i_{from} (i_{prev}). Part B then rotates the newly generated atoms $i_{\text{grow}}^{\text{x}}$ about the $i_{\text{from}} - i_{\text{prev}}$ axis in a manner similar to dihedral trial generation. These two steps require different growth strategies. The ideal probability density for part A trials is

$$p_{\text{trial}}^{\text{ideal}}(\theta_{i_{\text{prev}},i_{\text{from}},i_{\text{grow}}^{\text{x}}}) = \frac{1}{2}\sin(\theta_{i_{\text{prev}},i_{\text{from}},i_{\text{grow}}^{\text{x}}}) \tag{2.9}$$

on the range [0,π]. The Gaussian distribution is a reasonable choice for generating bending angle trials with an arbitrary trial distribution of

$$p_{\text{trial}}^{\text{Gauss}}(\theta_{i_{\text{prev}},i_{\text{from}},i_{\text{grow}}^{\text{x}}}) = \frac{1}{C_{\text{bendA}}^{\text{Gauss}}\sigma^{\text{Gauss}}\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\theta_{i_{\text{prev}},i_{\text{from}},i_{\text{grow}}^{\text{x}}} - x^{\text{Gauss}}}{\sigma^{\text{Gauss}}}\right)^2\right] \tag{2.10}$$

where x^{Gauss} and σ^{Gauss} are the mean and standard deviation of the Gaussian distribution, and $C_{\text{bendA}}^{\text{Gauss}}$ is an integration constant that accounts for the finite range of the Gaussian distribution when generating bending trials in a simulation.

$$C_{\text{bendA}}^{\text{Gauss}} = 0.5\,[\text{erf}(\pi) - \text{erf}(0.0)] \tag{2.11}$$

Again we use two different procedures to generate the means and standard deviations. The 'global Gaussian' approach used in previous work [35] sets the equilibrium bending angle

as the mean and specifies a global standard deviation. Here we also consider the 'autofit Gaussian' approach where the mean and standard deviation are fit to the true distribution of each angle, independent of all other interactions.

$$p^{\text{true}}_{\text{bendA}}(\theta_{i_{\text{prev}},i_{\text{from}},i^{\text{x}}_{\text{grow}}}) = C^{\text{true}}_{\text{bendA}} \sin(\theta_{i_{\text{prev}},i_{\text{from}},i^{\text{x}}_{\text{grow}}}) \exp\left[-\beta u_{\text{bendA}}(\theta_{i_{\text{prev}},i_{\text{from}},i^{\text{x}}_{\text{grow}}})\right] \quad (2.12)$$

Where we do not need to know the value of the normalization constant ($C^{\text{true}}_{\text{bendA}}$) in order to fit the mean and standard deviation.

The bending B trials have an ideal distribution that is uniform on $[-\pi,\pi]$. In contrast, the true bend B distribution is now more complicated and often contains multiple maxima.

$$p^{\text{true}}_{\text{bendB}}(\phi_{i^{y}_{\text{grow}}} - \phi_{i^{1}_{\text{grow}}}) = \exp\left[-\beta u_{\text{bend}}(\theta_{i^{1}_{\text{grow}},i_{\text{from}},i^{y}_{\text{grow}}})\right] \quad (2.13)$$

Where $\phi_{i^{y}_{\text{grow}}} - \phi_{i^{1}_{\text{grow}}}$ is the *pseudo*-dihedral angle between the i^{y}_{grow} atom grown during this step and the first atom that is grown during this step (i^{1}_{grow}) relative to the $i_{\text{from}} - i_{\text{prev}}$ axis, and $\theta_{i^{1}_{\text{grow}},i_{\text{from}},i^{y}_{\text{grow}}}$ is computed from the bond lengths, bend A angles, and the relative *pseudo*-dihedral angles. The Gaussian distributions we have used for the bond and bend A generation are not generally appropriate to represent a distribution that has more than one maximum. However, we can modify this distribution by breaking the $[-\pi,\pi]$ region into several subregions and then describing each subregion using a Gaussian distribution. The probability density at any point then becomes the product of the probability of selecting a particular region times the probability distribution inside of that region. This modified Gaussian distribution is

$$\begin{aligned} p^{\text{m.g.}}_{\text{bendB}}(\phi) &= p^{\text{reg.}}(\phi_{\text{low}}(j), \phi_{\text{high}}(j)) \\ &\times \frac{\sqrt{2}}{\left[\text{erf}\left(\phi_{\text{high}}(j)\right) - \text{erf}\left(\phi_{\text{low}}(j)\right)\right]\sigma(j)\sqrt{\pi}} \exp\left[-\frac{1}{2}\left(\frac{\phi - x(j)}{\sigma(j)}\right)^2\right] \end{aligned} \quad (2.14)$$

Our "global Gaussian" implementation for the bend B trial selection assigns subregions and means according to the hybridization pattern of the i_{from} atom as described in Table 2.1. The standard deviation is set globally for all of the angle types and subregions in the simulation. The subregions are selected uniformly with a probability density of $1/n^{\text{reg.}}_{\text{bendB}}$. While this procedure works reasonably well, it does require assignment of hybridization for all atoms in the simulation and also a determination of boundary and mean parameters for any new central atom hybridizations that might be encountered in future work. The 'autofit Gaussian' implementation rectifies these problems by automatically determining the maxima and minima in Equation 2.13 under the constraints that the bond lengths and bending A angles are set to their equilibrium values. The minima are used to determine the boundary regions, and then the mean and standard deviation is computed in each of the subregions for use during the simulation. If insufficient extrema are found then a uniform distribution is

Table 2.1. Part B bending angle region boundaries and means used with the global Gaussian option as a function of the hybridization of the central atom (i_{from}) in a bending angle. Hybrid patterns not listed here utilize a uniform distribution instead of the multiple Gaussian distribution.

Hybrid Pattern	$n^{\text{reg.}}_{\text{bendB}}$	j	$\phi_{\text{low}}(j)$	$\phi_{\text{high}}(j)$	$x(j)$
sp^3	2	1	$-\pi$	0.0	$-2\pi/3$
		2	0.0	π	$2\pi/3$
sp^2	1	1	0.0	2π	π
aromatic	1	1	0.0	2π	π

used instead of the multiple Gaussian distribution. In this work the probability of selecting each subregion was uniform with a probability density of $1/n^{\text{reg.}}_{\text{bendB}}$.

The dihedral angle selection is very similar to the bending angle B selection. The ideal distribution is uniform with a probability density of $1/2\pi$. The multiple Gaussian distribution for dihedral trials is

$$\begin{aligned} p^{\text{m.g.}}_{\text{dihed}}(\phi) &= p^{\text{reg.}}(\phi_{\text{low}}(j), \phi_{\text{high}}(j)) \\ &\times \frac{\sqrt{2}}{\left[\text{erf}\left(\phi_{\text{high}}(j)\right) - \text{erf}\left(\phi_{\text{low}}(j)\right)\right]\sigma(j)\sqrt{\pi}} \exp\left[-\frac{1}{2}\left(\frac{\phi - x(j)}{\sigma(j)}\right)^2\right] \end{aligned} \tag{2.15}$$

where the boundaries and means for the 'global Gaussian' implementation are listed in Table 2.2, and the standard deviation is set globally by the user.

Similar to the bending angle B selection, the dihedral 'global Gaussian' requires specification of the hybridization of each atom in the simulation. This is inconvenient to maintain so we have also implemented an 'autofit Gaussian' version that fits multiple Gaussian peaks automatically to the true dihedral distribution

$$p^{\text{true}}_{\text{dihed}}(\phi) \propto \exp\left[-\beta u_{\text{dihed}}(\phi)\right] \tag{2.16}$$

under the constraints that the bond lengths and bending angles are set to their equilibrium values. The minima in this distribution are used to break the 2π region into subregions that are then each fit to a Gaussian distribution. In both the 'global Gaussian' and 'autofit Gaussian' cases the subregions are chosen uniformly with a probability density of $1/n^{reg.}_{\text{dihed}}$.

A Gaussian distribution is analytically positive over its entire range, however it goes numerically to zero in a computer simulation roughly 38σ away from the mean. While this is a large deviation from the mean, it can occur during a molecular simulation if the conformation is extremely poor. This is an unfortunately common occurrence for molecules that

Table 2.2. Dihedral angle region boundaries and means used with the global Gaussian option as a function of the hybridization of the two central atoms (i_{from} and i_{prev}) in a dihedral angle. The ordering of the two central atoms is not important. Hybrid patterns not listed here utilize a uniform distribution instead of the multiple Gaussian distribution.

Hybrid Pattern	$n_{\text{dihed}}^{reg.}$	j	$\phi_{\text{low}}(j)$	$\phi_{\text{high}}(j)$	$x(j)$
$sp^3 - sp^3$	3	1	$-2\pi/3$	0	$-\pi/3$
		2	0	$2\pi/3$	$\pi/3$
		3	$2\pi/3$	$4\pi/3$	π
$sp^2 - sp^2$	4	1	$3\pi/4$	$5\pi/4$	π
		2	$-3\pi/4$	$-\pi/4$	$-\pi/2$
		3	$-\pi/4$	$\pi/4$	0
		4	$\pi/4$	$3\pi/4$	$\pi/2$
$arom - arom$	2	1	$\pi/2$	$3\pi/2$	π
		2	$-\pi/2$	$\pi/2$	0
$sp^3 - sp^2$	6	1	$5\pi/6$	$7\pi/6$	π
		2	$-5\pi/6$	$-3\pi/6$	$-4\pi/6$
		3	$-3\pi/6$	$-\pi/6$	$-2\pi/6$
		4	$-\pi/6$	$\pi/6$	0
		5	$\pi/6$	$3\pi/6$	$2\pi/6$
		6	$3\pi/6$	$5\pi/6$	$4\pi/6$

contain complex cyclic geometries, as the algorithms are not always sufficient to ensure loop closure. An arbitrary trial distribution of zero is fatal to the move as it creates a situation where a move is not reversible, and therefore the CBMC procedure is not provably correct. In order to avoid this problem one can use an arbitrary trial distribution that is a linear combination of other arbitrary trial distributions.

$$p^{\text{l.c.}}(z) = p^{\text{dist.a}}_{\text{select}} \times p^{\text{dist.a}}(z) + (1 - p^{\text{dist.a}}_{\text{select}}) \times p^{\text{dist.b}}(z) \tag{2.17}$$

Here z is the interaction being generated (bond length, bending angle, or dihedral angle), $p^{\text{dist.a}}_{\text{select}}$ is the probability of selecting distribution a, and $p^{\text{dist.a}}(z)$ and $p^{\text{dist.b}}(z)$ are the normalized probability densities for distributions a and b, respectively. Combining an ideal distribution with a Gaussian distribution realizes the benefits of the targeted distributions while removing the numeric zero probability problem.

We utilize additional biasing when regrowing cyclic molecules, or the interiors of large molecules. This fixed-endpoint biasing was originally proposed for lattice polymers [36] and subsequently generalized to continuous space molecules [37]. An additional biasing is employed when we are growing an atom that is on a path that is one ungrown atom (two bonds) or two ungrown atoms (three bonds) away from connecting with an already existing atom ($\mathbf{r}_{\text{exist}}$). This biasing is combined with the arbitrary trial distribution method at an appropriate stage of the coupled-decoupled CBMC growth and Equations 2.1, 2.2, and 2.3 are replaced by

$$P_{\text{select}}(\mathbf{r}_i) = \frac{p^{\text{arb}}_{\text{trial}}(\mathbf{r}_i) \times \frac{p^{\text{ideal}}_{\text{trial}}(\mathbf{r}_i)}{p^{\text{arb}}_{\text{trial}}(\mathbf{r}_i)} \times p_{\text{bias}}(\mathbf{r}_i, \mathbf{r}_{\text{exist}}) \exp[-\beta u(\mathbf{r}_i)]}{W_k} \tag{2.18}$$

$$W_k = \sum_{j=1}^{n^k_{\text{trial}}} \frac{p^{\text{ideal}}_{\text{trial}}(\mathbf{r}_j)}{p^{\text{arb}}_{\text{trial}}(\mathbf{r}_j)} p_{\text{bias}}(\mathbf{r}_j, \mathbf{r}_{\text{exist}}) \exp[-\beta u(\mathbf{r}_j)] \tag{2.19}$$

$$P_{\text{accept}} = \text{Min}\left[1, \frac{\prod_{\text{k}=1}^{n_{\text{step}}} W_k^{\text{new}}}{\prod_{\text{k}=1}^{n_{\text{step}}} W_k^{\text{old}}} \times \frac{\prod_{\text{k}=1}^{n_{\text{step}}} p_{\text{bias}}(\mathbf{r}_s, \mathbf{r}_{\text{exist}})^{\text{old}}}{\prod_{\text{k}=1}^{n_{\text{step}}} p_{\text{bias}}(\mathbf{r}_s, \mathbf{r}_{\text{exist}})^{\text{new}}}\right] \tag{2.20}$$

where $p_{\text{bias}}(\mathbf{r}_i, \mathbf{r}_{\text{exist}})$ is a probability that depends upon the distance between the trial site and all existing atoms in the molecule, and $p_{\text{bias}}(\mathbf{r}_s, \mathbf{r}_{\text{exist}})$ are the additional biasing probabilities for the atoms that were selected during the growth procedure. This method of incorporating an additional biasing and then removing it in the acceptance rate is fundamentally the same as that followed for anisotropic atom growth [38], dual-cutoff [39], and long range corrections like the Ewald sum [40]. In common with those methods, it no longer computes the chemical potential properly unless all of the trial positions (whether selected or not) have the same biasing value.

We consider several options for the two bond biasing, where the atom being grown ($\mathbf{r}_{\mathrm{g}}$) is separated by two bonds that pass through one atom that has not yet been grown ($\mathbf{r}_{\mathrm{x}}$) in order to contact one atom that has already exists ($\mathbf{r}_{\mathrm{e}}$). This is the second to last step in the closing of a ring. In all of the cases mentioned we employed a minimum bias value (10^{-40}) in order to make sure the biasing was always nonzero. The analytic Boltzmann using angles (aBua) two-bond biasing depends upon the distance between the atom being grown and the existing atom ($l_{\mathrm{g,e}}$). If $l_{\mathrm{g,e}}$ is less than the sum of the equilibrium values of the two missing bond lengths ($l_{\mathrm{g,x}}^{\mathrm{equil.}} + l_{\mathrm{x,e}}^{\mathrm{equil.}}$) then

$$p_{\mathrm{bias}}^{\mathrm{aBua}}(\mathbf{r}_i, \mathbf{r}_{\mathrm{exist}}) = \exp(\beta u_{\mathrm{bend}}(\theta_{g,x,e})) \tag{2.21}$$

where $\theta_{g,x,e}$ is described by $l_{\mathrm{g,e}}$, $l_{\mathrm{g,x}}^{\mathrm{equil.}}$, and $l_{\mathrm{x,e}}^{\mathrm{equil.}}$. Otherwise the bias is set to

$$p_{\mathrm{bias}}^{\mathrm{aBua}}(\mathbf{r}_i, \mathbf{r}_{\mathrm{exist}}) = \exp(\beta \left[u_{\mathrm{bend}}(\theta_{g,x,e}) + u_{\mathrm{bond}}(l_{\mathrm{g,x}}) + u_{\mathrm{bond}}(l_{\mathrm{x,e}})\right]) \tag{2.22}$$

where $\theta_{g,x,e} = \pi$, $l_{\mathrm{g,x}} = l_{\mathrm{g,x}}^{\mathrm{equil.}} l_{\mathrm{g,e}}/(l_{\mathrm{g,x}}^{\mathrm{equil.}} + l_{\mathrm{x,e}}^{\mathrm{equil.}})$, and $l_{\mathrm{x,e}} = l_{\mathrm{x,e}}^{\mathrm{equil.}} l_{\mathrm{g,e}}/(l_{\mathrm{g,x}}^{\mathrm{equil.}} + l_{\mathrm{x,e}}^{\mathrm{equil.}})$.

We have previously used [34] an analytic Boltzmann dihedral energy sum (aBdes) as the two-bond bias probability. If $l_{\mathrm{g,e}} > l_{\mathrm{g,x}}^{\mathrm{equil.}} + l_{\mathrm{x,e}}^{\mathrm{equil.}}$ then the biasing is set to the minimum biasing, otherwise

$$p_{\mathrm{bias}}^{\mathrm{aBdes}} = \exp\left(\beta \left[u_{\mathrm{dihed}}(\phi_{\mathrm{f,g,x,e}}(1)) + u_{\mathrm{dihed}}(\phi_{\mathrm{f,g,x,e}}(2))\right]\right) \tag{2.23}$$

where $\phi_{\mathrm{f,g,x,e}}(1)$ and $\phi_{\mathrm{f,g,x,e}}(2)$ are the two possible solutions given the constraints that $l_{\mathrm{g,x}}$, $l_{\mathrm{x,e}}$, and $\theta_{\mathrm{f,g,x}}$ are all set to their equilibrium values.

We also considered two variations of a single Gaussian distribution based upon the $l_{\mathrm{g,e}}$ distance. The first used a mean and standard deviation that was automatically determined by fitting to the true distribution of this distance considering only the bond lengths and angles.

$$p_{\mathrm{fit}}(l_{\mathrm{g,e}}) = l_{\mathrm{g,x}}^2 l_{\mathrm{x,e}}^2 \theta_{\mathrm{g,x,e}} \exp\left(\beta \left[u_{\mathrm{bond}}(l_{\mathrm{g,x}}) + u_{\mathrm{bond}}(l_{\mathrm{x,e}}) + u_{\mathrm{bend}}(\theta_{\mathrm{g,x,e}})\right]\right) \tag{2.24}$$

This computation is done at the beginning of the simulation by sampling the bond lengths uniformly on $[0.5l^{\mathrm{equil.}}, 1.5l^{\mathrm{equil.}}]$ and the bond angle uniformly on $[0, \pi]$. The self-adapting fixed-endpoint method [41] is efficient for refining this Gaussian distribution as only the sum of the observed distances and the sum of the observed distances squared are required to self-adapt the mean and standard deviation for all of the two-bond interactions in the system.

The three-bond biasing procedure is dependent upon the positions of the atom being grown ($\mathbf{r}_{\mathrm{g}}$) in this step, the atom it is growing from ($\mathbf{r}_{\mathrm{f}}$) and the estimated positions of the

missing atoms ($\mathbf{r}_x$ and $\mathbf{r}_y$) that bridge between the growing atom and the existing endpoint atom ($\mathbf{r}_e$). The three-bond biasing probability utilized in previous work [34] is a complicated analytic function dependent upon the estimated maximum and minimum $l_{x,e}$ distance. The maximum ($l_{x,e}^{max}$) and minimum ($l_{x,e}^{min}$) distances are estimated by using the $\mathbf{r}_f$, $\mathbf{r}_g$, and $\mathbf{r}_e$ positions and the equilibrium values for $l_{g,x}$ and $\theta_{f,g,x}$. They are then compared with the equilibrium $l_{x,e}$ distance, computed by using the equilibrium values of $l_{x,y}$, $l_{y,e}$, and $\theta_{x,y,e}$. If $l_{x,e}^{equil.} > l_{x,e}^{max}$ then the bias is a penalty for compressing the x-y-e angle

$$p_{bias}^{a24}(l_{x,e}^{equil.}, l_{x,e}^{min}, l_{x,e}^{max}) = \exp\left(\beta u_{bend}(\theta_{x,y,e}^{compress})\right) \tag{2.25}$$

where $\theta_{x,y,e}^{compress}$ is computed using $l_{x,e}^{max}$, $l_{x,y}^{equil.}$, and $l_{y,e}^{equil.}$. If $l_{x,e}^{min} > l_{x,e}^{equil.}$ then the biasing is an expansion penalty analogous to Equations 2.21 and 2.22. Otherwise the three-bond biasing is set to one.

We also combined the maximum and minimum distance approach with the automatically fit Gaussian from the two-bond bias schemes. $l_{x,e}^{max}$ and $l_{x,e}^{min}$ are computed as before and then used as the integration bounds for the two-bond Gaussian probability.

$$p_{bias}^{ug24}(l_{x,e}^{min}, l_{x,e}^{max}) = \int_{l_{x,e}^{min}}^{l_{x,e}^{max}} \frac{1}{C_{2b}^{Gauss}\sigma_{2b}^{Gauss}\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{l - x_{2b}^{Gauss}}{\sigma_{2b}^{Gauss}}\right)^2\right] dl \tag{2.26}$$

Here x_{2b}^{Gauss}, σ_{2b}^{Gauss}, and C_{2b}^{Gauss} are the mean, standard deviation, and normalization constant for the two-bond Gaussian biasing.

The final three-bond biasing we employed is a single self-adapting Gaussian distribution based upon the $l_{g,e}$ distance. The initial mean and standard deviation are fit to the dihedral distribution, using the equilibrium bond lengths and bending angles. These values are periodically updated based upon a linear combination of the previous settings and the mean and standard deviations observed during the course of the simulation.

Two different coupled-decoupled strategies are compared in this report. The first is the original formulation [32] extended to work for flexible bond lengths [34]. The second is an alternative formulation that couples the bond stretch, bending angles, and dihedral selection steps to a new selection step that occurs immediately before the nonbonded selection. Both methods are summarized in Table 2.3.

2.3 Results

In this section we describe some preliminary evaluations of the new algorithms. There are three criteria to consider. The first, and most important, is that the new algorithms provide

Table 2.3. Summary of the two coupled-decoupled configurational-bias methods employed in this work. Each step is either decoupled from all subsequent steps, or is coupled to a single subsequent selection step.

Selection Step	Original [32, 34]	Coupled to pre-nonbond
Bond length	decoupled	pre-nonbond
Bend angle A	decoupled	pre-nonbond
Bend angle B	decoupled	pre-nonbond
Dihedral angle	nonbond	pre-nonbond
Pre-nonbond	N/A	nonbond
Nonbond	decoupled	decoupled

the same answers as previously developed algorithms. The second, is a rough measure of the computational cost and sampling efficiency of the new algorithms relative to previous algorithms. The third is a measure of the robustness of the algorithms, with a focus on suggesting the optimal settings for a broad spectrum of molecules and force fields. All of the results shown here were generated using version 4.12 of the freely available MCCCS Towhee Monte Carlo simulation program [42].

First, we calculated the distribution of C-C bond lengths for a single molecule of Gromos 43A1 [43] united-atom ethane in a 30Å cubic box observed from 100,000 Monte Carlo moves using one of several variants of ATD-CBMC or the Metropolis [44] single atom translation move. The distribution of bond lengths is clearly independent of the type of Monte Carlo moves employed. The canonical insertion chemical potential of Gromos 43A1 ethane was computed in an empty 30Å box at 300 K. This chemical potential contains all of the energetic terms (intramolecular and intermolecular), but does not contain any number density terms. The NVT insertion chemical potentials using the autofit Gaussian with the default standard deviation [549.499_3K$/k_B$], autofit Gaussian with double the default standard deviation [549.9_2K$/k_B$], r^2 with bounds of (0.0,2.0 $l_{\text{equil.}}$) [549.7_9K$/k_B$] and r^2 with bounds of (0.85 $l_{\text{equil.}}$,1.15 $l_{\text{equil.}}$) [549.5_4K$/k_B$], where the subscripts indicate the standard deviation in the final digits, are all statistically identical. For the acceptance rate as a function of the number of bond trials, we find that the autofit Gaussian using the default standard deviation is already above 99% when using a single trial, while the r^2 with bounds methods still have not quite reached that level when using 10,000 trials. The autofit Gaussian method with the default standard deviation fitted to Equation 2.8 provides the same answers and a higher acceptance rate with a few orders of magnitude less effort.

We explore the sampling algorithms for part A of the bending angle selection by studying a single molecule of TraPPE united atom [45] propane at 300 K in a 30Å box. We compared the C-C-C angle distribution observed by numerically integrating Equation 2.12 to determine $C^{\text{true}}_{\text{bendA}}$, with Monte Carlo simulations utilizing several different bend A trial distributions for 1,000,000 CBMC moves and 100 bending trials. The distributions are in

excellent agreement, with the notable exception of the automatically fit Gaussian using a standard deviation scaled by 0.5 as that method results in poor agreement in the tails of the angle distribution. This illustrates the danger of generating trials from a distribution that is narrow compared to the actual distribution. This problem is eliminated by using a linear combination of the ideal distribution and the automatically fit Gaussian distribution to generate trial distributions for bending angle A. We find that an ideal fraction of 0.001 is sufficient to converge to the correct distribution for the TraPPE-UA propane case studied above, but we consider an ideal fraction of 0.01 a reasonable compromise of acceptance rate and sampling accuracy.

Finally, we used the new algorithm to build the peptide that was the original focus of this work, the nano-1 peptide of Dieckmann and coworkers. To do so, we used a similar structure from the Protein Data Bank (PDB; http://www.rcsb.org/pdb/). The nano-1 peptide is based on a previously designed peptide called coil-V_aL_d, whose crystal structure has been determined [46] and entered into the PDB (PDB id 1COI). The sequence of nano-1 is identical to that of coil-V_aL_d, with the exception that the leucine residues in the "d" position of the heptad of coil-V_aL_d are replaced by phenylalanine residues in nano-1. To build a starting configuration, we downloaded the structure of coil-V_aL_d from the PDB. We created an initial set of coordinates that contained only the backbone units. Configurational-bias MC as implemented here was then used to grow in all the side chains of the peptide. The new algorithms were able to successfully grow all the side chains, including the cyclic side chains of phenylalanine (Phe) and histidine (His) which are both present in nano-1.

We also built an (8,8) CNT using the carbon nanotube builder implemented into Towhee. We placed the nano-1 peptide next to the CNT in vacuum using a dielectric constant of 80 to mimic that of water. We used the Charmm-27 force field for the peptide and CNT-Charmm-27 force field for the peptide/tube interactions as described in Sec. 1.2.1. Because nano-1 carries a net charge of -2, we included two sodium counterions in the simulation. We ran a few simulations, all at a temperature of 300K, and utilizing the new CBMC moves as well as atom translations for the peptide and sodium ions, and molecule center-of-mass translations and rotations for the peptide. We kept the CNT fixed at all times.

In general, the CBMC moves were fairly successful, and were able to regrow the peptide with acceptance rates greater than 2% for fragments up to 12 atoms long. Regrowth for segments up to 7 atoms all occurred with acceptance rates greater than 17%, and we had some regrowth of long fragments up to 28 atoms long in some of the runs. Thus, we are able to sample the side chains of the peptide.

The peptide did associate with the CNT in vacuum, although we did not achieve a configuration similar to that constructed by Dieckmann et al. [5] in the current set of simulations. We ran several simulations in a rather small box of dimensions $40 \times 40 \times 61.525$ Å. In these simulations, the box was too small so that the peptide associated with the tube on one end, and the periodic image of the tube on the other end of the peptide. We then placed the system in a larger box of dimensions $100 \times 100 \times 61.525$ Å and ran another simulation for 50,000 MC cycles. In that time, the peptide did not rotate around to become parallel to the CNT. Nevertheless, in all our simulations we found that the peptide appeared

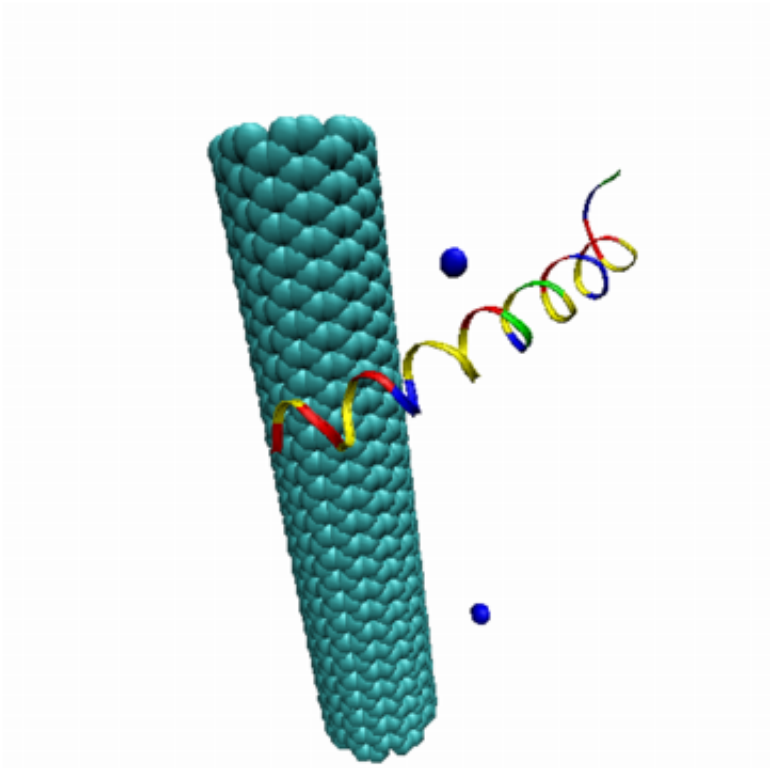

Figure 2.1. The nano-1 peptide (helix) next to an (8,8) CNT, with two sodium counterions shown in dark blue. The helix is colored according to residue type: yellow is nonpolar, blue basic, red acidic, and green polar.

to associate with the CNT through the phenylalanine (Phe) residues, as it was designed to do. In Fig. 2.1 we show a snapshot from the last set of simulations, showing the $\alpha-$helical peptide next to the CNT, along with the two sodium counterions. One end of the peptide appears to be bound to the nanotube. A close-up of the bound region is shown in Fig. 2.2.

Given more time, we could in principle explore further simulations of this system, to attempt to verify the energy-minimized model built by Dieckmann et al. We expect that a judicious initial condition with the peptide aligned along the axis of the tube, followed by CBMC simulations, would very likely find similar configurations as those in Ref. 5. However, we note that including the solvent is extremely important in protein simulations, and in particular we would expect solvent effects to be critical for nano-1 since it is an amphiphilic peptide. Further algorithmic advances will be needed to perform efficient simulations of peptides in explicit water.

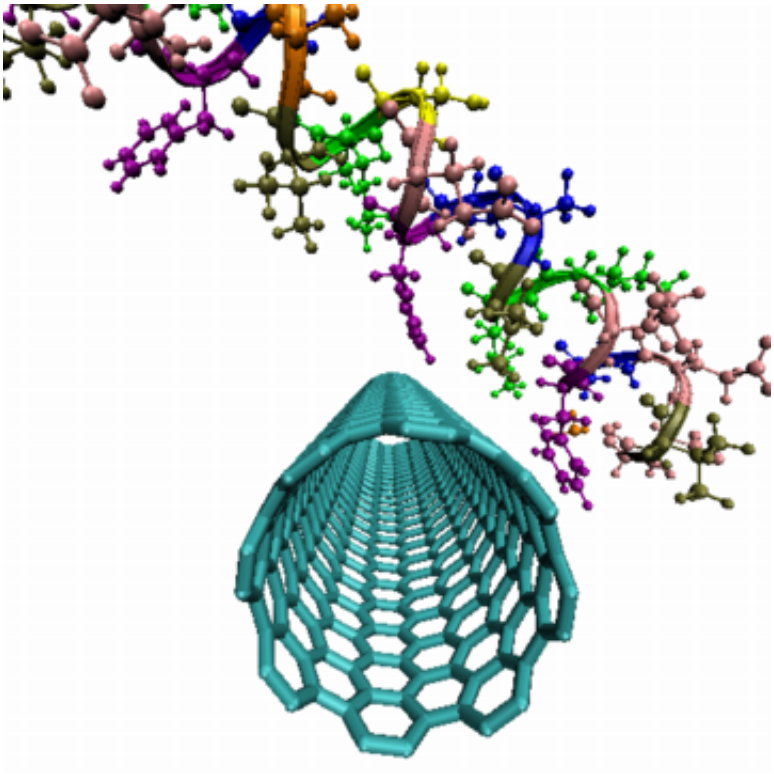

Figure 2.2. A close up of the nano-1 peptide next to the CNT. The side chains are now shown, colored by residue. One of the Phe residues (dark purple) is closest to the tube, with the aromatic ring of the Phe nearly flat to the tube.

Chapter 3

Conclusion

We began by investigating the force fields available in the literature to simulate peptide/CNT interactions, and implemented these in the Monte Carlo code Towhee. Ab initio calculations focused on understanding the interaction between water and CNTs, due to controversy in the field about this interaction. We have calculated the interaction of water with the exterior of the CNT and find a similar value to that for water inside a CNT. Because of the complexity and large size of the nano-1/CNT complexes, a brief survey of the literature on coarse-grained models for peptides was conducted. There have been many different approaches to modeling peptides explored in the literature. Most of these models would require too much time to implement than was available for the current project, but a knowledge of them is useful for future work.

The major accomplishment for the year was the development and implementation of new algorithms to sample peptide conformations. A new formulation of configurational-bias Monte Carlo using arbitrary distributions to generate trial bond lengths, angles, and dihedrals was developed and shown to provide similar acceptance rates with substantially less computational effort. When combined with an alternative formulation of coupled-decoupled configurational bias Monte Carlo, significant improvements in the acceptance rate are obtained for cyclic molecules. This allows the simulation of complex biomolecules such as polypeptides and DNA as these both contain a significant number of cyclic subunits. When tested on the peptide nano-1, we were able to grow the side chains using the new algorithms given an initial configuration for the backbone atoms. Thus, we expect the algorithm to be useful in future simulations of complex molecules such as biomolecules, particularly if combined with a good implicit solvent method.

References

[1] Michael J. O'Connell, Peter Boul, Lars M. Ericson, Chad Huffman, Yuhuang Wang, Erik Haroz, Cynthia Kuper, Jim Tour, Kevin D. Ausman, and Richard E. Smalley. Reversible water-solubilization of single-walled carbon nanotubes by polymer wrapping. *Chem. Phys. Lett.*, 342:265–271, 2001.

[2] M. Zheng, A. Jagota, E.D. Semke, Bruce A. Diner, Robert S. McLean, Steve R. Lustig, Raymond E. Richardson, and Nancy G. Tassi. DNA-assisted dispersion and separation of carbon nanotubes. *Nature Materials*, 2:338–342, 2003.

[3] M. Zheng, A. Jagota, M.S. Strano, A.P. Santos, P.W. Barone, S.G. Chou, Bruce A. Diner, M.S. Dresselhaus, Robert S. McLean, G.B. Onoa, Ge. G. Samsonidze, E.D. Semke, M. Usrey, and D. J. Walls. Structure-based carbon nanotube sorting by sequence-dependent DNA assembly. *Science*, 302:1545–1548, 2003.

[4] Werder, T., Walther, J. H., Jaffe, R.L., Halicioglu, T., and Koumoutsakos, P., *J. Phys. Chem. B*, 107:1345, 2003.

[5] Gregg R. Dieckmann, Alan B. Dalton, Paul A. Johnson, Jian Chen, Geoff M. Giordano, Edgar Muñoz, Inga H. Musselman, Ray H. Baughman, and Rockford K. Draper. Controlled assembly of carbon nanotubes by designed amphiphilic peptide helices. *J. Am. Chem. Soc.*, 125:1770–1777, 2003.

[6] A. Kolinski and J. Skolnick. Reduced models of proteins and their applications. *Polymer*, 45:511-524, 2004.

[7] K. F. Lau and K. A. Dill. *Macromolecules*, 22:3986-3997, 1989.

[8] D. K. Klimov, D. Newfield, and D. Thirumalai. Simulations of β-hairpin folding confined to spherical pores using distributed computing. *Proc. Natl. Acad. Sci. USA*, 99:8019-8024, 2002.

[9] S. R. Lustig and A. Jagota. Selectivity of polypeptide binding to nanoscale substrates. *Mat. Res. Soc. Symp. Proc.*, 724:N4.6.1-6, 2002.

[10] S. Wang, E. S. Humphresy, S.-Y. Chung, et al. Peptides with selective affinity for carbon nanotubes. *Nature Mat.*, 2:196-200, 2003.

[11] Siepmann, J. I., 1990, *Mol. Phys.* , **70**, 1145.

[12] Siepmann, J. I. and Frenkel, D., 1992, *Mol. Phys.* , **75**, 59.

[13] Laso, M., de Pablo, J. J. and Suter, U. W., 1992, *J. Chem. Phys.* , **97**, 2817.

[14] Frenkel, D., Mooij, G. C. A. M. and Smit, B., 1992, *J. Phys. Condens. Matter*, **4**, 3053.

[15] Mooij, G. C. A. M., Frenkel, D. and Smit, B., 1992, *J. Phys. Condens. Matter*, **4**, L255.

[16] Panagiotopoulos, A. Z., 1987, *Mol. Phys.* , **61**, 813.

[17] Panagiotopoulos, A. Z., Quirke, N., Stapleton, M. and Tildesley, D. J., 1988, *Mol. Phys.* **63**, 527.

[18] Smit, B., de Smedt, Ph. and Frenkel, D., 1989, *Mol. Phys.* , **68**, 931.

[19] Siepmann, J. I. and McDonald, I. R., 1992, *Mol. Phys.*, **75**, 255.

[20] Siepmann, J. I., Karaborni, S. and Smit, B., 1993, *Nature*, **365**, 330.

[21] Siepmann, J. I., Karaborni, S. and Smit, B., 1993 *J. Am. Chem. Soc.*, **115**, 6454.

[22] Smit, B., Karaborni, S. and Siepmann, J. I., 1995, *J. Chem. Phys.*, **102**, 2126.

[23] Escobedo, F. A. and de Pablo, J. J., 1996, *Mol. Phys.*, **87**, 347.

[24] Siepmann, J. I. and McDonald, I. R., 1993, *Mol. Phys.*, **79**, 457.

[25] Siepmann, J. I., Martin, M. G., Mundy, C. J. and Klein, M. L., 1997, *Mol. Phys.*, **90**, 687.

[26] Zhuravlev, N. D. and Siepmann, J. I., 1997, *Fluid Phase Equilib.*, **134**, 55.

[27] Vlugt, T. J. H., Martin, M. G., Smit, B., Siepmann, J. I. and Krishna, R., 1998, *Mol. Phys.*, **94**, 727.

[28] Cui, S. T., Cummings, P. T. and Cochran, H. D., 1997, *Fluid Phase Equilib.*, **141**, 45.

[29] Dijkstra, M., 1997, *J. Chem. Phys.*, **107**, 3277.

[30] Vlugt, T. J. H., Krishna, R. and Smit, B., 1999, *J. Chem. Phys.*, **103**, 1102.

[31] Macedonia, M. D. and Maginn, E. J., 1999, *Mol. Phys.*, **96**, 1375.

[32] Martin, M. G. and Siepmann, J. I., 1999, *J. Phys. Chem. B*, **103**, 4508.

[33] Snurr, R. Q., Bell, A. T. and Theodorou, D. N., 1993, *J. Phys. Chem.*, **97**, 13742.

[34] Martin, M. G. and Thompson, A. P., 2004, *Fluid Phase Equil.*, **217**, 105.

[35] Martin, M. G. and Biddy, M. J., 2005, *Fluid Phase Equil.*, **236**, 53.

[36] Dijkstra, M., Frenkel, D. and Hansen, J-P., 1994, *J. Chem. Phys.*, **101**, 3179.

[37] Vendruscolo, M., 1997, *J. Chem. Phys.*, **106**, 2970.

[38] Smit, B., Karaborni, S. and Siepmann, J. I., 1995, *J. Chem. Phys.*, **102**, 2126.

[39] Vlugt, T. J. H., Martin, M. G., Smit, B., Siepmann, J. I. and Krishna, R., 1998, *Mol. Phys.*, **94**, 727.

[40] Ewald, P., 1921, *Ann. Phys.*, **64**, 253.

[41] Wick, C. D. and Siepmann, J. I., 2000, *Macromolecules*, **33**, 7207.

[42] http://towhee.sourceforge.net

[43] Scott, W. R. P., Hunenberger, P. H., Tironi, I. G., Mark, A. E., Billeter, S. R., Fennen, J., Torda, A. E., Huber, T., Kruger, P. and van Gunsteren, W. F., 1999, *J. Phys. Chem. A*, **103**, 3596.

[44] Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E., 1953, *J. Chem. Phys.* **21**, 1087.

[45] Martin, M. G. and Siepmann, J. I., 1998, *J. Phys. Chem. B* **1998**, **102**, 2569.

[46] Ogihara, N. L., Weiss, M. S., Degrado, W. F. and Eisenberg, D., 1997, *Protein Science*, **6**, 80.

DISTRIBUTION:

3 MS 1114
Amalie Frischknecht, 1814

1 MS 1110
Marcus Martin, 1435

1 MS 0310
Susan Rempe, 8333

1 MS 1415
Kevin Leung, 1114

2 MS 9018
Central Technical Files, 8945-1

2 MS 0899
Technical Library, 9616

1 MS 0123
D. Chavez, LDRD Office, 1011