CHARACTERIZATION OF POLYMER-SUPPORTED HOMOGENEOUS CATALYSTS BY MOLECULAR MODELING

A Dissertation
Presented to
The Academic Faculty

by

Andrew T. Swann

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Chemical & Biomolecular Engineering

Georgia Institute of Technology
December 2008

CHARACTERIZATION OF POLYMER-SUPPORTED HOMOGENEOUS CATALYSTS BY MOLECULAR MODELING

Approved by:

Peter J. Ludovice, Advisor
School of Chemical & Biomolecular Engineering
Georgia Institute of Technology

Christopher W. Jones
School of Chemical & Biomolecular Engineering
Georgia Institute of Technology

C. David Sherrill
School of Chemistry & Biochemistry
Georgia Institute of Technology

Martha Grover
School of Chemical & Biomolecular Engineering
Georgia Institute of Technology

Matthew J. Realff
School of Chemical & Biomolecular Engineering
Georgia Institute of Technology

Date Approved: 14 November 2008

The question isn't who is going to let me; it's who is going to stop me.

Ayn Rand, *The Fountainhead*

To Jennifer

for this would not have been possible without her support.

ACKNOWLEDGEMENTS

I would especially like to thank my advisor, Pete Ludovice, for his support throughout my term at Georgia Tech. I would like to acknowledge my committee for their involvement, as well as the Department of Energy for funding. I would like to thank William Sommer, John Sears and Tait Takatani, and John Richardson, all of whom collaborated with me and assisted me in various aspects of this work, mainly through experiments, *ab initio* calculations, and general knowledge, respectively. I thank the Chemical Computing Group, particularly Paul Labute, Alex Cross, and Martin Santavy, for both their software and their assistance with it, without which this work would have taken much longer. I would also thank my parents for their encouragement throughout this work, as well as Jennifer Minnow, for her continuing support. Finally, I thank all of my friends, the academy, and anyone else I may have forgotten (sorry).

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS OR ABBREVIATIONS

ES	Family of thermostats for NVT MD that use an extended system (integral control) of the thermostat to the molecular system. This extended system is described by its own set of coordinates, momentum, and mass.
LFB	Thermostat for NVT MD that uses the Berendsen temperature controller and is integrated by the leapfrog Verlet algorithm.
MC	Monte Carlo.
MD	Molecular Dynamics.
NH	Thermostat for NVT MD that uses the Nosé-Hoover temperature controller and is integrated by an implicit Verlet algorithm.
NH-SV	Thermostat for NVT MD that uses the Nosé-Hoover temperature controller and is integrated by the Störmer-Verlet leapfrog algorithm.
NP	Thermostat for NVT MD that uses the symplectic Nosé-Poincaré temperature controller and is integrated by the generalized leapfrog algorithm.
NPT	Isothermal-isobaric ensemble. Number of particles, temperature, and pressure are all held constant.
NVE	Microcanonical ensemble. Number of particles, system volume, and total system energy are all held constant.
NVT	Canonical ensemble. Number of particles, system volume, and temperature are all held constant.
PCO	Poly(cyclooctene).
PNB	Poly(norbornene).
PNBDC	Poly(N-undecyl-*exo*-norbornene-5,6-dicarboximide).
PTBA	*cis*-t-butyl poly(acetylene).
PTMSP	Poly(1-trimethylsilyl-1-propyne).
ROMP	Ring-opening metathesis polymerization.
TSMC	Torsional States Monte Carlo.
WAXD	Wide Angle X-ray Diffraction.
WC	Family of thermostats for NVT MD that use a weak coupling (proportional control) of the thermostat to the molecular system.

$\boldsymbol{F}$	The force vector of an atom.
$\boldsymbol{a}$	The acceleration vector of an atom.
$\boldsymbol{p}$	Momentum vector of an atom.
$\boldsymbol{v}$	The velocity vector of an atom.
$\boldsymbol{x}$	The position vector of an atom.
Δt	Change in time from t_1 to t_2. Denotes the timestep used in MD algorithms.
η	The coordinate associated with an ES thermostat.
$\langle q \rangle$	The thermal average of the pseudorotational amplitude of cyclopentane.
$\langle r^2 \rangle_0$	Unperturbed mean square end-to-end distance.
$\langle s^2 \rangle_0$	Unperturbed mean square radius of gyration.
$\langle \delta K^2 \rangle$	Mean-squared kinetic energy fluctuations.
Φ	Torsion angle of the central bond of four atoms.
π	The momentum of the thermostat of an extended system MD algorithm.
$\tilde{\boldsymbol{p}}$	Canonically transformed momentum for the NP algorithm.
ξ	The momentum associated with an ES thermostat.
E_{bend}	Angle bend energy between three bonded atoms.
E_{es}	Electrostatic pair interaction energy.
$E_{stretch}$	Bond stretch energy between two bonded atoms.
$E_{torsion}$	Torsion energy between four bonded atoms.
E_{vdw}	van der Waals pair interaction energy.
ee	The enantiomeric excess of one enantiomer over another of a mixture of enantiomers.
K	Kinetic energy.
k_b	Boltzmann constant.
m	The mass of an atom.
Q	The thermal coupling parameter for a molecular dynamics thermostat. This parameter is the mass associated with the thermostat.
q	The pseudorotational amplitude of cyclopentane.

q_e	The mechanical equilibrium (potential energy minimum) value for the pseudorotational amplitude of cyclopentane.
s	The coordinate for the thermostat of an extended system MD algorithm.
t	Time.
$U(\boldsymbol{x})$	The potential energy of the simulation based on the atomic coordinates.
v_{rms}	The pseudorotational velocity of cyclopentane.
C_∞	Limit of the characteristic ratio of a given polymer as the polymer length goes to infinity.
C_n	The characteristic ratio of a given polymer.
P	Pressure.
T	Temperature.
T_g	Glass transition temperature.

SUMMARY

Simulations were used to assist in both the optimization and experimental support of polymer-supported immobilized homogeneous catalysts. This work is a starting point for using molecular modeling to assist in the design of immobilized homogeneous catalysts, where the broader impact is the use of such catalysts which offer high reactivity and selectivity while also providing improved separability and recyclability over heterogeneous catalysts. ROMP poly(norbornene) was examined because it was hypothesized that one of its isomeric configurations might have a helical conformation like vinylic PNB. Alpha shapes were used to determine the accessibility of these polymers with an approximated catalyst group attached to the backbone. The polymer size, reactant size, catalyst size, and linker length were all varied. The simulations were validated by reproducing the expected trends of a random coil for accessibility across the range of the varied properties. Structural analysis of the final conformations showed that these structures were all random coils. It was found that the assumption that the backbone cyclopentane ring was a non-rotatable bond was invalid, which was most likely the largest contributing factor in the lack of a helical structure. It was also found that increasing the size of the virtual catalyst group caused this polymer to have a regions with a local helical conformation. The backbone cyclopentane ring of ROMP PNB was stiffened by adding a dicarboximide group to the ring. The simulation results showed that the TR configuration produced a broad helical conformation. This helix is broad, so its radius of gyration is indistinguishable from that of an equivalent random coil with less than 100 repeat units. Additionally, accessibility did not properly capture this structural difference, but that was mainly because these simulations were pre-optimized for accessibility by having a long linker length and relatively small polymer dimensions. Co(III)salen catalysts were simulated to determine a way to use

simulations to optimize polymer supports for these catalysts. The supports examined were an oligomer synthesized by Jacobsen, poly(cyclooctene) polymerized as a macrocycle, and PCO polymerized as a straight chain polymer. The MMFF94 force field was extended to accommodate cobalt terms based on the ESFF force field, X-ray diffraction structures, and *ab initio* quantum calculations. In order to compare the supports, the individual catalyst efficiency and the overall catalyst efficiency were combined into a "reaction score." The results showed that the PCO macrocycle was the optimal support in the range of 3-5 repeat units, which was consistent with experimental work.

CHAPTER I

INTRODUCTION

1.1 Immobilized Homogeneous Catalysis

Immobilized homogenous catalysis has been present in catalysis research for a number of years as a potential solution to the drawbacks offered by traditional homogeneous and heterogeneous catalysis. These two families of catalysts have been the dominating catalysis systems, but there is a distinct divide between the two. Homogeneous catalysis occurs when the catalyst and reactants are in the same phase. This type of catalysis is broadly characterized by mild reaction conditions, high yields, and high selectivity. Homogeneous catalysts tend to be small organic or organometallic compounds. This type of catalyst would normally be ideal for all reactions, if not for the difficulty of separation of the catalyst from the reaction mixture at the conclusion of the reaction. The separation of these catalysts can often be difficult on a laboratory scale and nearly impossible on the industrial scale without additional and expensive equipment. One possible solution to this problem is to use heterogeneous catalysis, which occurs when the catalyst and reactants are in different phases. Heterogeneous catalysis generally requires harsher reaction conditions (increased temperature and/or pressure) and increased reaction times while producing lower yields and selectivities than its homogeneous counterpart, but the catalysts are much easier to separate from the products. Additionally, these catalysts are more easily regenerated and re-used. These catalysts are generally inorganic and supported on some kind of substrate (e.g., platinum particles supported on carbon nanoparticles for fuel cell catalysts)[1, 2, 3].

For many applications in the mid- to late-twentieth century, heterogeneous catalysts were the catalyst of choice. However, as research progressed in nearly every chemical application, it was discovered time and again that the optimal chemical solutions for a diverse

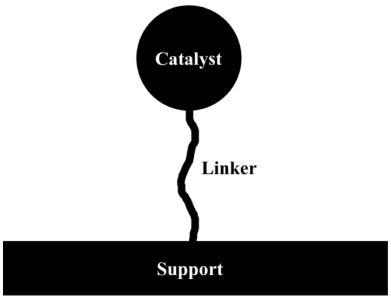


Figure 1: A rough sketch of an immobilized homogeneous catalyst.

application set were to be found in specific enantiomers of individual molecules. Such chemical applications range from pharmaceuticals to consumer products such as perfumes and even paper. This focus on stereospecific chemicals highlights the drawbacks of the two major classifications of catalysis. For enantioselective reactions, homogeneous catalysis would normally be the best choice, except that on the industrial level, added separation steps increase both capital and operating costs. Therefore, heterogeneous catalysts continue to be used simply due to reduced capital cost. One advantage to using either system is that they have both been heavily studied, and they both have well-documented design principles. This means that each type of catalyst can be optimized for the system in which it is used to the full potential of the catalyst[1, 2, 3].

Immobilized homogeneous catalysis is one method that has been used in an attempt to develop a catalyst that combines the reactivity and selectivity of homogeneous catalysts with both the separability and reusability of heterogeneous catalysts. The type of immobilized homogeneous catalysis in this work is commonly called supported or tethered homogeneous catalysis. Essentially, a supported homogeneous catalyst is a homogeneous catalyst which has been attached, or immobilized, on a solid support by a series of atoms that is commonly referred to as either a tether or a linker[4]. A simple diagram of this can be found in Figure 1. Common supports include meso- and microporous silica and alumina[5], as well as various polymers[6]. Unfortunately, immobilized homogeneous catalysts have a history of being expensive to design, mainly because these catalysts have to be

designed specifically for the system in which it is involved. This is largely due to the lack of existing design principles that are broad enough to cover a number of systems. The design principles that would need to be developed would have to cover methods of design from both traditional catalysis categories combined in such a way that would make the design of a new immobilized homogeneous catalysis system cheaper to develop and optimize.

The potential impact of this type of catalysis is great. Every process that currently uses traditional homogeneous catalysis can benefit from the decrease in cost that comes with having fewer and/or simpler separation steps. Traditional heterogeneous catalysis can also benefit by switching to these systems. Higher yields and selectivities mean that there will be a smaller amount of recycled reactants and waste products, especially for products which require high enantiomeric purity, and this will also decrease costs in utilities required for recycling and waste treatment. This has the added benefit of overall decreased energy cost, which is coupled with the decreased energy cost of the lower reaction temperatures and pressures that normally accompany heterogeneous catalysis.

1.2 Polymer Supports

Polymers have previously been used in the application of immobilized homogeneous catalysis and provide many advantages as supports for such systems[4]. Insoluble polymers, such as cross-linked polymers, that are used as catalyst supports tend to behave like heterogeneous catalyst systems. Additionally, specialized techniques are required for characterization and reaction assessment of insoluble polymers. For this reason, soluble polymer supports are a much more attractive support. They tend to behave much more like homogeneous catalyst systems. They are easier to characterize to a reasonable degree using IR and NMR techniques. There are also a number of options in the method of separating the immobilized catalyst from solution. These options include dialysis, filtration, size exclusion chromatography, precipitation, and liquid phase separation[6].

There are a number of types of polymeric chains that can be used as catalyst supports.

One such type includes linear chains with functionalized end groups. Linear chains with a catalyst either on every monomer unit or some other configuration that offers higher loading are also available for use. Hyperbranched polymers and dendrimers can be used as well[7]. However, each polymer type has its potential disadvantages. A polymer with the catalyst only on the end of the chain has a very low loading, and thus requires more chains if a minimum amount of catalyst is required. Linear chains with functionalized monomer units tend to have limited solubility. Dendrimers would seem to be the optimal choice, as they are very easily separable, and there are a number of potential active sites. However, care must be taken to ensure a positive dendritic effect, that is, steric crowding and undesired site-site interactions must be avoided[7]. Additionally, these polymers tend to have a very intricate multi-step synthesis, making them a more expensive option. There are many hyperbranched polymers used as catalyst supports, and they often require only a single step polymerization. However, it is difficult to design such systems. Additionally, the chemical stability of these supports depends on the backbone used, which are typically poly(amines) and poly(esters). These types of polymers may not be stable in certain chemical reaction systems[7].

In the end, there is still not a single polymer that can be generally used as a support. Additionally, each catalyst separation method has its potential limitations, and some methods may not be applicable to specific systems due to the nature of the reactants, products, and/or the interactions thereof. This is one of the main reasons why polymer supported catalysts are not commonly found in industrial applications. Fundamental knowledge of how these systems behave can expedite and simplify the design of such catalytic systems.

1.3 Molecular Mechanics

Molecular mechanics describes a method of molecular modeling which consists of an atomistic mechanical model. Basically, all of the atoms of a system are modeled by balls and springs, where the forces and energies are calculated from this model. A force field is used to describe the collection of force constants associated with each "spring." A typical

diagonal force field has the energy form

$$E_{total} = E_{stretch} + E_{bend} + E_{torsion} + E_{es} + E_{vdw}, \quad (1)$$

where $E_{stretch}$ is the energy associated from a bond stretching between two bonded atoms; E_{bend} is the energy associated from the angle bend between three bonded atoms; $E_{torsion}$ is the energy associated with the torsion angle between four bonded atoms; E_{es} is the energy associated with the nonbonded electrostatic pair interactions; and E_{vdw} is the energy associated with the van der Waals dispersion interactions, though usually bonded atoms do not include pair interactions for the nearest three bonded neighbors. Appendix A shows examples of the force field equations that are used in this work. Essentially, conformation space sampled by a molecule using any method is dictated by the force field, so this is the cornerstone of molecular mechanics.

There are three major methods that sample conformation space. These will be summarized here, though each one has many variants. Energy minimization is used to find the minimum energy conformation that is representative of a static system. Due to the multidimensional nature of polymer conformations, these minimum energy conformations are often local minima. This highlights the importance of the starting conformation, as different starting conformations can achieve different minima if there are a number of local minima in the conformation space. The major minimization algorithms, in increasing order of complexity, are steepest descent, conjugated gradient, and some variant of Newton's method, which are often employed in a sequence to conserve computational resources. All of these algorithms find minima which may or may not be the global minimum. There are methods, such as those employed by Floudas[8], that are designed to find only the global minimum, but these are much less efficient. In evaluating the conformational minima obtained from the starting conformation, methods which find local minima are adequate so long as the conformational potential energy surface is known or multiple starting conformations are used.

Another method for sampling conformation space is molecular dynamics (MD). This is simply integrating Newton's equations of motion ($\sum \mathbf{F} = m_i \boldsymbol{a}$) over time. This method is also dependent upon the starting conformation, however, it is less so, in that any molecular dynamics method samples a particular thermodynamic ensemble that can potentially allow it to pass over energy barriers between local minima. The most common ensembles used for polymeric systems are NVE, NVT, and NPT, where thermostats and barostats are employed to keep temperature and pressure constant.

The third method is called Monte Carlo. A Monte Carlo method uses randomly generated moves of atoms to sample conformation space. Depending on the algorithm used, Monte Carlo can be the least dependent of the three methods on the starting conformation, and it is one of the most flexible methods as far as implementation. There are several different Monte Carlo implementations, and they range from those that can have general applications and those which have been developed for a specific system. For polymers, particularly those in a bulk simulation, Monte Carlo can be very time-consuming compared to energy minimization and MD. This is because polymers have a number of constraints that are present in such a simulation. Separate chains cannot cross, and an individual chain cannot cross parts of itself. This is in addition to the normal internal bond constraints that are present in other bonded molecules. With all of these constraints that are present, Monte Carlo methods had to be developed that performed non-realistic moves that produced realistic conformations with decreased simulation times[9]. One such method is described and used later in this work.

1.4 Molecular Modeling of Polymers

Modeling polymers in some form or another has been done for several decades. Probably the most famous was the original mathematical models of polymers done by Paul Flory, which were developed in the 1930s and 1940s[10, 11]. Since the use of computers in academia, polymers have been modeled by techniques ranging from simple statistical

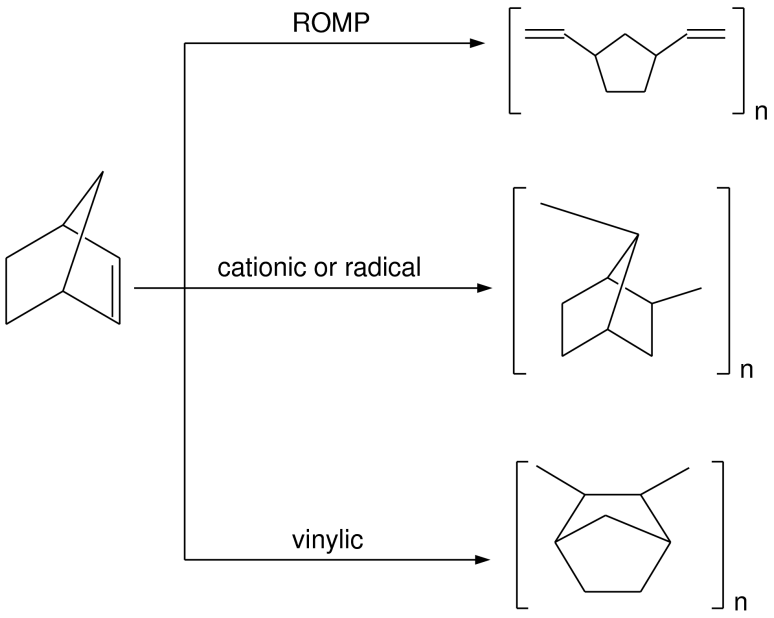


Figure 2: The three polymerization pathways of norbornene.

mesoscopic models to complex quantum models. At the current technology level, small to medium-size polymers, up to 10,000-20,000 total atoms, can be modeled by the appropriate molecular mechanics methods using a reasonable amount of computation time. Even more atoms can be adequately modeled by reducing the complexity of the simulation. This can give a lot of insight into polymer behavior and properties quickly. Generally, modeling polymers has two applications – explaining experimental behavior and designing polymers for specific properties. A good example of both of these applications from previously performed molecular modeling studies is vinylic poly(norbornene).

1.4.1 Vinylic Poly(norbornene)

Bicyclo[2.2.1]hept-2-ene, or poly(norbornene) (PNB), has three possible polymerization pathways, each forming a different polymer, as shown by Figure 2[12]. The first is created via ring-opening metathesis polymerization (ROMP)[13]. The second is created using radical/cationic polymerization[14]. The third is a vinylic polymer created using metal-coordinated[15] or metallocene[16] catalysts. In this case, the vinylic polymer is examined. Its major application is as a photoresist in the microelectronics industry, though it

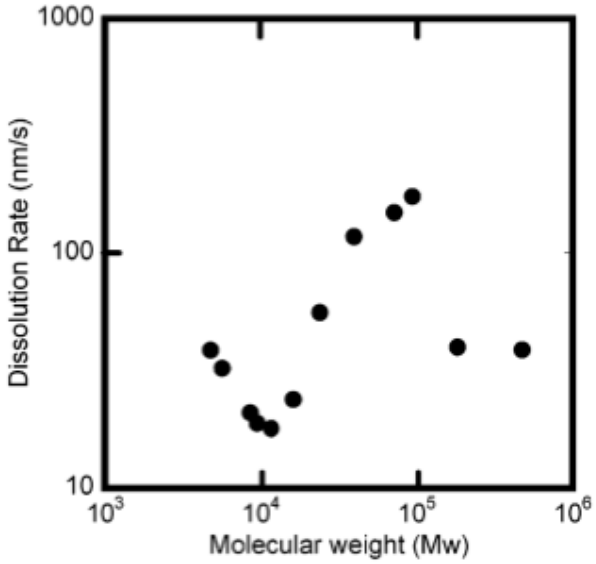
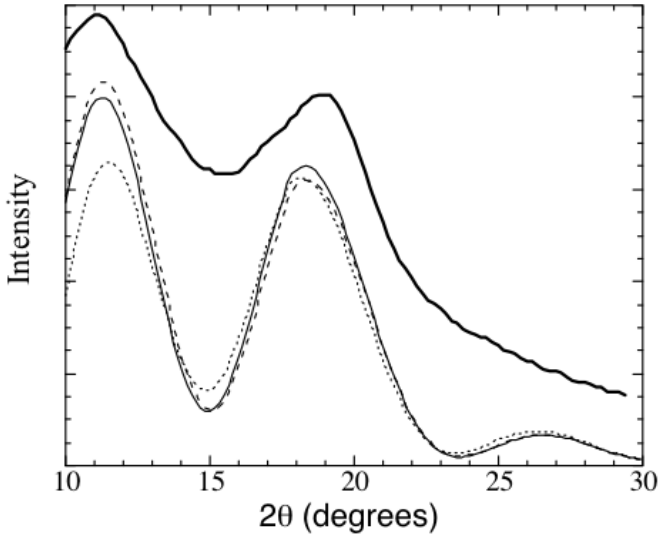


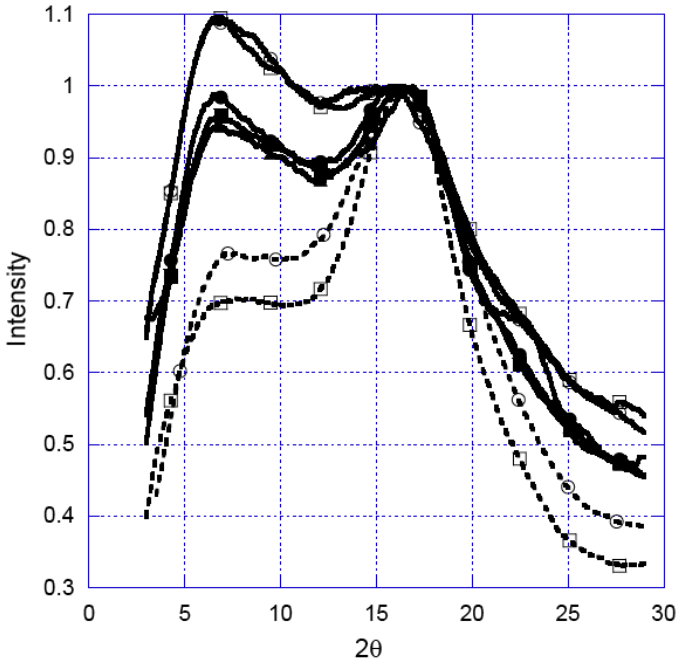
Figure 3: Log-log plot of the dissolution rate of bis(trifluoromethyl)carbinol-substituted poly(norbornene) films[20].

has other applications in the microelectronics industry[17, 18]. The properties that make it a good photoresist are that it has a high glass transition temperature (T_g), transparency to 157 and 193 nm light, and low optical birefringence. It also has a low dielectric constant ($2.2 < \epsilon < 2.4$), which also makes it a good interlayer dielectric[19]. The high T_g is useful because it can go through multiple processing cycles as a photoresist without the polymer structure being compromised. Optical transparency at the exposure wavelength is necessary for any photoresist. However, it has a unique dissolution rate (Figure 3) that cannot be explained from its other properties[20]. Normally, the log of the dissolution rate decreases linearly with the log of increasing molecular weight, though at high molecular weights, the dissolution rate will plateau. Dissolution rate is a key property of photoresists, and it is imperative that its behavior is known for a given photoresist because it is dissolved in the development step of the photolithography process.

Using a Monte Carlo method called Torsional States Monte Carlo (TSMC), it was found that there were repeating backbone angles of ±120°[23]. These backbone torsion angles are periodically interrupted by a "kink." These kinks are found periodically, so the shorter the



(a)



(b)

Figure 4: WAXD spectra for PNB. (a) The dark solid line represents the experimental WAXD pattern for PNB, the dashed line represents a four chain model of PNB, the solid line represents a six chain model of PNB, and the dotted line represents the WAXD pattern for a poly(ethylene)-PNB 50:50 random copolymer[21]. (b) WAXD spectra for MW from 8.5×10^3 (top solid line) to 4.74×10^5 (lowest dotted line) g mol^{-1}[22].

polymer chain, the fewer kinks in the backbone. One method to potentially characterize this type of behavior is wide angle X-ray diffraction (WAXD). The normal diffraction pattern for an amorphous polymer is a broad peak, called an "amorphous halo." For 2,3-exo,exo-erythro-diisotactic PNB, the X-ray diffraction pattern shows two peaks (Figure 4), and molecular simulations were found to produce WAXD spectra which matched that of the experimental data (also shown in Figure 4(a)). This WAXD pattern can be explained with the helix-kink model that was employed creating the simulation data. The model shows a split between the intermolecular (low angle) and intramolecular (high angle) interactions from two distinct patterns of order, and further work has shown that the intermolecular order decreases with increasing MW (Figure 4(b)). This is consistent with a helix-kink model where one might expect small, ordered chains that will increase kink content with larger chains. The increased kink content would appear as a decrease in intermolecular order[21, 23, 22, 20].

The helix-kink model appears to explain the unusual dissolution rate. The dissolution rate initially decreases linearly, which can be explained by the helices with few kinks that can pack tightly enough to make it difficult for solvent to enter the polymer matrix. The dissolution rate then increases, which is caused by the increased number of kinks at higher molecular weights creating solvent accessible pockets. It decreases once again from the polymer chain obtaining a critical size in which the the solvent accessible areas that are created are no longer enough to assist the dissolution. The model also explains the high T_g. The rigid structure afforded by the helices in the backbone decreases the amount of flexibility of the backbone. The packing allowed by the helical conformation of the polymer is such that the chains are more tightly packed than the chains of a random coil. Both of these properties will require a large amount of kinetic energy (high temperature) in order for the polymer chains to move freely.

This example has so far only demonstrated the contribution of modeling to explaining

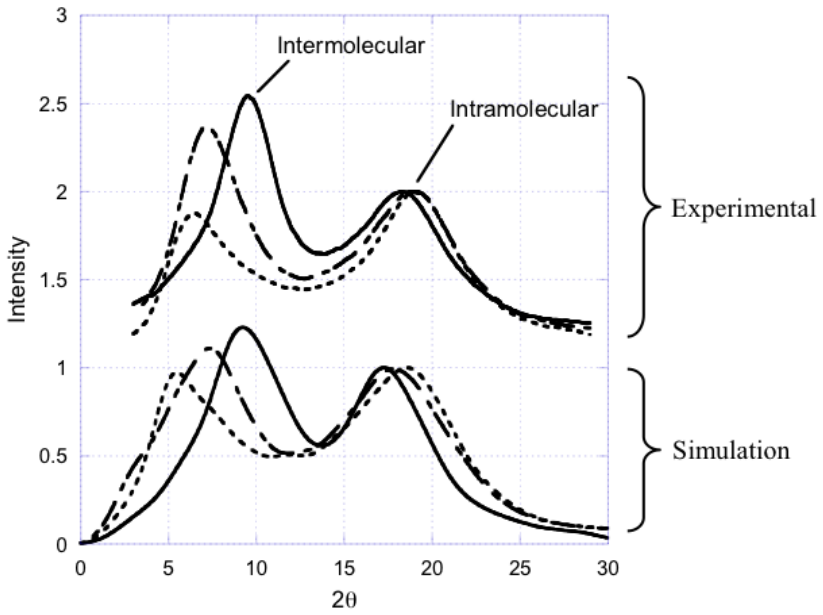


Figure 5: Comparison between simulated and experimental WAXD pattern (shifted for comparison) for vinylic PNB with methyl, butyl, and hexyl side groups on the norbornene ring. The simulated and experimental WAXD pattern for the methyl, butyl, and hexyl PNB are represented by the solid line, short-long dashed line, and dotted line, respectively[24].

experimental behavior. However, modeling has also been used to improve upon the properties of this polymer for its specific application. This was done by determining the effect of different side groups on these same properties of the polymer through modeling. Figure 5 shows a comparison of experimental WAXD data with two 100 repeat unit simulations of vinylic PNB with methyl, butyl, and hexyl side groups. There were two peaks for each polymer, but as the side group increased in size, the peak corresponding to the intermolecular portion of the system decreased in intensity. This is important because it shows that there will be a change in intermolecular order as the side group increases in size due to the less ordered nature of the larger alkane side groups.

1.5 Molecular Modeling of Immobilized Homogeneous Catalysts

1.5.1 Motivation

The lack of concrete design rules or general heuristics for immobilized homogeneous catalysis systems is the biggest drawback to their use. The effect of this is that these systems take longer to optimize because it is generally done in a trial-and-error fashion. Molecular modeling can be a powerful tool in the optimization of immobilized homogeneous catalysis systems. It will not replace experimental work, but it can supplement the experimental work by explaining the results of the experiments, or by decreasing the amount of experimental work by eliminating dead-end optimization routes. Insight supplied by molecular modeling can improve the speed at which these systems are optimized by potentially determining important aspects of the polymer backbone that could improve its use as an immobilized homogeneous catalyst support. The use of molecular modeling might not pinpoint the optimal system, but the insight provided by molecular mechanics simulations can reduce the number of possible variables. For example, simulations can provide reasons why one polymer support might be better than another, or at the very least, they can be used to rank polymer supports so that subsequent experimental work is performed on the highest ranking supports. Additionally, when the proper tools are available, molecular modeling can be faster and use less manpower than equivalent experimental work. There is no raw material use for a simulation, and with computing power being more affordable and readily available, molecular modeling techniques can be used with minimal capital cost. By using simulations to guide experiments, the overall experimental cost would be reduced, which would potentially reduce the overall development cost of such catalysts.

1.5.2 Scope

The overall scope of this work is to apply molecular modeling to potential polymer supports for immobilized homogeneous catalysis. To do this, molecular modeling techniques were examined to determine the best method for analyzing these systems. This includes

the force field, molecular modeling methods, and analysis methods. As polymer supports, two systems were examined. The first is ring opening metathesis polymerized (ROMP) poly(norbornene). This system has been used as a polymer support for immobilized homogeneous catalysts, and it is thought that it might share conformational characteristics with its vinylic cousin. The desired characteristic is that of a highly expanded helix which, in turn, could maximize access to the catalytic side groups of the polymer. The other is poly(cyclooctene), which has been used as a support for a Co-salen catalyst that is used for the hydrolytic kinetic resolution of racemic terminal epoxides. This system was studied because the reaction mechanism requires two catalyst sites to interact with the reactant for a reaction to occur. Ultimately, each system will be used as a basis to show the development and applicability of the molecular modeling tools in the design of immobilized homogeneous catalysts.

1.5.3 Impact

The potential impact from this work includes contributions to the molecular modeling community from findings in algorithm accuracy and modeling considerations, as well as new uses for previously used techniques. From a polymer standpoint, there is new information about polymer conformations and design considerations for property-specific polymer development. In the field of immobilized homogeneous catalysis, the impact will be in the form of what might be necessary to design a polymer support for this type of catalysis. Additionally, this work will be useful as a starting place for modeling other supports. This work will also highlight a method that can be useful in modeling traditional catalytic systems, particularly heterogeneous, as a way to examine accessibility on a molecular level.

CHAPTER II

CYCLOPENTANE

2.1 Introduction

2.1.1 Objectives, Motivation, and Scope

2.1.1.1 Objectives

1. Determine the most accurate force field(s) for the simulation of cyclopentane

2. Determine the appropriate algorithm for a molecular dynamics study of cyclopentane

2.1.1.2 Motivation

The main reason that cyclopentane is of interest is because it is the most dynamic and difficult component of the ROMP poly(norbornene) backbone to simulate. The ROMP PNB backbone consists of a cyclopentane ring alternating with a double bond, the two connected by a single bond (Figure 2). Generic sp2 and sp3 carbons appear in all commonly used force fields, and they all tend to offer similar behavior. However, the specific dynamics of a cyclopentane ring have been observed experimentally and documented, so a force field which properly describes these dynamics would be necessary for modeling ROMP PNB. Because polymer chain dynamics are dominated by the dynamics of their backbones, the dynamics of ROMP PNB will depend heavily on the dynamics of the cyclopentyl groups contained therein. Therefore, accurately simulating the dynamics of these cyclopentyl groups is essential for the determination of the conformational space of the polymer backbone, as a cyclopentane ring will offer a greater degree of flexibility to the backbone than just a typical bond, and this flexibility is primarily due to the pseudorotation of cyclopentane.

Additionally, there are many different algorithms to use for molecular dynamics. Since

cyclopentane was already being studied to determine a proper force field, it makes sense that its use could be extended to examining MD algorithm differences. This is something that has previously been done to a certain extent, but it is typically done with simple Lennard-Jones fluids, and not with real, experimentally verifiable bonded compounds.

2.1.1.3 Scope

This work is a systematic investigation of the effects of components of the simulation on reproducing accurate dynamics of the cyclopentane ring which will presumably increase the accuracy of polymer simulations that contain cyclopentyl and cyclopentyl derivative groups. The force field comparison will be limited to force fields compatible with the Molecular Operating Environment (MOE) software from the Chemical Computing Group. This limits the force fields to ones that are diagonal. The only non-diagonal term that is available in this software is for stretch-bend interactions. This unfortunately excludes Allinger's MM series and Accelrys's PCFF and COMPASS force fields, all of which are considered superior for polymers. However, while the MOE force fields are traditionally used for biological modeling, it will be later shown that they are adequate for polymer modeling.

The algorithm comparison will be limited to commonly used MD algorithms. The Berendsen algorithm is very commonly used with biological systems. The Nosé-Hoover algorithm is the most commonly used extended system algorithm, and Nosé-Poincaré algorithm was created to improve upon Nosé-Hoover. The only other commonly used MD algorithms are stochastic algorithms. These were not examined because the NVT ensemble is not the result of such an algorithm – it is a side-effect. The study of MD algorithms was to be limited to only those which are commonly used to purposefully model in the NVT ensemble.

2.1.2 Background

2.1.2.1 Cyclopentane

It is well known that cyclopentane undergoes pseudorotation between two degenerate low-energy conformations[25]. Both experimental and theoretical results support this, with estimates for the barrier to planarity, the pseudorotational velocity, and the pseudorotational amplitude. Not only is understanding cyclopentane important for polymers like ROMP PNB, understanding the conformational dynamics of cyclopentyl rings is also important for carbohydrates and polynucleotides that contain cyclopentyl derivatives (e.g. ribose) in their backbones[26, 27, 28, 29].

The accurate modeling of cyclopentane dynamics depends on the ability of the simulation to reproduce the metrics that describe the pseudorotation, such as the barrier to planarity and the pseudorotational velocity and amplitude. While the reproduction of the barrier to planarity depends only on the forcefield used in the simulation, the pseudorotational velocity and amplitude depend on the MD algorithm as well. This dependence results from the fact that the pseudorotation is a dynamic process and can be affected by the molecular dynamics algorithm used. If the dynamics algorithm generates different types of fluctuations, the pseudorotational velocity and associated amplitude may be different than the reported experimental or theoretical value. Previous work has been done examining the applicability of the MM3[30] and AMBER[31] force fields to cyclopentane. While the AMBER force field had a better agreement with the barrier to planarity, both force fields showed similar pseudorotational amplitudes and pseudorotational velocities, and both simulations were carried out using the Berendsen temperature control algorithm with similar thermal inertia parameters that determine the coupling of the Berendsen thermostat to the simulated system. This previous work did not examine the role played by the details of the simulation algorithm in the resulting dynamics of cyclopentane. A sensitivity of the cyclopentane dynamics to the simulation algortihm could result in an inaccurate simulation should the wrong algorithm be chosen.

2.1.2.2 Pseudorotational Parameters

Cyclopentane pseudorotates between two degenerate low-energy conformations, the twist (C_2), or half-chair, and envelope (C_s) conformations. It is experimentally unknown which conformation has the lowest energy, but *ab initio* quantum calculations have estimated that the envelope is more stable by only 3 cal mol^{-1}[32]. However, another calculation[33] at even higher levels of theory, found that there is virtually no energy difference between the two conformations. Additionally, the pseudorotation of cyclopentane is considered unhindered, meaning that the pseudorotation barrier is small compared to the thermal energy at 300 K. The main reason for this is because the highest energy conformation of cyclopentane is the planar conformation, and the pseudorotation pathway never includes this conformation. Also, for the molecule to pseudorotate at room temperature, it would indicate that the pseudorotation barrier cannot be high compared to k_bT.

There are three parameters used to describe cyclopentane: barrier to planarity, pseudorotational amplitude (also called puckering amplitude), and pseudorotational velocity. The barrier to planarity is simply the energy difference between the planar (D_{5h}) conformation and either of the low-energy conformations. The pseudorotational amplitude is the amount of displacement from the planar conformation, and the pseudorotational velocity is the rate of rotation of the torsional C-C bonds in the cyclopentane ring during the pseudorotation. While the pseudorotational velocity is clearly a function of the simulation dynamics, the associated amplitude can also be a function of the dynamics because it is a thermal average of the deviation from the planar conformation.

2.1.2.3 Molecular Dynamics Algorithms

As previously mentioned, molecular dyamics is simply based on solving Newton's equations of motion. A MD algorithm generally contains two parts: an ensemble "controller" and an integrator. The most widely used integrator (and the one used in all cases in this work) is the Verlet integrator (or a modification thereof). The original Verlet algorithm[34],

of which the original mathematical algorithm is attributed to Störmer[35], is obtained by a Taylor expansion about $\boldsymbol{x}(t)$, where $\boldsymbol{x}$ is the position vector at a given time (t). The original algorithm is solely based on position, i.e., there is no need to calculate a velocity to solve for the trajectories. The major disadvantages are that the velocity calculation can only be done after the new trajectories are calculated, which makes the integrator difficult to use in other ensembles, and there is a possibility of numerical imprecision in solving the trajectory due to the addition of terms with different precision[36]. The major modifications to this algorithm in order to overcome these drawbacks are called leap-frog Verlet[37] and velocity Verlet[38]. Both modifications can be solved for the original Verlet algorithm, but they are written in a specific way that overcomes the disadvantages in the Verlet algorithm. The leap-frog Verlet algorithm uses half-step velocities in order to generate trajectories, where there is a disadvantage in that the temperature and kinetic energies are not calculated with the full step values. The velocity Verlet algorithm also uses half-step velocities, but it ends each step with all values at the same time so that the next step begins with all values evaluated to the beginning of that step. The disadvantage to the velocity implementation is that it is harder to couple to a controller.

The ensemble "controller" part of the MD algorithm is essentially a thermostat or barostat that is coupled to the integrator in order to produce a specific ensemble. The Verlet integrator by itself solves a system in the NVE ensemble. A thermostat or barostat is attached to the integrator in order to produce constant temperature (T) or constant pressure (P) algorithms, respectively. Normally, the algorithm is named after the controller implementation that is used, however, it is very important that the integrator is known. It will be shown later that controllers will behave differently when they are used with different integrators.

This work will focus solely on the NVT ensemble. The main reason that NVT is the optimal ensemble for this work is because normally experiments are performed at a fixed temperature and volume. The two main types of thermostats for NVT are classified as weak

coupling (WC) and extended system (ES). The two differ greatly in their implementation, and they can be likened to proportional and integral controllers, respectively.

Four different MD algorithms were used in these simulations: the Berendsen temperature controller integrated with the leapfrog Verlet algorithm (LFB)[39], the Nosé-Hoover temperature controller integrated by the Störmer-Verlet leapfrog algorithm (NH-SV)[40], the Nosé-Hoover temperature controller integrated by an implicit integration scheme based on a modification of the velocity Verlet algorithm (NH)[9], and the symplectic Nosé-Poincaré temperature controller integrated by the generalized leapfrog algorithm (NP)[41]. LFB is a WC algorithm, and the remainder employ the ES approach[42]. The NP approach maintains the Hamiltonian form and is symplectic. The use of a symplectic algorithm should allow a simulation to perform more accurately over very long term simulations by avoiding the interference of the thermostat algorithm in the system dynamics. Each algorithm has a thermal inertial parameter, which is denoted by Q, and describes the strength of the coupling of the thermostat to the system. This parameter acts as the controller time constant for the various algorithms. For LFB, each atomic velocity is periodically scaled by a factor of λ, which is described by Equation 2,

$$\lambda = [1 + \frac{\Delta t}{Q}(\frac{T_0}{T} - 1)]^{\frac{1}{2}}, \tag{2}$$

where Q is the coupling parameter, Δt is the integration time step, T_0 is the target temperature, and T is the current temperature of the simulation. Equation 3 shows the relationship for the thermostat scaling factor in the NH and NH-SV algorithms,

$$\dot{\xi} = \frac{f}{Q}(k_B T - k_B T_0), \tag{3}$$

where k_b is Boltzmann's constant, f is the degrees of freedom in the system, and ξ is the scaling factor that determines what fraction of the momentum, $\boldsymbol{p}$, is adjusted to keep the temperature fixed. This is carried out using the relationship

$$\dot{\boldsymbol{p}} = \boldsymbol{F} - \xi \boldsymbol{p}, \tag{4}$$

where $\boldsymbol{F}$ is the total force exerted on the particular atom. For the NP algorithm,

$$\dot{s} = s\frac{\pi}{Q}, \tag{5}$$

where s is used in every step to transform the velocity from the extended ($\boldsymbol{x}$, $\tilde{\boldsymbol{p}}$, s, π) space to ($\boldsymbol{x}$, $\boldsymbol{p}$) space, where $\boldsymbol{x}$ are the real coordinates, $\boldsymbol{p}$ is the real momentum, $\tilde{\boldsymbol{p}}$ is the canonically transformed momentum, s is the extended system coordinates, and π is the momentum associated with s[41] (full equations for all algorithms can be found in Appendix B). When the value of Q is reported, it will be reported in the units appropriate for its algorithm. For LFB, the unit for Q is ps, it is kcal ps mol^{-1} for both NH algorithms, and for NP, its units are kcal ps^2 mol^{-1}.

2.2 Cyclopentane Data and Parameter Calculation

2.2.1 Experimental and Theoretical Data

Experimental data from spectroscopic studies of cyclopentane has shown cyclopentane to have a barrier to planarity of 5.21±0.14 kcal mol^{-1}[45], however, this value was obtained using a model that interpreted spectroscopic data[51], and not directly from the experiment because the planar conformation is not sampled under typical conditions. This value is accepted as reasonable, however, and has been recently reproduced (within the experimental error) by *ab initio* quantum calculations at the MP2/cc-pVTZ level of theory[33].

The value for the pseudorotational velocity that is used is 400 deg ps^{-1}[52]. This theoretical value is entirely based on a calculation that assumes that there is no barrier to pseudorotation and that equipartition in one degree of freedom is the total contributor of the kinetic energy. While these assumptions are reasonable, particularly at room temperature, this value likely represents a lower bound to the pseudorotational velocity. Coupling

Table 1: Summary of previous experimental and simulated values of the pseudorotational amplitude (q)

Exptl. Method	$E_{barrier}$ (kcal mol^{-1})	q (Å)
Fit Enthalpy Data[43]	-	0.47
IR[44]	-	0.479
Raman[45]	5.21±0.14	0.47±0.04
Raman[46]	5.16	0.458
Electron Diffraction[47][1]	-	0.427±0.015
ED w/ Anharmonicity[47][2]	-	0.438
NMR[48]	-	0.463
HF/6-31G[49][3]	-	0.389
HF/6-31G*[49][3]	-	0.403
HF/6-31G**[49, 50][3]	-	0.403
B3LYP/6-31G**[33][3]	3.68	0.403
MP2/cc-pVDZ[33][3]	5.56	0.432
MP2/cc-pVTZ[33][3]	5.10	0.429
AM1[49][3]	-	0.217
AM1[32][3]	-	0.160
PM3[49][3]	-	0.284
PM3[32][3]	-	0.244
MM2[50][3]	4.68	-
MM3[30][4]	4.2	0.50±0.03
AMBER[31][4]	4.7	0.46±0.02
AMBER[49][3]	-	0.441
MM+[49][3]	-	0.444
MMX[32][3]	-	0.414

[1]The thermal average pseudorotational amplitude ($\langle q \rangle$)

[2]The mechanical equilibrium pseudorotational amplitude (q_e)

[3]Values for q are from optimized structures. Where the C_2 and C_s values vary, the C_s values are reported

[4]Values for q are from molecular dynamics and are averaged over the simulation

with other non-rotational degrees of freedom in the cyclopentane ring should help facilitate rotation and increase this velocity. The one-dimensional assumption in the theoretical calculation does not account for this coupling.

There have been several studies of the pseudorotational amplitude (q) of cyclopentane, which are summarized in Table 1, along with the reported barrier to planarity. The spectroscopic studies have used infrared[44], Raman[45, 46], NMR[48], and electron diffraction[47]. *Ab initio*[32, 33] and semi-empirical quantum calculations have also been performed[32, 49], as well as several classical molecular modeling studies[30, 31, 32, 49, 50]. Table 1 indicates that the values for the amplitude differ depending on the study, regardless of whether the study is computational or experimental. On the computational side, one reason for the disparity is that the *ab initio*, semi-empirical, and molecular mechanics simulations all have geometries that are optimized at 0 K, whereas the molecular dynamics studies are performed at room temperature. Additionally, there are very different levels of theory being used, and they have varying degrees of accuracy.

The spectroscopic studies all have one thing in common – they all use a rigid body model. This will simplify the mathematics, but it means that they typically over-predict the value of the pseudorotational amplitude. This occurs because allowing bond lengths and angles to deviate will reduce the amplitude of the change in torsion angles, and this deviation is not included in the rigid model. This explains why the spectroscopic values are all higher than the electron diffraction values. The electron diffraction studies are likely the most accurate, as there is less of a dependence on a model because the bond distances are derived experimentally. The two values given for electron diffraction are both useful as well. One, the thermal average ($\langle q \rangle$), corresponds to the observed average amplitude at 298 K. This $\langle q \rangle$ value is the thermodynamic equilibrium value, and is taken directly from the electron diffraction experiment. The other value (q_e), applies a correction for the anharmonic vibration in the ring to estimate the value of the amplitude at 0 K[47]. This q_e value corresponds to the amplitude at the low energy minimum and should be compared

to energy minimized structures, while $\langle q \rangle$ should be compared to the results of molecular dynamics experiments. Therefore, q_e will be continued to be called the equilibrium value to be consistent with the original electron diffraction study. However, it should be noted that this is not the thermodynamic equilibrium value, but it is the mechanical equilibrium value that corresponds to the potential energy minimum. The experimental values that will be compared to this work will focus on the barrier to planarity found by Carreira and coworkers, the pseudorotational amplitude found by Adams and coworkers using electron diffraction, and the pseudorotational velocity calculated by Henry and coworkers.

All of the previous computational studies show an interesting trend. The *ab initio* values get closer to the experimental values for the barrier and the amplitude as the level of theory increases. Additionally, it is apparent from Table 1 that semi-empirical methods are not adequate for this molecule. Energy minimization methods with AMBER and variations of the MM series of force fields seem to perform close to the same level as the *ab initio* studies at the highest level of theory performed. The molecular dynamics studies, however, do not seem to perform nearly as well, even with the same or similar force fields. This discrepancy occurs because MD is preferentially sampling regions of higher pseudorotational amplitude as seen in a higher value of $\langle q \rangle$ from the MD simulations relative to the q_e value from the energy minimization calculations in Table 1. This result is the opposite of what the experimental electron diffraction results indicate. This effect could be due, in part, to the force field if the energy hypersurface about the pseudorotational minimum is asymmetric. However, one of the results of this work will show that the sign of the difference between q_e and $\langle q \rangle$ can be changed by altering MD parameters and algorithm, which underscores the importance of these aspects of the simulation.

2.2.1.1 Parameter Calculation

The barrier to planarity is easily calculated by finding the potential difference between a planar cyclopentane molecule and one of the low-energy conformations ($E_{\text{barrier}} = E_{\text{planar}} -$

$E_{\text{envelope/twist}}$). Additionally, it is trivial to calculate the pseudorotational amplitude and velocity from the endocyclic torsion angles. The Altona-Sundaralingam formulism[27] for the pseudorotational properties was used, but it has been shown that this is similar to other calculation methods[29, 53], which indicates that the calculation methods are not a significant factor in the calculation of these parameters. To calculate the pseudorotational amplitude (q), the following formulae are utilized,

$$q = \frac{\theta_m}{102.5}, \tag{6}$$

$$\theta_m = \frac{2}{5}\sqrt{\left(\sum_{j=0}^{4}\theta_{\text{j}}\cos\alpha_{\text{j}}\right)^2 + \left(\sum_{j=0}^{4}\theta_{\text{j}}\sin\alpha_{\text{j}}\right)^2}, \tag{7}$$

$$\alpha_{\text{j}} = \frac{4\pi}{5}j, \tag{8}$$

where θ_m is the maximum deviation of the endocyclic torsion angles from 0°, and θ_{j} is the one of the five endocyclic torsion angles. The pseudorotational velocity (v_{rms}) is found using

$$v_{\text{rms}} = \sqrt{\sum_{i=1}^{N}\left(\frac{v_{\text{i}}^2}{N}\right)}, \tag{9}$$

$$v_{\text{i}} = \frac{\text{d}P}{\text{d}t} = \frac{\Delta P}{\Delta t}, \tag{10}$$

$$P = \tan^{-1}\left[\frac{(\theta_2+\theta_4)-(\theta_1+\theta_3)}{2\theta_0(\sin 36+\sin 72)}\right], \tag{11}$$

where P is the pseudorotational phase angle, the θ terms denote each of the torsion angles, and v_{i} is the instantaneous velocity defined by changes in phase angle and time, which are denoted by ΔP and Δt, respectively.

2.3 Simulation Methodology

2.3.1 Barrier to Planarity

In order to determine the barrier to planarity, a planar cyclopentane molecule was minimized to find the energy difference between the planar structure and the lowest energy conformation. This was done using version 2005.06 of the Molecular Operating Environment (MOE) from the Chemical Computing Group (Montreal, Canada, http://www.chemcomp.com). Subsequent energy minimization and molecular dynamics were also carried out using the MOE software. The total potential energy of this molecule was minimized using a truncated Newton algorithm[54], and the molecule was minimized to a gradient of 0.01 kcal mol^{-1} $Å^{-1}$. The initial planar conformation used was obtained by creating a cyclopentane molecule that was arranged in a single plane and had bond lengths equal to the equilibrium bond lengths of the force field being examined. This is a metastable conformation which was minimized to produce the initial planar molecule used in subsequent simulations. The final structure of the molecule was obtained by making a small perturbation in one of the carbon atoms and subsequently minimizing the energy. The energy was recorded for the initial planar structure and after the minimization to yield the barrier to planarity. Additionally, the final conformation that was obtained depended on how the carbon atom was perturbed, so it was possible to examine differences between the half-chair and envelope conformations. The force fields used were AMBER94[55], MMFF94[56], PEF95SAC[57], OPLS-AA[58], CHARMM22[59], and Engh-Huber[60]. It should be noted here that the MOE implementation for CHARMM22 is slightly different from the original in that the atom-centered charges have been approximated to match the charges in the original CHARMM22 force field. Unfortunately, because MOE does not include all of the types of cross terms necessary for MM3, calculations were not performed for MM3, so no direct comparison to MM3 can be made, however, the results obtained here will be compared to the previous study.

2.3.2 Molecular Dynamics

All MD simulations were performed at a temperature of 300 K, and the calculations used a time step of 0.02 fs in order to match the simulations used in the previous work. This time step would normally be considered far too small for typical production runs, however, it allows for direct comparison to previous work, and it reduces integration error to ensure that the differences between various algorithms is not just the difference in integration error. The MD trajectories were sampled every 100 time steps, or every 2 fs.

2.3.2.1 Equilibration

Analysis of a 1 ns simulation of a single cyclopentane molecule confirmed that a simulation length of 100 ps was more than sufficient for the equilibration of these isolated molecule simulations. In all cases, the system equilibrated in only a fraction of 100 ps. It is expected that a small, isolated molecule like cyclopentane will undergo rapid equilibration, and in this case, equilibration occurs in under five picoseconds. The results of this study for the AMBER94 force field are shown in Table 2, where the initial and final values correspond to data from the first 100 ps and last 100 ps from the 1 ns simulation. The largest change over one nanosecond is in the mean-squared fluctuations in the kinetic energy $\langle \delta K^2 \rangle$, followed by the pseudorotational velocity. The largest deviation over the nanosecond simulation is with $\langle \delta K^2 \rangle$ for NP at 11%, but the results from Table 2 indicate that this does not seem to have a substantial correlation with the deviation in the pseudorotational parameters, so this is assumed to be acceptable. These simulations reveal that equilibration occurs quickly, and a 100 ps simulation is a long enough simulation for the purposes of this study.

2.3.2.2 Simulations

The majority of the simulations that were performed used a single isolated cyclopentane molecule. This was useful for several reasons, mainly the ease of comparison with previous simulations, which were all single molecule simulations. In addition, it allows for ease of

Table 2: Results of equilibration study. Comparison between the initial 100 ps and final 100 ps of a 1 ns simulation for fluctuations ($\langle \delta K^2 \rangle$, kcal^2 mol^{-2}) and pseudorotational amplitudes ($\langle q \rangle$, Å) and velocities (v_{rms}, deg ps^{-1}) using the AMBER94 force field.

Algorithm	Initial $\langle \delta K^2 \rangle$	Final $\langle \delta K^2 \rangle$	% Change	Initial $\langle q \rangle$	Final $\langle q \rangle$	% Change	Initial v_{rms}	Final v_{rms}	% Change
LFB	-	-	-	0.488	0.495	1%	1301	1356	4%
NH	7.1	6.8	4%	0.412	0.410	1%	649	663	2%
NH-SV	6.6	6.8	3%	0.421	0.419	0.4%	678	620	9%
NP	8.9	8.0	11%	0.403	0.410	2%	604	626	4%

Table 3: Q and $\langle \delta K^2 \rangle$ (kcal^2 mol^{-2}) for each algorithm. The expected value of the fluctuations based on statistical mechanics is 8.0 kcal^2 mol^{-2}.

	LFB		NP		NH-SV		NH	
Force Field	Q	$\langle \delta K^2 \rangle$	Q	$\langle \delta K^2 \rangle$	Q	$\langle \delta K^2 \rangle$	Q	$\langle \delta K^2 \rangle$
CHARMM22	0.2	-	0.002	7.4±0.4	0.0002	7.0±0.1	0.0002	7.1±0.2
OPLS	0.2	-	0.001	7.4±0.3	0.0002	7.0±0.2	0.0002	7.1±0.2
AMBER94	0.2	-	0.001	7.1±0.4	0.0002	6.9±0.1	0.0002	7.0±0.1
MMFF94	0.2	-	0.001	6.9±0.3	0.0002	7.0±0.1	0.0002	6.9±0.1
PEF95SAC	0.2	-	0.0008	7.2±0.6	0.00015	7.0±0.1	0.00015	6.9±0.1
Engh-Huber	0.2	-	0.0005	7.1±0.1	0.0002	7.0±0.1	0.0002	6.8±0.1

comparison to the experimental data, where most of the data was obtained using the gas phase. A single molecule simulation is the equivalent of the ideal gas case. Because, for the most part, the experiments were performed at ambient temperature and pressure, it is assumed that an ideal gas is an accurate depiction of the gas phase cyclopentane system. Simulations were carried out using the Berendsen temperature controller integrated with the leapfrog Verlet algorithm (LFB) with a coupling constant of 0.2 ps in order to match the previous AMBER results in the literature. The next simulations were carried out with the NP, NH, and NH-SV algorithms, and for each force field, the value of Q that was used reproduced the appropriate level of kinetic energy fluctuations for the NVT ensemble from statistical mechanics, as given by Equation 12,

$$\langle \delta K^2 \rangle_{NVT} = \langle (K - \frac{3}{2} N k_b T)^2 \rangle = \frac{3}{2} N (k_b T)^2, \tag{12}$$

where K is the instantaneous kinetic energy in the simulation, $\langle \delta K^2 \rangle$ is the mean-squared fluctuations in the kinetic energy about the mean, N is the number of molecules, k_b is Boltzmann's constant, and T is the temperature[36]. For cyclopentane at 300 K, $\langle \delta K^2 \rangle_{NVT}$ = 8.0 kcal^2 mol^{-2}. By using this relationship for the kinetic energy fluctuations, a value for Q was found that correctly reproduces fluctuations for each force field and algorithm. These values are reported for each force field in Table 3. The results for $\langle \delta K^2 \rangle$ are not shown for LFB, as it has been shown that the LFB algorithm does not sample the canonical ensemble. Therefore the fluctuation relationships for the canonical ensemble are not applicable to LFB, though fluctuation relationships have been published for the ensemble sampled by weak coupling algorithms[61]. This value for Q was used with the LFB algorithm because it matched the value used in previous work[31]. For the ES algorithms, while the values for $\langle \delta K^2 \rangle$ are not exactly those of cyclopentane from statistical mechanics, they are within 1.2 kcal^2 mol^{-2}. While this variation represents a 15% difference, the sensitivity analysis, described later, indicates that such a variation has a negligible effect on the cyclopentane dynamics. Using these values for Q, a simulation was run for each force field with each

algorithm for 100 ps. This was done a total of 15 times with different initial velocity distributions in order to determine the variation of the calculated quantities.

2.3.2.3 Liquid Phase

The liquid case was also studied to see if there was a difference between the pseudorotational dynamics between the gas and liquid phases. The previous simulations were isolated molecules, and these are taken to be equivalent to the gas phase, assuming the gas is ideal. For the liquid case, a periodic cube 15 Å on a side was created with 21 cyclopentane molecules in it, giving the system the experimentally measured liquid density with a specific gravity of 0.745 at 300 K[62]. The simulation was run with periodic boundary conditions with the NP algorithm with value of 0.001 kcal ps^2 mol^{-1} for Q. The minimum image convention was enforced such that the effective potential energy cutoff was 7.5 Å, or half of the box width. This was done for both MMFF94 and AMBER94 force fields. This system was run for 1 ns, and the first and final 100 ps were analyzed.

2.3.2.4 Long Simulations

To examine the effects of the algorithms over long simulation periods, 1 μs simulations were performed. The first set of simulations were done with cyclopentane using a time step of 0.02 fs with NP, NH, and NH-SV in order to remove as much round-off error as possible so that the vast majority of the accumulation of the relative energy error (REE) is due to the details of the algorithm rather than the round-off error. This also allows a direct comparison of the algorithms studied. The REE as a function of time was examined to determine the amount of error introduced by each ES controller algorithm. The REE is defined as

$$\mathrm{REE} = \left| \frac{\mathrm{E(t)} - \mathrm{E(0)}}{\mathrm{E(0)}} \right|, \tag{13}$$

where E is the conserved quantity for the algorithm, t is the current timestep, and 0 denotes the energy at the start of the simulation. For NH and NH-SV, this quantity is

$$E = \sum_i \frac{\boldsymbol{p}_{\mathrm{i}}^2}{2m_{\mathrm{i}}} + U(\boldsymbol{x}) + \frac{Q\xi^2}{2} + gk_bT\eta, \tag{14}$$

and for the NP algorithm this quantity is

$$E = H_{\mathrm{Nosé}} = \sum_i \frac{\tilde{\boldsymbol{p}}_{\mathrm{i}}^2}{2m_{\mathrm{i}}s^2} + U(\boldsymbol{x}) + \frac{\pi^2}{2Q} + gk_bT\ ln\ s, \tag{15}$$

where m is the mass of an atom, $\boldsymbol{x}$ are the atomic coordinates, g denotes the degrees of freedom, $U(\boldsymbol{x})$ denotes the potential energy, η is the integrated value of ξ, and $H_{\mathrm{Nosé}}$ is the Nosé Hamiltonian. All other variables were defined previously.

As long as the temperature is kept relatively constant, i.e., the average temperature for the simulation is close to the desired temperature, the REE measures the amount of the controller contribution to the conserved quantity for the particular algorithm. An increase in the REE occurs as the dynamics of the system become dominated by the controller algorithm, and this can change the dynamics of the simulated system over long periods of time. Since this effort was to extend the work done by Bond and coworkers in comparing the different ES algorithms, simulations examining the time step were also performed. To benchmark the results obtained here with the work done by Bond and coworkers, a simulation of 108 argon atoms in a cube with 16.5 Å sides was reproduced. The temperature was 720 K, and each simulation was performed for 100 ps with Q set to 0.1 for each algorithm. The time step ranged between 0.01 and 4 fs in order to determine the sensitivity of the REE to the integration time step. The maximum REE was recorded for each simulation, as long as the average simulation temperature was within 5% of the desired temperature. This corresponds to the simulations by Bond and coworkers[41], which used a Lennard-Jones fluid with Q = 0.1, T^* = 6.0, ρ^* = 0.95, and N = 108, where T^* and ρ^* are reduced quantities for the temperature and density that have been scaled by the Lennard-Jones potential parameters. The main reason for this simulation was two-fold. First, it allowed a comparison to previous work, and second, simulations using the explicit Nosé-Poincaré algorithm (NPe)[63] were performed. This algorithm was not used up to this point due to time and

resource constraints, however, there is enough data presented on NPe to warrant a concrete conclusion on the performance of implicit versus explicit integration of the Nosé-Poincaré algorithm.

Additionally, three cyclopentane systems were simulated over the same time step range as the previous argon simulations, simply to examine the REE for different types of systems. The first was an isolated cyclopentane molecule at 300 K. The second simulation used the liquid cyclopentane system described above with a temperature of 300 K. The third also used the liquid system, except the simulation was performed at 900 K. The Q parameter was set to 0.0002 for all of the algorithms for all three systems in the units of Q for that algorithm.

2.3.3 Simulation Averages

For a 100 ps simulation sampled every 2 fs, there are 50,000 total data points. Each pseudorotational parameter and simulation average is calculated using all of the data points. Generally, the values reported will be the average of the quantity over all of the repeated simulations, except where, like the condensed phase simulations, the simulation was only performed once. The value that is reported for each force field is the simulation average with its 90% confidence interval. This was calculated from the standard deviation of the pseudorotational parameters over the fifteen repeated simulations. It is assumed that fifteen simulations for a small molecule system will give a reasonable average, and the standard deviation will accurately reflect the amount of variance between simulations. The 90% confidence interval is used throughout this chapter to denote the "simulation error," and this work uses the confidence interval to denote the limit of the variability between simulations.

Table 4: Energy barrier to the planar conformation of cyclopentane compared to experiment (5.21 kcal mol^{-1}) and pseudorotational amplitude compared to minimum potential energy value (q_e = 0.438 Å)

Force Field	$E_{barrier}$ (kcal mol^{-1})	Exptl. Deviation (kcal mol^{-1})	q_e (Å)	Exptl. Deviation (Å)
MP2/cc-pVTZ[33]	5.10	-0.11	0.429	-0.009
MMFF94	5.43	0.22	0.401	-0.037
AMBER94	4.75	-0.46	0.419	-0.019
AMBER[31]	4.7	-0.51	-	-
Engh-Huber	4.46	-0.75	0.298	-0.140
PEF95SAC	6.16	0.95	0.402	-0.036
MM3[31]	4.2	-1.01	-	-
OPLS-AA	3.90	-1.31	0.376	-0.062
CHARMM22	2.29	-2.92	0.304	-0.134

2.4 Results and Discussion

2.4.1 Barrier to Planarity

Because it is purely an energetic barrier, the barrier to planarity is only a function of the force field, and therefore not the MD algorithm. Despite this, the dynamics can be affected by the barrier to planarity. While some of the force fields that were used were not parametrized with small organic molecules like cyclopentane (e.g. AMBER94 and CHARMM22), their use gives some insight into the effect that force field choice can have on cyclopentane modeling. The barrier to the planar conformation was found for each force field, and it is listed in Table 4. The results for the AMBER94 force field are the same as the results published previously[31], so these results should allow for direct comparison to the previous work. As expected, the high level *ab initio* calculations of Wu and coworkers has the closest value to experiment, but of the force fields, it is obvious that MMFF94 is the closest of the tested force fields to the experimental value of 5.21±0.14 kcal mol^{-1}, with AMBER94 also producing reasonably close results. However, looking at the value for q_e for each force field shows that the barrier to planarity has no direct correlation with the ability to accurately reproduce q_e. This is likely due to the fact that the planar conformation

is a very high energy conformation that is not sampled in typical dynamics. Therefore, this barrier is not a good indication of how the pseudorotation of the ring will behave.

The minimized structures were examined to find the cause of this discrepancy. The envelope conformation was used to compare the structures of Wu and coworkers to the energy minimization results. While only the results for the envelope conformation are reported here, the results have an identical trend for the twisted conformation. For each force field, the bond lengths, bond angles, and torsion angles for the minimized structures were different from the *ab initio* values. However, bond lengths and bond angles are not involved in calculating the pseudorotational properties, and additionally, these are usually going to be reasonably accurate for a given force field. The values for the torsion angles can be found in Table 5 (the naming convention in this table follows the convention used by Wu and coworkers for atom assignment). These data show that the closer the torsion angles are to the *ab initio* values, the closer the value of q_e is to the experimental value. Also, the deviation in q_e is proportional to the total deviation of all three torsion angles. The only exception to this is the Engh-Huber force field. This is most likely due to Engh-Huber using unified atoms instead of hydrogens. This is logical given the fact that the torsion angles at the local energy minimum actually determine the maginitude of the pseudorotational amplitude, and the lack of additional torsion terms that would be associated with torsion angles from H-C-C-H and H-C-C-C bonds will change the effect of the torsion energy on q_e.

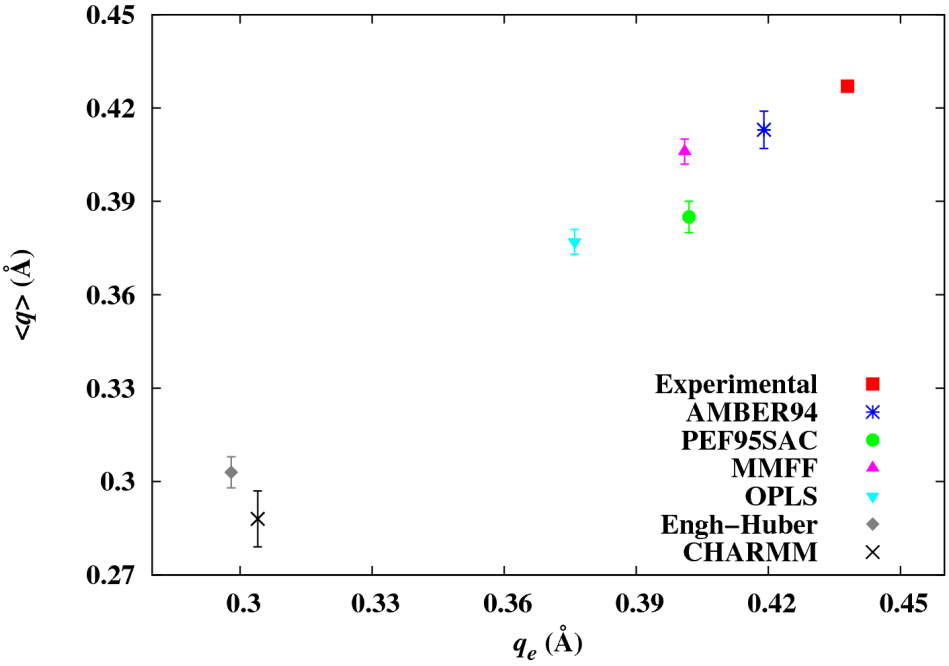
2.4.2 Pseudorotational Parameters

2.4.2.1 Effect of Minimized Structure Accuracy

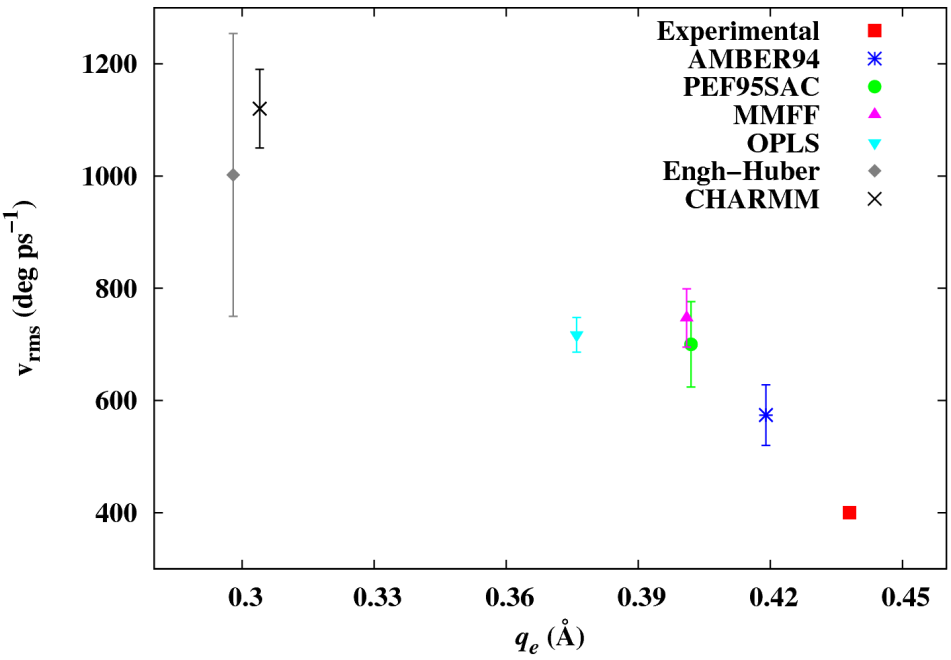
Each force field will have a different pseudorotational amplitude and velocity because each force field will vary in its energetics. As was discussed above, a minimized structure that matches the *ab initio* geometry produces a more accurate value of q_e, however, it is not obvious whether this is important for dynamic pseudorotational parameter accuracy. To

Table 5: Comparison of torsion angles (Φ) of each force field to *ab initio* torsion values for the C_s conformation. The total deviation column lists the total difference between the force field values and the *ab initio* values over all three torsion angles ($\sum |\Phi_i^{ai} - \Phi_i^{ff}|$). All values of the torsion angles are in degrees.

Force Field	q_e (Å)	Φ(C5C1C2C3)	Φ(C1C2C3C4)	Φ(C2C3C4C5)	Total Deviation
MP2/cc-pVTZ[33]	0.429	-42.1	25.9	0	-
AMBER94	0.419	-41.7	27.7	-3.1	5.3
PEF95SAC	0.402	-40.3	27.3	-4.0	7.2
MMFF94	0.401	-38.3	22.2	2.4	9.9
OPLS-AA	0.376	-33.2	15.2	8.6	28.2
CHARMM22	0.304	-27.9	14.4	4.6	30.3
Engh-Huber	0.298	-29.5	19.1	-1.4	20.8



(a)



(b)

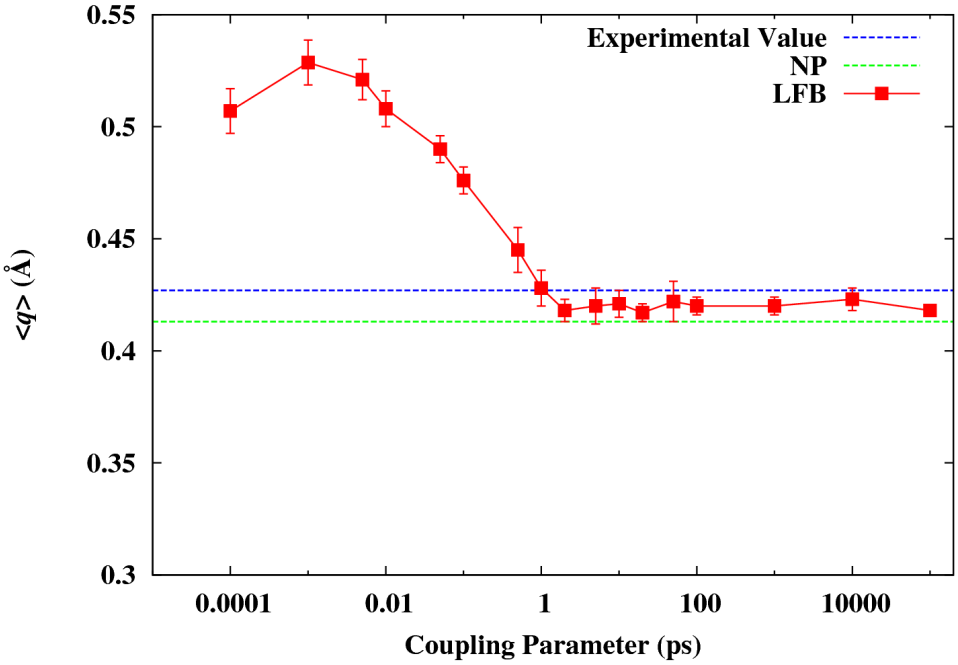
Figure 6: The effect of the minimum potential pseudorotational amplitude for each force field on (a) the mean observed pseudorotational amplitude and (b) the pseudorotational velocity.

examine this relationship, each force field was used to determine the effect of the relationship between the q_e and both $\langle q \rangle$ and the pseudorotational velocity. The results can be found in Figure 6, and it is apparent that AMBER94 produces the closest values to the experimental values for all of the measured properties. The only ones that come close to this performance are MMFF94 and PEF95SAC, which produce similar results. From this result, MMFF94 would be recommended over PEF95SAC, mainly because it has a closer value of $\langle q \rangle$ than PEF95SAC. Looking at the velocity results, both force fields perform identically within the error. Since there is only a 0.001 Å difference between q_e and a 0.021 Å difference between $\langle q \rangle$ for MMFF94 and PEF95SAC, the deciding factor should fall to $\langle q \rangle$, which is where the two force fields differ the most. Therefore, MMFF94 is recommended over PEF95SAC because of its closer value of $\langle q \rangle$. Because of the accuracy of the AMBER and MMFF94 force fields, the rest of this work will mainly focus on AMBER and MMFF94.

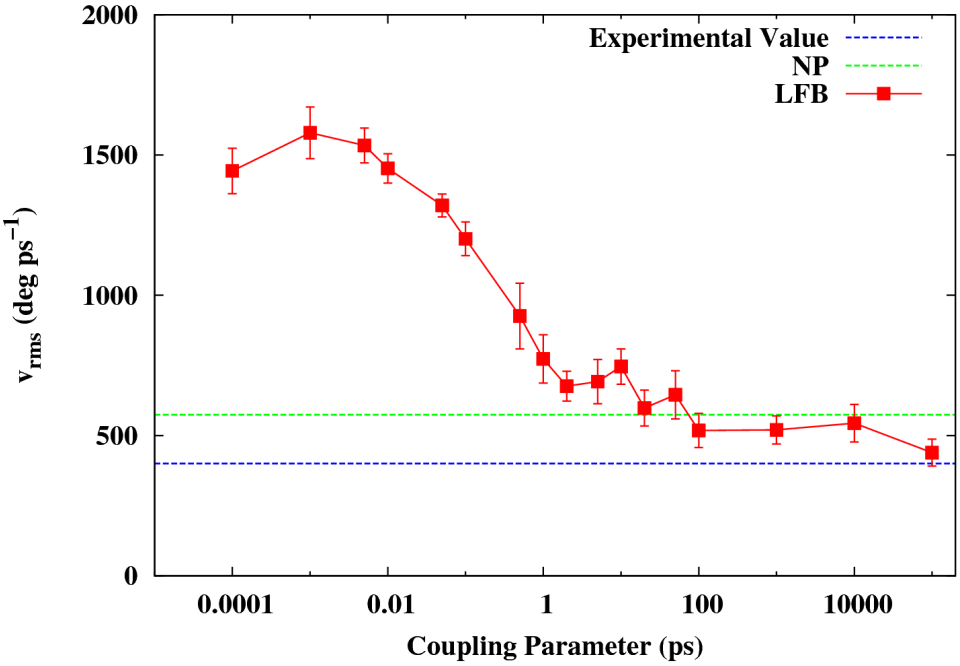
Since q_e is calculated at the local energy minimum in potential energy, and the *ab initio* geometry produced a good estimate of q_e, it is reasonable that the force field whose energy minimum comes closest to the *ab initio* geometry would produce the most accurate value of q_e. This trend is observed in the data and indicates the ability of the force field to reproduce the barrier to planarity may be of secondary importance given the negligible degree to which this very high energy planar conformation is sampled at typical temperatures.

2.4.2.2 Effect of the Thermal Coupling Parameter

The first simulation was an attempt to reproduce the results from Cornell and coworkers[31]. Using a thermostat coupling constant of 0.2 ps (the same as the reported value from this previous study) with the AMBER94 force field, it was found that the puckering amplitude was 0.461±0.008 Å and the pseudorotational velocity was 1086±68 deg ps^{-1}. This is in good agreement with the previous study's results of 0.46±0.02 Å and 1036±300 deg ps^{-1}. Figures 7(a) and 7(b) show the effect the coupling parameter has with the LFB algorithm



(a)



(b)

Figure 7: The effect of the coupling parameter for the AMBER94 force field and LFB algorithm on (a) the pseudorotational amplitude and (b) the pseudorotational velocity. The line labeled "Experimental Value" represents the accepted experimental value, and the line labeled "NP" shows the mean value obtained from NP simulations.

using AMBER94. These plots show that for both the pseudorotational amplitude and velocity, the values for both parameters change as Q is increased. It is important to note that for values of Q greater than 1 ps, the temperature is no longer effectively controlled to 300 K. At a Q value of 1 ps, the mean temperature is 297 K, which is within 1% of the target temperature, but with a Q of 100000 ps, the mean temperature is only 152 K. However, for values of Q less than 1 ps, which produce a consistently controlled temperature, the pseudorotational parameters are quite sensitive. This indicates that for WC algorithms, the choice of Q is very important. The LFB algorithm has been used in a large number of MD simulations since it was published in 1984, and most molecular simulations use the original recommended value of 0.1-0.4 ps for Q, which was originally optimized for a periodic water simulation. However, it is seen here that this is system-dependent, and it is imperative that a proper coupling constant be found. Just a 0.8 ps difference in the coupling parameter produced a deviation in the pseudorotational amplitude of more than 10% in this case. This entire deviation is wholly artificial and should be avoided as much as possible to ensure proper dynamics. Additionally, it must be noted that the pseudorotational velocity values here do not agree with those of Cornell and coworkers[31]. They consistently obtained higher values for the pseudorotational velocity for different values of the coupling parameter. This could be caused from a number of factors, from simulation details to implementation details. In this case, the most likely reason is that they were using averages over only 10 ps of data, while these simulations use 100 ps of data. Due to the nature of the pseudorotational velocity calculation, the amount of available data will have a large impact, particularly with smaller sample sizes.

For the ES MD algorithms, the variation of the dynamics with Q is different, and it is summarized in Table 6. While this table focuses on the results from the AMBER94 force field and the NP and NH-SV algorithms, these results followed the same trend for all three ES algorithms and all of the force fields. These results show that Q has little effect on the pseudorotational parameters. The value of Q does have a large effect on the kinetic energy

Table 6: The effect of Q for the AMBER94 force field using ES algorithms. The units of Q for NP are kcal ps^2 mol^{-1}, and they are kcal ps mol^{-1} for NH-SV. The average of all of the shown values are reported, along with the maximum deviation from that average and the 90% confidence interval.

Force Field	MD Algorithm	Q	$\langle \delta K^2 \rangle$ (kcal^2 mol^{-2})	$\langle q \rangle$ (Å)	v_{rms} (deg ps^{-1})
AMBER94	NP	0.00001	2.4	0.408	717
AMBER94	NP	0.001	7.1	0.419	724
AMBER94	NP	0.01	12.0	0.410	730
AMBER94	NP	0.1	19.0	0.407	734
AMBER94	NP	1	27.8	0.419	671
Median				0.413	703
Max Deviation				0.006	32
90% CI			0.4	0.006	54
AMBER94	NH-SV	0.00001	0.6	0.413	712
AMBER94	NH-SV	0.001	6.9	0.415	679
AMBER94	NH-SV	0.01	15.4	0.414	670
AMBER94	NH-SV	0.1	21.7	0.419	677
AMBER94	NH-SV	1	22.7	0.419	677
Median				0.416	691
Max Deviation				0.003	21
90% CI			0.1	0.003	32

fluctuations, while the pseudorotational parameters remain relatively constant. For example, for the NH-SV algorithm, Q ranges from 0.0001-1 kcal ps mol^{-1}, which causes $\langle \delta K^2 \rangle$ to range from 0.6-22.7 kcal^2 mol^{-2}. However, the pseudorotational amplitude only ranges from 0.413-0.419 Å, and the pseudorotational velocity ranges from 670-712 deg/ps. These drastic changes in fluctuations, while not affecting the local pseudorotation dynamics, will likely affect the bulk thermodynamic and transport properties that are also coupled to dynamics. This is a significant advantage to the ES algorithms because the Q parameter may be adjusted to produce the appropriate properties without significantly affecting the dynamics. In the NVT ensemble, the mean squared kinetic energy fluctuations are constrained to the previously described relationship. This constraint can be fulfilled by adjusting Q while not affecting the pseudorotational dynamics significantly. This is different from the LFB results, which show that over the same range of Q, the pseudorotational amplitude deviates from 0.42-0.53 Å, and the pseudorotational velocity deviates from 520-1580 deg ps^{-1}. It is also important to note that the values of the pseudorotational amplitude in Table 6 correspond to the thermal or time average $\langle q \rangle$. These values should be lower than the low energy minimum value based on electron diffraction results presented previously. All of these values of $\langle q \rangle$ in Table 6 for ES algorithms are lower than the value of q_e of 0.44 Å for the AMBER force field.

The MD results are important because they highlight that while for LFB, the value for Q must be found to reproduce cyclopentane dynamics, for an ES algorithm, this value will not affect these dynamics. It has previously been shown that extended systems methods are ergodic for a wide range of Q[64], and since cyclopentane dynamics do not change with Q, they must be a property of cyclopentane in the canonical ensemble, indicating that neither Q nor MD algorithm should change its pseudorotational dynamics, as long as the MD algorithm is a canonical ensemble algorithm. However, because $\langle \delta K^2 \rangle$ is a fundamental property of the canonical ensemble, if Q does not reproduce the fluctuations for the ensemble, statistical mechanical fluctuation relationships will not be valid, which will cause any

Table 7: The effect of the MD algorithm for the AMBER94, MMFF94, and MM3 force fields. The units of Q for NP are kcal ps^2 mol^{-1}, and they are ps for LFB.

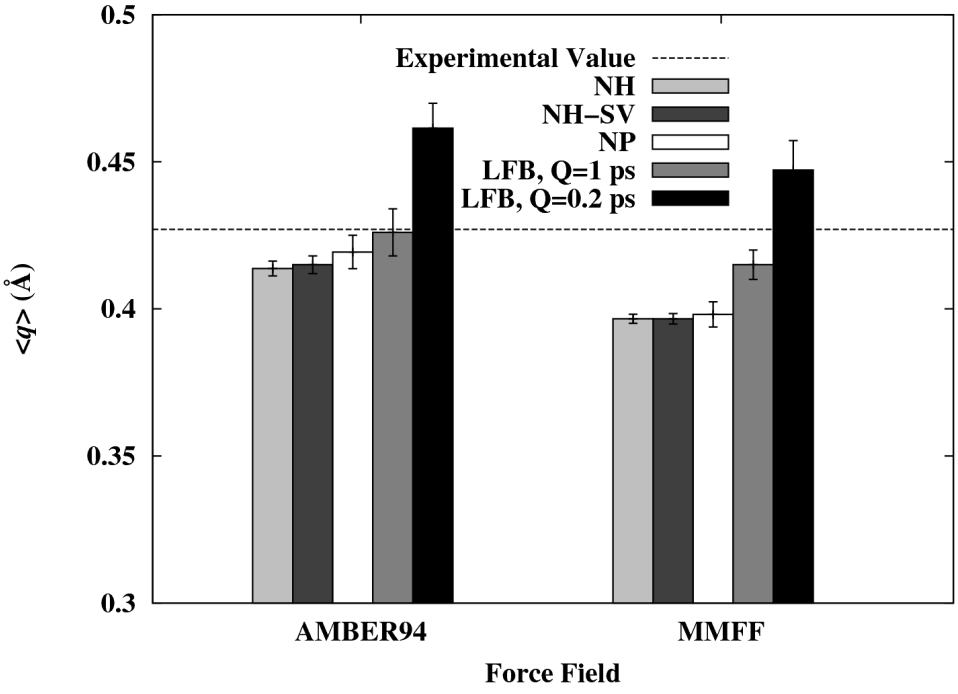
Force Field	MD Algorithm	Q	$\langle \delta K^2 \rangle$ (kcal^2 mol^{-2})	$\langle q \rangle$ (Å)	v_{rms} (deg ps^{-1})
Experimental			7.995	0.427	400
MM3[30]	LFB	0.1	-	0.5±0.03	1700±300
AMBER94[31]	LFB	0.2	-	0.46±0.02	1036±300
AMBER94	LFB	0.2	-	0.461±0.008	1086±68
MMFF94	LFB	0.2	-	0.450±0.01	1265±118
AMBER94	LFB	1	-	0.426±0.008	753±84
MMFF94	LFB	1	-	0.415±0.005	874±63
AMBER94	NP	0.001	7.1±0.4	0.419±0.006	724±54
MMFF94	NP	0.001	6.9±0.3	0.398±0.004	744±52

statistical mechanical property relationships, e.g., diffusivity, to be erroneous.

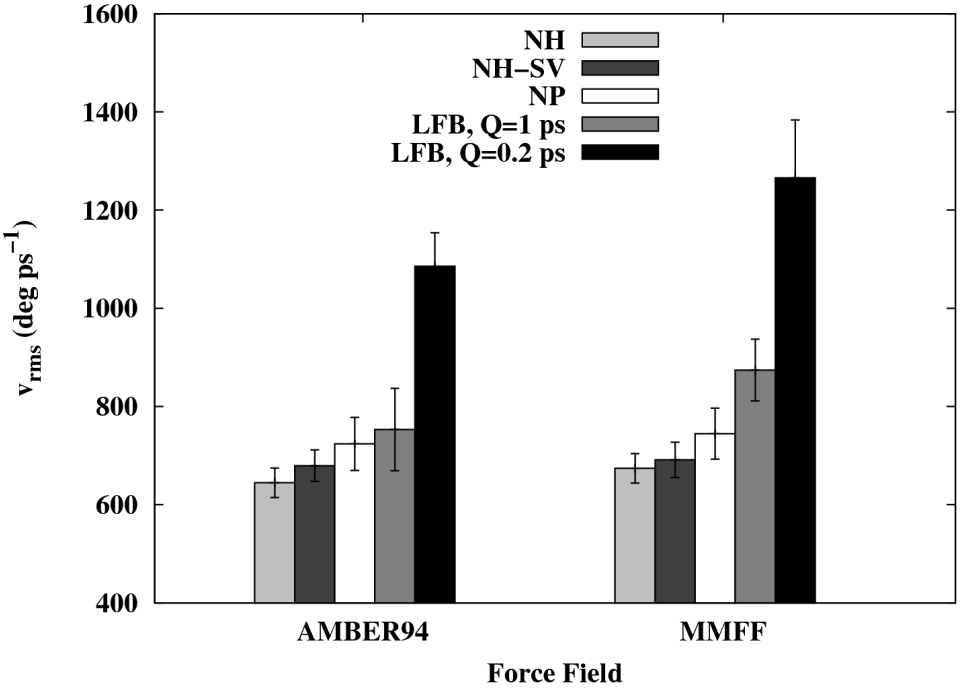
2.4.2.3 Effect of MD algorithm

The results for the different MD algorithms are shown in Figure 8. While Figure 8(b) shows more variance in v_{rms} than $\langle q \rangle$, it is clear that NH, NH-SV, and NP all perform similarly, with NP consistently producing the best results for the pseudorotational parameters. It would also appear that with the proper Q, LFB can produce local dynamics similar to its ES counterparts. From Figure 7, it shows that it is possible to choose a value for Q that produces similar results compared to the other algorithms, which can be seen in Figure 8 with Q set at 1 ps. Table 7 shows that the results reported in the literature for AMBER94 were able to be reproduced. However, the fact that Q must be adjusted to produce the correct pseudorotational dynamics means that it may not be possible to meet the aforementioned fluctuation criterion independently.

The data in Table 7 also show the effect of Q and the MD algorithm on the symmetry of motion characterized by the electron diffraction data. The electron diffraction data and its correction for the asymmetric bond vibration indicate that $\langle q \rangle$ should be lower than q_e. This was the case for ES algorithms in both Table 6 and Table 7, but not for the LFB algorithm in Table 7. This difference underscores the important effect of MD algorithm



(a)



(b)

Figure 8: The effect of the MD algorithm on (a) the pseudorotational amplitude and (b) the pseudorotational velocity for all force fields. The dashed lines represent the experimental data. If the value of Q is not given, it is the value from Table 3.

on the dynamics. The symmetry of the motion about the energy minimum is affected by the MD algorithm. Note that the correct symmetry of motion can be achieved by changing Q for the LFB algorithm, but only when using values of Q that do not provide reasonble temperature control.

2.4.2.4 Sensitivity of the Dynamics to Force Field Parameters

Havrvey and Prabhakaran have previously shown that the torsion energy makes the largest contribution to the dynamics of a ribose molecule[29]. This makes sense when comparing the pseudorotational amplitudes and velocities between force fields, as these parameters are solely calculated by the torsion angles. While the bond length and angle parameters do couple to the torsional modes, this is a secondary effect. Therefore, in an attempt to get closer to experimental values, the torsion parameters for AMBER94 and MMFF94 were modified to investigate the sensitivity of the pseudorotational paramters to the torsional force field parameters. For AMBER94, the torsion is the same for any combination centered around a carbon bond, e.g., torsion parameters for H-C-C-H are the same as those for C-C-C-C. Also, these torsions only have one force constant (a two-fold potential), so this was increased from 1.4 kcal mol^{-1} to 1.74 kcal mol^{-1}. This brought the torsion angles to values that were much closer to the *ab initio* values, and the minimum potential energy amplitude was 0.438 Å, which is identical to the value derived from experiment. However, the dynamical properties did not match, as $\langle q \rangle$ was 0.437 Å. Also, the pseudorotational velocity decreased to 640 deg ps^{-1}, which is closer to the expected value. For MMFF94, there are different torsion parameters for each type of torsion combination, e.g., there are different parameters for C-C-C-C and H-C-C-H. Since the amplitude is calculated by the C-C-C-C dihedrals, this was the dihedral that was focused on. MMFF94 uses three torsion parameters (one-, two-, and three-fold) for each dihedral type. Manipulating the two- and three-fold parameters proved to be the most fruitful, and these were changed from -0.274 and 0.563 kcal mol^{-1} to -0.825 and 0.9 kcal mol^{-1}, respectively. This led to a q_e of

Table 8: Comparison of gas and liquid phase simulations using the NP algorithm and Q is set at 0.001 kcal ps mol^{-1}

Force Field	Phase	$\langle q \rangle$ (Å)	v_{rms} (deg ps^{-1})
AMBER94	Liq.	0.413	644
AMBER94	Gas	0.419±0.006	724±54
MMFF94	Liq.	0.399	654
MMFF94	Gas	0.398±0.004	744±52

0.438Å, torsion angles almost exactly those predicted by *ab initio*, a $\langle q \rangle$ of 0.437, and a pseudorotational velocity of 639 deg ps^{-1}.

This variation of the force field parameters shows potential, as the difference between q_e for the simulations and experiment went from approximately 0.02 Å to the exact experimental value. However, the mean pseudorotational amplitude had a value that was 0.01 Å too high for both force fields. This is most likely due to other force field parameters playing a role in the cyclopentane dynamics. It would make sense that increasing the torsion parameters to such an unreasonable value would fail to fully replicate dynamic properties while completely agreeing with the minimum potential energy model. However, this does underscore that the torsion parameters do need to be found specifically for this molecule, but not at the expense of ignoring the other parameters. The best way to find the optimal force field parameters, which is beyond the scope of this work, would be to first fit all of the force field parameters to produce the *ab initio* geometries for both low energy conformations in order to match the proper value for q_e. Once this is complete, the force constants should be adjusted to match $\langle q \rangle$ to the experimental value.

2.4.2.5 Condensed Phase Effects

One major assumption that was made for the aforementioned simulations was that cyclopentane was in the ideal gas phase, i.e., that one single molecule that does not interact with anything else is adequate to describe cyclopentane. In order to test whether this is true, simulations were performed at liquid density. A small system was used mainly to decrease

the large simulation time caused by using such a small time step. A comparison of the gas and liquid simulations can be found in Table 8. It shows that pseudorotational amplitudes were obtained that are not statistically different for the liquid and gas phases for both force fields. This is not too surprising, since it has been shown that the torsional contribution to the potential energy of cyclopentane is the most important. Therefore, it is expected that the addition of van der Waals interactions will not substantially effect the cyclopentane dynamics. Because this in a non-polar aliphatic hydrocarbon system, no long range electrostatic forces were included. However, variation of Q to accurately reproduce the bulk thermodynamic and transport effects in such liquid phases systems will be required. Because the local pseudorotation dynamics are not sensitive to Q, such an adjustment is possible for this system.

2.4.2.6 Long-term Simulations

By looking at simulations that run for a very large number of timesteps, the benefits of one MD algorithm over another in the accumulation of error should be apparent. The main reason that this was examined was because the NP algorithm was claimed to be a superior algorithm because it was symplectic and it reduced the amount of the error in the algorithm's conserved quantity. Bond suggested that the accumulation of this REE, discussed earlier, can change the nature of the dynamics over time[41]. Because there are experimental parameters for cyclopentane that can be derived from MD, this is a good test molecule. By calculating these parameters at the beginning and end of the simulation, it can be determined how much this energy error affects the local dynamics of the system.

The first test was a 0.02 fs time step run for 1 μs, and the relative energy error is plotted for each algorithm in Figure 9. After 100 ns, the relative energy error for NH-SV and NP is very stable. The REE for the NH algorithm, however, continuously climbs over the simulated time. Table 9 shows the results of the pseudorotational parameters at the first and last 100 ps of the simulation. From this, it is shown that NP and NH-SV have similar

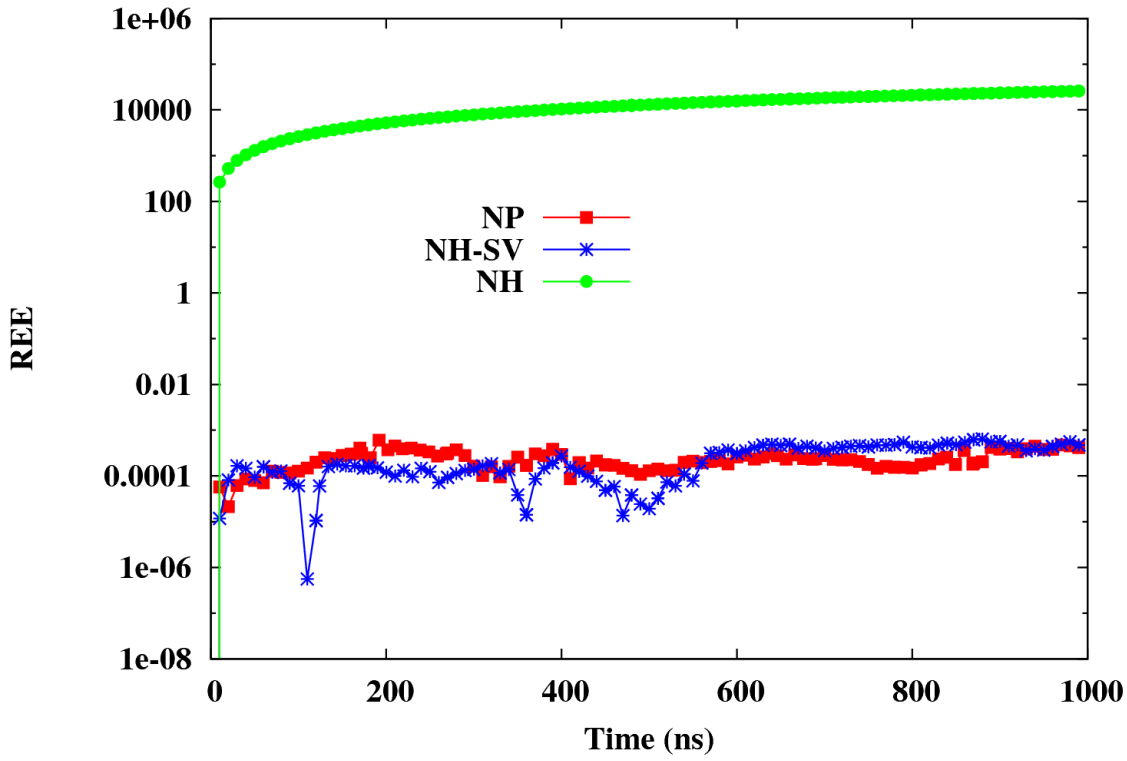


Figure 9: The effect of the MD algorithm on the relative energy error (REE) for NP, NH, and NH-SV over time.

Table 9: Comparison of pseudorotational parameters for the first and last 100 ps of a 1 µs simulation

Force Field		$\langle q \rangle$ (Å)	v_{rms} (deg ps^{-1})
NP	start	0.392	563
	end	0.396	665
	difference	0.004	102
	90% CI	0.004	52
NH-SV	start	0.392	589
	end	0.395	640
	difference	0.003	51
	90% CI	0.002	36
NH	start	0.407	844
	end	0.393	705
	difference	-0.014	-139
	90% CI	0.002	30

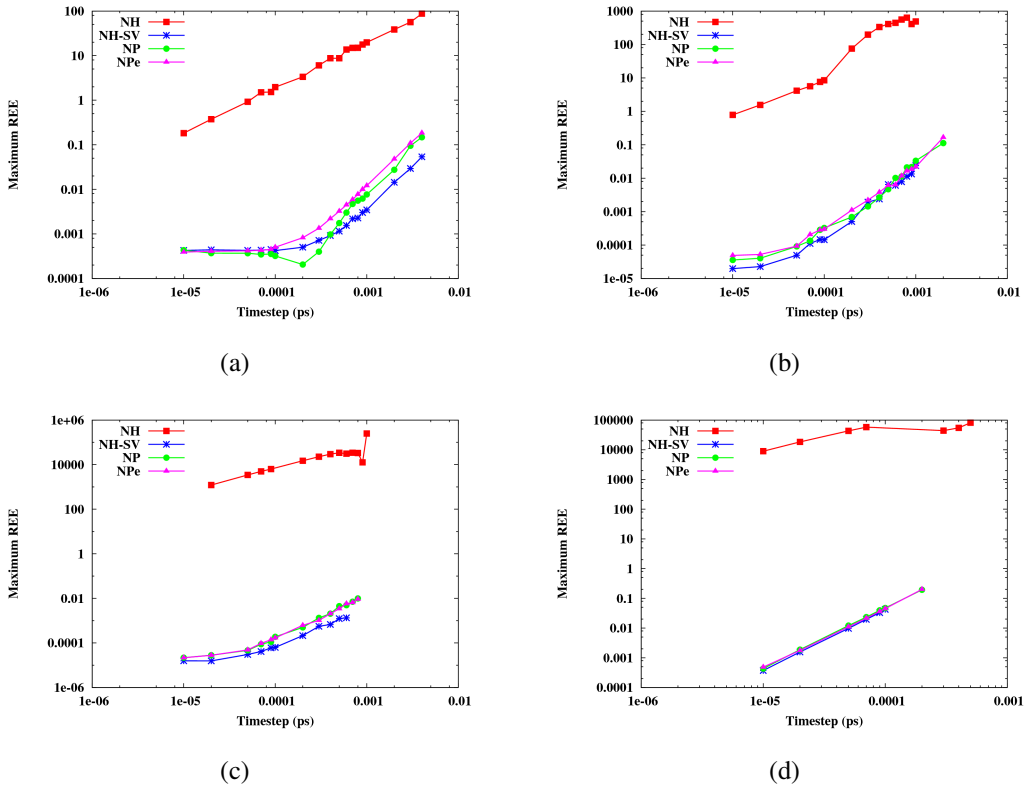


Figure 10: The effect of the time step on the maximum relative energy error for each algorithm on (a) argon, (b) cyclopentane at 300 K, (c) liquid cyclopentane at 300 K, and (d) liquid cyclopentane at 900 K.

results as far as the difference between the beginning and the end of the simulation. For the pseudorotational parameters, each is within overlapping confidence intervals, which indicates that they are not statistically different. For NH, however, the parameters between the first and final 100 ps differ significantly. This would indicate that the NH controller is applying artificial dynamics to the system. It would also indicate that the relative energy error is important because a large drift in REE can cause a change in the dynamics.

The next set of simulations that were performed were on par with those of Bond and coworkers[41] to determine the maximum REE for a given time step. This would show the effect of the time step for each algorithm. Not only was their experiment repeated, Nosé's explicit version of NP (NPe)[63] was added, and the effect of temperature and bonding to the system was also examined. The results can be found in Figure 10. The results for

argon, compared to Bond and coworkers agree reasonably well as far as algorithm ordering, however, the magnitude for NH is different. This is most likely due to the fact that a smaller value for Q is used, and they showed that the maximum REE will change depending on the value of Q. They found that an increased value for Q will cause NH to have similar results to the other algorithms, but this is logical, since this would basically bring all of the algorithms closer to the NVE limit. However, the ordering of the algorithms is the most important, so even with a slightly different Q, the previous data should be matched to a reasonable degree. It should also be noted that NPe has the highest maximum REE among NP, NPe, and NH-SV as the time step increases.

The effect of adding bonds to the system causes the algorithms to be almost imperceptibly different for a small system at a relatively low temperature (cyclopentane at 300 K, Figure 10(b)). However, looking at Figure 10(c), adding more molecules and making the system relatively dense causes this to change. Now, NH-SV is the optimal algorithm, and both NP and NPe display identical results. The effect of changing the temperature can be seen in Figure 10(d), and NH-SV is no longer noticeably better. This would lead to the conclusion that NP is superior to NPe in that it performs better for small systems and identically for larger systems and at increased temperatures. Also, NH-SV will be the optimal algorithm to use at lower temperatures, but as the temperature increases, NP becomes increasingly optimal, most likely due to the stability from its symplectic form. Additionally, and probably most important, the result of looking at different systems to compare algorithms shows that a simple Lennard-Jones system is not enough to adequately compare MD algorithms.

2.5 Summary

Cyclopentane was simulated to determine the proper force field and MD algorithm for simulations of ROMP PNB. This molecule is important because it is the most dynamic part of the polymer backbone. Simulations were performed that compared the results to both

experimental work and previous simulation studies. It was found that the AMBER94 and MMFF94 force fields were the most accurate force fields of those that were compared. It was also found that previous MD simulations that used the Berendsen algorithm used the wrong value for the thermal inertia parameter, so they exhibited dynamics which are inconsistent with experimentally validated cyclopentane behavior. Additionally, it was shown that the REE is an important quantity to be aware of, as MD algorithms which generate high values of the REE will exhibit a non-physical variation in the dynamic behavior of a molecule.

CHAPTER III

MODELING ROMP POLY(NORBORNENE) USING TORSIONAL STATES MONTE CARLO

3.1 Introduction

3.1.1 Objectives

1. Determine the most accurate force field for the simulation of ROMP PNB

2. Use Torsional States Monte Carlo to sample conformation space of ROMP PNB

3.1.2 Motivation

Previously, TSMC was used to sample the conformation space of vinylic PNB to determine the repeat angles in the backbone[23]. Once these repeat angles were found, it was trivial to determine the pattern of the repeat angles and show not only that the polymer was helical, but that the helix had a distinct kink with a specific geometry. Because vinylic PNB has a similar molecular structure as far as linked cyclic groups on the backbone, it is theorized that ROMP PNB could share this helix-kink structure. Applying the same method used with vinylic PNB to this system could yield the same information.

3.1.3 Scope

This work will be confined to choosing a force field for ROMP PNB, as well as using TSMC to examine the backbone conformation of ROMP PNB. Choosing a force field is imperative in obtaining reasonable conformations, however, a force field that reproduces the conformation exactly is not necessary. A force field that produces reasonable conformations, as matched by experiment, will be adequate. Also, TSMC is the only method that will be used, as its usefulness is tested on polymers which may not adhere to the same

conformation type as vinylic PNB.

3.1.4 Background

3.1.4.1 Torsional States Monte Carlo

Torsional States Monte Carlo is an attempt to increase the speed of Metropolis Monte Carlo, mainly for polymer systems. The original Metropolis Monte Carlo method[65] is very robust and commonly used, however, it does have its drawbacks. It can be very time consuming to effectively sample the conformation space of a large system because individual moves must be done. It becomes even slower in systems such as polymers, which have a series of bonded atoms that make the moves even more important to get the right displacement value. The random displacement becomes even harder to put bounds on because atoms in the backbone of the polymer have less freedom than those on sidechains. Also, regardless of whether an atom is on a sidechain or on the backbone, these displacements must be small to be able to be accepted. Essentially, polymers will have a very large increase in energy for even small perturbations that produce an unrealistic conformation. By only focusing on the degrees of freedom that fundamentally change the conformation, extra moves can be avoided.

The molecular motion of polymers is unique compared to that of small molecules. Small molecules, like cyclopentane, have conformations that are dependent on all of the force field terms that are present. While the torsion term was the most important for the cyclopentane ring structure, the bond stretch and angle bend terms also had an impact on the conformation, though not quite as significant as the torsion terms. In order to generate conformations of cyclopentane, MD was used, and all of the force field terms were used to generate conformations. This is similar to most other small molecules. Exploring the conformation space requires the use of all available force field terms.

Polymers are a different case. All of the force field terms are necessary to generate

the initial conformation, i.e., generating an initial conformation requires the correct equilibrium bond lengths, angles, and torsions, along with the proper potential constants associated with each. However, unlike small molecules, exploring the conformation space is mainly dependent on torsional changes, as bond stretch and angle bend motion is very minor compared to that of the torsion angles. Essentially, the conformation space can be adequately explored by deviations in the torsions for polymers.

Torsional States Monte Carlo (TSMC) is a method that is used to take advantage of this assumption for polymer conformations. Both MD and translational Monte Carlo (MC) use all energy terms in the force field, so determining the potential is the most time consuming step for both methods. For MD, decoupling the energy terms from the potential calculation and only solving Newton's equations of motion using the torsion energy would be unrealistic for an all-atom simulation. For MC, it is much easier. TSMC uses the pivot algorithm[66] on top of the Metropolis MC algorithm to sample conformation space. Instead of perturbing atom positions, as would be the case for a standard translational Monte Carlo algorithm, the pivot algorithm perturbs the torsion angles. The algorithm for TSMC is

1. Randomly choose a torsion angle
2. Perturb the torsion angle a random amount ($-\pi < \Phi < \pi$)
3. Perform local minimization of all atoms
4. Determine the potential energy
5. If $\chi \leq e^{-\frac{\Delta U}{k_b T}}$ then

   Accept the move and repeat with the new conformation

   Else

   Reject the move and start over with the original conformation

where χ is a random number between zero and one, Φ is the random angle value, ΔU is the energy difference between the initial state and the state after the move ($\Delta U = U_2 - U_1$), k_b is Boltzmann's constant, T is the temperature, and $e^{-\frac{\Delta U}{k_b T}}$ is the Boltzmann factor. The

Boltzmann factor is designed so that it is always greater than one if ΔU is negative, and it is between zero and one for positive ΔU. This means that the new conformation is always accepted if the potential energy decreases with a move. It also means that the likelihood of acceptance for a move that increases the potential energy goes down as the energy difference increases, and it goes up as the temperature increases.

This algorithm differs from a traditional MC algorithm in one very important way. The system is minimized after every move. Typically, MC is used to generate conformation space for a given temperature, and it correlates very well with the conformation space generated by MD with a thermostat set to the same temperature. With MC, the system is never minimized between moves. The reason for this is that with MC, the goal is to get several different conformations in a similar conformation space. The sampled conformation space will be that of a minima (global or local) with the walls of the energy well greater than k_bT. However, the purpose of this method is to absolutely determine the low-energy conformation, as it is unknown what the actual low-energy conformation is. Therefore, the current conformation is minimized after each move to ensure that the energy that is used to calculate the Boltzmann factor is always the lowest it can be so that the lowest available energy conformation is always found for each move. If the minimization were not performed, then there would be no way for the energy to decrease. One major advantage of this is that it greatly decreases the role of the starting conformation.

For the systems that are being studied here, TSMC seems to be a particularly good choice, as a major advantage to using this method for polymeric systems is that if there are repeating backbone torsion angles in polymers which have ordered conformations (e.g., helical polymers), they will be readily apparent. This is especially useful when the goal is to determine whether or not these repeating backbone angles exist.

3.1.4.2 Vinylic Poly(norbornene)

Previous work on vinylic PNB has shown that for a particular configuration of PNB, the polymer forms a helix-kink conformation. The 2,3-exo,exo-erythro-diisotactic poly(norbornene) configuration was the only one that was found to form this conformation[23]. Because of this, three hypotheses were formed as to what was required of a helix-kink polymer.

The first is that the backbone bond types must alternate between rotating and non-rotating bonds. This is consistent in vinylic PNB with the alternating norbornane ring on the backbone of the polymer. The part of the backbone that is also part of the norbornane ring cannot rotate due to the fact that it is constrained by the norbornane ring itself. If the bond in the backbone was free to rotate, the norbornane ring would have to be flexible enough to allow the rotation. Since the norbornane ring is highly constrained, this bond cannot rotate. The bond that connects norbornane rings is free to rotate.

The second hypothesis is that the polymer must be cis across the backbone. In the case of vinylic PNB, the exo,exo designation means that the norbornane ring attached to the backbone has both bonds on the same side of the ring, i.e., if the bridging carbon was considered the "top" of the ring, both bonds that continue the backbone are facing the top the top of the configuration. If the polymer were exo,endo, the polymer would not form a helix.

The final hypothesis is that the polymer must have a bulky side side group. The best example for this is poly(acetylene). Poly(acetylene) does not have a helical conformation, yet *cis*-t-butyl poly(acetylene) (PTBA)does[67]. The addition of the side group provides the steric hindrance that prevents the torsion angle of the rotatable bond to adopt a more linear conformation. For vinylic PNB, the side group can be considered to be the norbornane ring. Additionally, the 2,3-exo,exo-erythro-disyndiotactic poly(norbornene) configuration did not form a helix, which leads to the conclusion that the side group needs to alternate sides of the backbone.

There are a class of polymers that adopt this helix-kink conformation. The most notable

example is poly(1-trimethylsilyl-1-propyne) (PTMSP). This polymer is used as a membrane in gas separation applications, mainly because of its high reverse selectivity. This property is most likely due to the fact that it also forms a helix-kink conformation[68]. All of these types of polymers, including this example, are used for special applications that are enabled due to properties that are present from the polymer adopting this helix-kink conformation.

One additional property that this type of polymer could provide is accessibility. If a catalyst is tethered to a polymer support, a random coil polymer has the potential to block the catalyst site to the reactant. A random coil polymer has no underlying structure, so it is possible that the catalyst site could be on the interior of the polymer molecule. This is similar to how an active site on an enzyme must be accessible to a substrate. If a polymer were helical, then there would be an overall order to the polymer backbone. This would make it less likely that a catalyst site would be inaccessible due to the backbone of the polymer.

TSMC was used previously on vinylic PNB with very positive results. The results showed very distinct repeatable angles at ±120° torsion angles for the rotatable single bond on the backbone as moves were performed. They also showed that the accepted moves transitioned between these two angle values. A kink that moved along the chain as moves progressed was responsible for the torsion angle pattern changing from 120° to 240°. Applying this same procedure to ROMP PNB could yield a similar result.

3.1.4.3 ROMP Poly(norbornene)

ROMP PNB has a very different backbone compared to its vinylic counterpart. Rather than an edge of a norbornane ring as part of the backbone, ROMP PNB contains a cyclopentane ring that is integral to the backbone. However, it is possible that one of the configurations of ROMP PNB would fulfill all of the hypotheses for a helical polymer. For the polymer itself, there are three different possible configurations in the backbone. The double bond

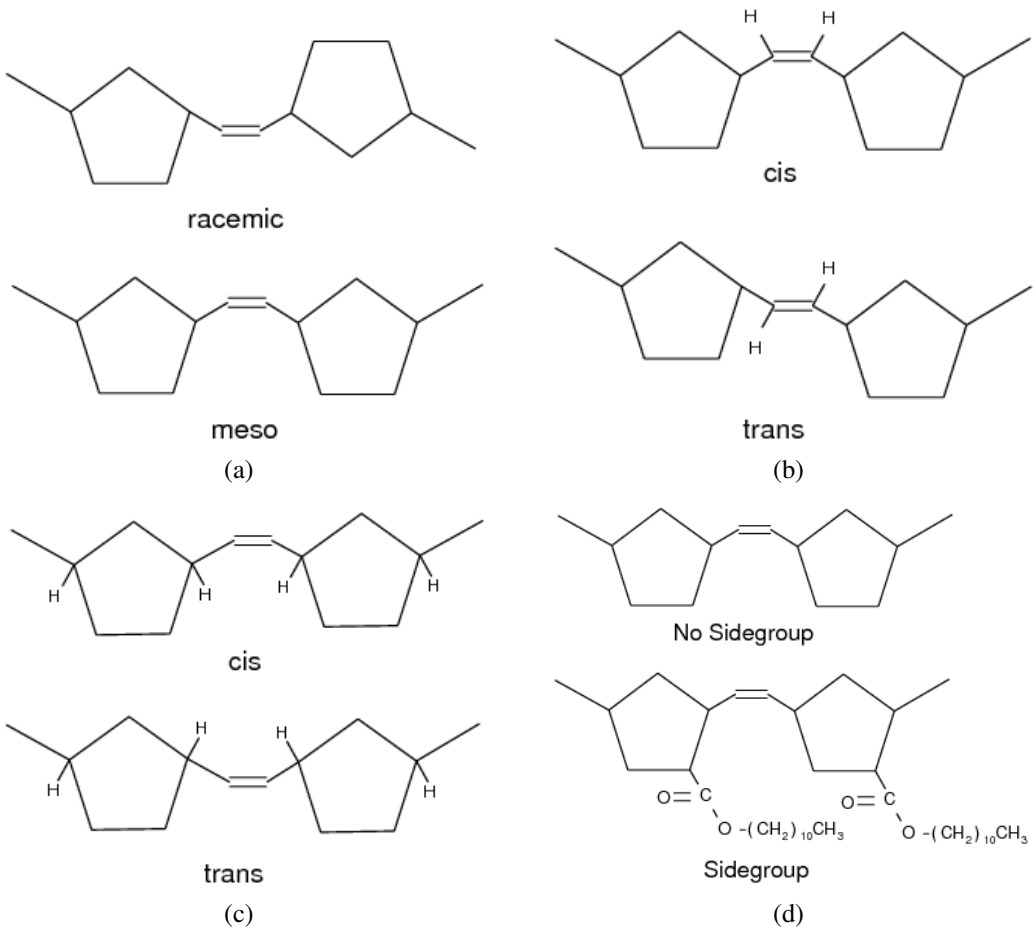


Figure 11: The different configurations of ROMP PNB. ROMP PNB can have (a) racemic or meso diads, can be (b) cis or trans across the double bond or (c) the cyclopentane ring, and can have (d) an attached sidechain on the cyclopentane ring.

can be cis or trans, the backbone can be cis or trans across the cyclopentane ring, and the polymer could have meso or racemic diads (Figure 11). Additionally, to test the bulky side group hypothesis, a side group can be added to the cyclopentane ring. The side group shown in Figure 11(d) was chosen because it can be found in the literature as a side group which was used to tether catalysts to ROMP PNB[69, 70].

From the three hypotheses for a helical polymer, there is a possibility that ROMP PNB can fulfill all three of them. The first characteristic of ROMP PNB that is important is that the backbone alternates as double bond / single bond / cyclopentane ring. If the cyclopentane ring is rigid enough, this would constitute the required alternating rotatable / non-rotatable bonds. The cyclopentane ring will always be cis due to the fact that the ROMP mechanism only acts on the double bond of the norbornene ring. This has been validated experimentally with a Ruthenium-based (Grubbs) catalyst[71]. As long as the double bond is cis, then the all cis non-rotatable bond hypotheses will be fulfilled. A side group will be necessary for the third hypothesis, and racemic diads will ensure that the side group alternates sides of the backbone. Additionally, there is evidence that the configurations could produce different conformations due to the fact that the double bond cis/trans tacticity produces polymers with different melting points[72].

For the remainder of this work, these configurations will be referred to in an abbreviated notation. The backbone is always assumed to be cis across the cyclopentane ring, so cis/trans will only designate the geometric isomerism of the double bond. This will be given as a *C* for cis and a T for trans. Meso and racemic diads will be designated by *M* and *R*, respectively.

3.2 Simulation Methodology

3.2.1 Force Field

From the previous work on cyclopentane, it was found that either MMFF94 or AMBER94 were adequate force fields. However, their performance for polymers like ROMP PNB

was unknown, as they were parameterized with proteins and nucleic acids, respectively. Poly(*cis*-1,4-butadiene) was used to test the behavior of each force field. This particular polymer was used for two reasons. First, it has a known, experimentally determined radius of gyration. Second, it is comprised solely of single and double bonds on the backbone, which is the remaining part of the ROMP PNB backbone now that force fields have been evaluated for cyclopentane.

Poly(*cis*-1,4-butadiene) was modeled using MD with chains of 5, 10, 20, 30, 50, and 100 monomer units. Each chain length was simulated once as one isolated chain. MD was performed using the NH-SV algorithm with a timestep of 1 fs at a temperature of 300 K. The hydrogens were constrained using the LINCS algorithm, and Q was set to 1 kcal ps mol^{-1}. The simulation was sampled every 0.5 ps. These simulations were repeated using both AMBER94 and MMFF94, and the unperturbed radius of gyration (s_0) was recorded for each sample. These simulations were performed using version 2005.06 of the Molecular Operating Environment (MOE) from the Chemical Computing Group (Montreal, Canada, http://www.chemcomp.com).

It is important to discuss the major assumption of simulating an isolated polymer chain. Basically, it is assumed that the polymer is in the unperturbed, or θ, state. Normally when polymers are in solution, there are intramolecular effects, generally from the bulk, that cause changes in the polymer's conformation. Much of this is due to excluded volume, which contributes to an increase in the distance between the ends of a polymer in solution. The θ-state corresponds to the case when these effects are negated, and the polymer is considered unperturbed. A solvent which leaves a polymer unperturbed is known as a θ-solvent. Generally, unperturbed dimensions are denoted by a subscript 0.

Once all of this data was collected, it was used to determine the limiting value of the characteristic ratio (C_∞) of poly(*cis*-1,4-butadiene) from each force field. This is done by plotting C_n versus n^{-1}, where C_n is the characteristic ratio and n is the number of monomer units. A simple model of ideal polymer chains will be used to determine the characteristic

ratio for each force field, and the force field with the best match to the experimental data will be used for future simulations.

The radius of gyration is generally defined as

$$s^2 = \frac{\sum_{i=0}^{N} m_i s_i^2}{\sum_{i=0}^{n} m_i}, \tag{16}$$

where m_i is the mass of a single backbone atom and s_i is the magnitude of its distance from the chain's center of mass. The mean square radius of gyration ($\langle s^2 \rangle$) is an experimental observable that characterizes a polymer's dimensions, which makes it a useful parameter. However, in most polymer models, the unperturbed mean square end-to-end distance ($\langle r^2 \rangle_0$) is used to calculate the characteristic ratio by

$$C_n = \frac{\langle r^2 \rangle_0}{nl^2}, \tag{17}$$

where l is the length of the bonds in the backbone. C_∞ is the specific limiting value as n approaches infinity ($C_\infty = \lim_{n\to\infty} C_n$). The characteristic ratio is indicative of the increase in actual chain dimensions compared to the equivalent unperturbed chain of the same bond lengths. One model that describes a relationship between the radius of gyration and the end-to-end distance is a random walk model. This model involves a polymer chain where each successive step has equal probability and is independent of previous steps. For a random walk chain, the unperturbed mean square end-to-end distance is related to the unperturbed mean square radius of gyration ($\langle s^2 \rangle_0$) by

$$\langle s^2 \rangle_0 = \frac{\langle r^2 \rangle_0}{6}, \tag{18}$$

and this is considered true for all polymers of "sufficient" length. To determine the characteristic ratio for AMBER and MMFF94, the freely jointed chain model was used. This model is a random walk chain which has a fixed bond length. This gives a relationship between $\langle s^2 \rangle_0$ and $\langle r^2 \rangle_0$ that is

Table 10: Data for the characteristic ratio for different polymer sizes for poly(*cis*-1,4-butadiene) using AMBER and MMFF94.

Chain Length	AMBER C_n	MMFF94 C_n
5	8.1	10.1
10	6.8	8.2
20	6.0	6.3
30	4.8	5.7
50	3.9	4.3
100	3.4	4.1

$$\langle s^2 \rangle_0 = \left(\frac{n+2}{n+1} \right) \frac{\langle r^2 \rangle_0}{6}, \tag{19}$$

which holds true for any value of n and will recover a random walk chain (Equation 18) as n approaches infinity[11].

It was earlier mentioned that plotting C_n versus n^{-1} yields the limiting value for the characteristic ratio. This is done by extrapolating this data using linear regression to $n = \infty$, or $n^{-1} = 0$. Essentially, the intercept found from this linear fit is C_∞. The results from these simulations for poly(*cis*-1,4-butadiene) with AMBER and MMFF94 can be found in Table 10. The values used for l were 1.485 and 1.475 Å for AMBER and MMFF94, respectively. This value was found by averaging all of the bond lengths over all of the simulations for each force field. This also includes the double bond in the backbone, which is why the value is smaller than the roughly 1.5 Å bond length for an $sp3$ carbon, but larger than the roughly 1.3 Å bond length for a double bond.

The experimental value for C_∞ for poly(*cis*-1,4-butadiene) is 4.9 ± 0.2, and this was calculated from experiments performed near 300 K[73]. C_∞ for each force field was calculated by performing linear regression on the simulation data. For AMBER, this value is 3.92, and it is 4.27 for MMFF94. From this data, a force field can be chosen for PNB. The previous cyclopentane work showed that AMBER was the optimal force field to model cyclopentane, however, MMFF94 was a very close second. Modeling a structure that has

a backbone which is similar to the rest of the ROMP PNB backbone (single and double bonds) yields a larger difference between the two force fields for the characteristic ratio. In this case, MMFF94 yields structures that are 0.35 closer to the experimental value of C_∞. Because of this, MMFF94 was chosen as the force field to model ROMP PNB.

3.2.2 Torsional States Monte Carlo

All of the TSMC simulations were performed using version 2005.06 of the Molecular Operating Environment (MOE) from the Chemical Computing Group (Montreal, Canada, http://www.chemcomp.com). All of the combinations of ROMP PNB (cis / trans, meso / racemic, with / without sidechains) were modeled by this technique, though not all simulation types were performed on all of the combinations. Some simulations were performed simply as proof-of-concept and so did not include all of the combinations.

Due to the work with the force fields, MMFF94 was chosen as the force field for all of the simulations. The energy minimization process in MOE is the sequential use of three minimization algorithms. Each method will be described below by accuracy and computational time. Methods that are considered accurate are those that can reach the actual minima with the greatest confidence. Computational time involves the length of time that an algorithm will run before it is solved. In general, computational time and accuracy are inversely related, i.e., methods which have increased accuracy take more time to compute each iteration.

The first of the three methods used is the fastest and least accurate method, called Steepest Descent. This algorithm minimizes the system by ensuring that the gradient of the energy is always negative by always moving in a direction that decreases with energy. In this case, the minimization gradient is the change in system energy per change in position vectors. Next, the conjugate gradient method is used. This algorithm is similar to steepest descent, but it performs an iterative optimization that makes it more accurate, though slower, than steepest descent. Finally, a truncated Newton method is used. The truncated

Newton method is slightly different from the previous two methods, which only use the gradient to calculate the next step. The truncated Newton method uses both the gradient and the derivative of the gradient to determine the next step of the iteration. It is the most accurate method, as it avoids saddle points, but due to the calculation of the second derivative, it requires the most computational resources[74].

For each method, there is both a gradient cutoff value and an iteration cutoff value, and whichever is reached first is the cause for either switching to the next method or, in the case of the third minimization method, ending the minimization altogether. The gradient cutoff is the gradient at which the method will stop once it is reached. A value that is too large will return conformations that would not be considered realistically minimized. Unfortunately, the desired value, which is infinitely small, cannot be used because each full minimization of a structure would take an unacceptable amount of time. The iteration cutoff value is the number of iterations a given method will perform. A very low number of iterations will not allow the method to reach the desired gradient. A value which is too large will allow an unrealistic starting conformation too long to find a minimum. The optimal values of each depend on both values, as well as the goal of minimization. In this case, the goal of the simulation is the low energy conformation. Therefore, the iteration cutoff was such that it would never be reached by a method. The gradient cutoff was determined in a different manner. The gradient cutoff must be such that the minimized structure is indeed minimized, otherwise conformations could be rejected which are actually lower in energy but are not fully minimized. However, the other criteria is that it must be set as high as possible to decrease the overall simulation time.

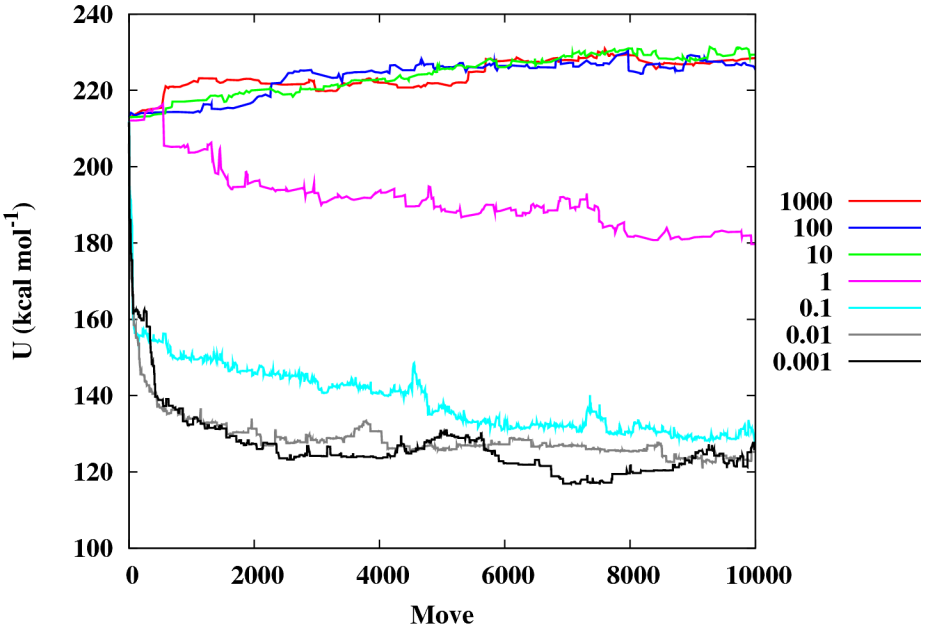
3.2.2.1 Gradient Cutoff Determination

The gradient cutoff that was used was found by running TSMC on the same starting conformation using different minimization parameters to determine the effect of the gradient cutoff. Theoretically, there will be a point where there will be a negligible difference in

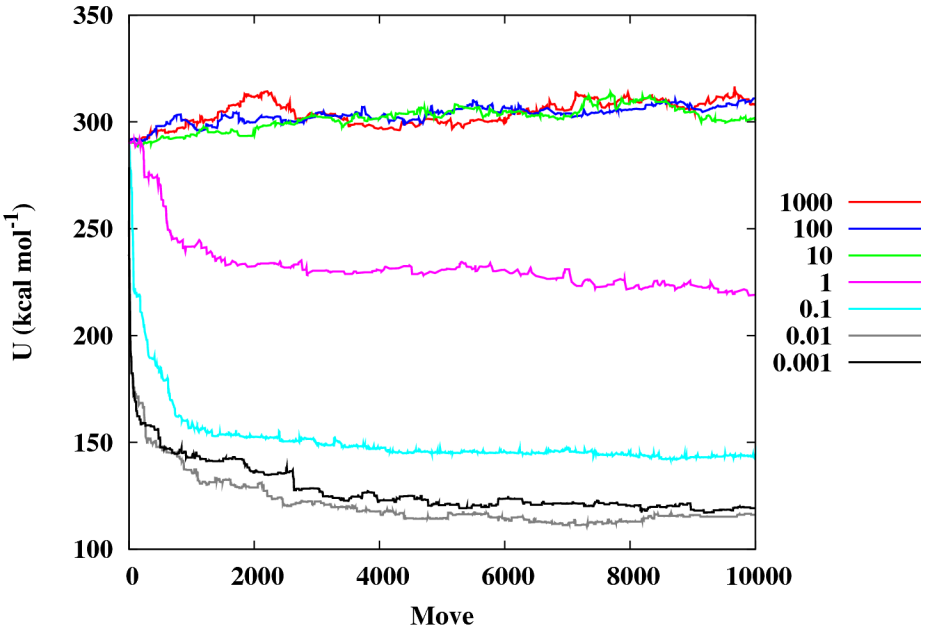
equilibrated energy between gradient cutoffs. The purpose for finding the highest possible gradient cutoff is that going to lower cutoffs required more iterations per method, which requires more computational time. In each simulation, the number of iterations for each method was chosen to be 4000. It is assumed that 4000 iterations for each method would allow the method to reach the gradient cutoff. For most simulations, the maximum number of iterations is usually 400-500, and this is generally considered a large number of iterations for systems with reasonable starting conformations. By increasing this value by an order of magnitude, it is assumed that iteration cutoff will never be reached.

The gradient cutoffs for each method were chosen based on the truncated Newton method. This method had cutoff values of 1000, 100, 10, 1, 0.1, 0.01, and 0.001 kcal mol^{-1} $Å^{-1}$. Since the first two methods are less accurate and much faster, they were not optimized. They were chosen to be 1000 kcal mol^{-1} $Å^{-1}$ for steepest descent and 100 kcal mol^{-1} $Å^{-1}$ for conjugate gradient. When the truncated Newton cutoff was 100 kcal mol^{-1} $Å^{-1}$ or greater, than the cutoff for the conjugate gradient method was increased to twice the cutoff of the truncated Newton method, and the cutoff for steepest descent was increased to ten times the cutoff of the truncated Newton method.

This cutoff analysis was performed on 20 monomer unit CR and TM polymers with side groups that had helical starting conformations. These two were chosen because they have complete opposite configurations, and the side group will produce the largest systems. An additional CR polymer with side group was also modeled that had an elongated starting conformation. This was done to see the effect of the gradient cutoff on different starting conformations. The resulting energy value from an accepted conformation was used to characterize the "quality" of the conformations generated by each cutoff. The cutoff which sampled the lowest energies is the one that is considered optimal. If multiple cutoffs sampled the same conformation space (had similar energy values), the optimal gradient is one with the highest value. The overall optimal cutoff is the one with the highest value that is common for each case.

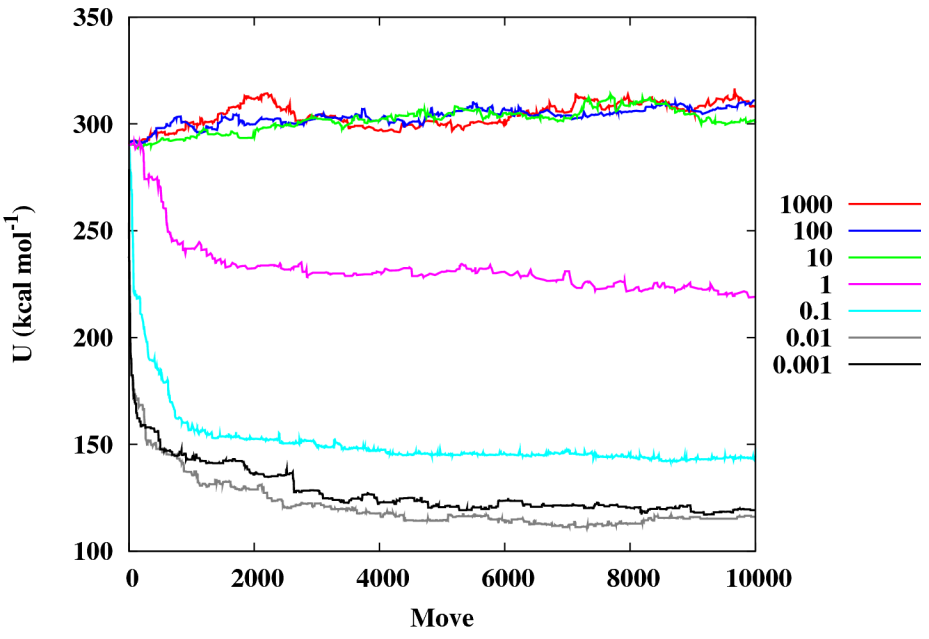


(a)



(b)

Figure 12: The effect of the gradient cutoff on (a) CR, (b) CR with an elongated starting conformation, and (c) TM, all with sidegroups.



(c)

Figure 12: The effect of the gradient cutoff on (a) CR, (b) CR with an elongated starting conformation, and (c) TM, all with sidegroups, *cont.*.

The TSMC runs were set up such that 10000 moves were performed for each of the three systems. The results are shown in Figure 12. Each system is slightly different, but there are some similarities that are carried through each. Gradients between 1000 and 10 kcal mol^{-1} $Å^{-1}$ either showed no change or an increase in the energy over the course of the simulation. A gradient value of 1 kcal mol^{-1} $Å^{-1}$ consistently showed a decrease in energy, but the decrease did not approach that of the lower gradient values. A gradient of 0.1 kcal mol^{-1} $Å^{-1}$ came close to that of the lower values, but the potential energy values were still between 10-30 kcal mol^{-1} from energies obtained by the lower gradients.

Beyond the conformational results, there are two more variables that are involved in choosing the proper gradient cutoff to use. The first is acceptance ratio, and the second is simulation time. The acceptance ratio is simply the number of accepted moves divided by the total number of moves, and this value is often converted into a percentage. In traditional MC, this value should be around 50%. Because TSMC removes translational movement

Table 11: Acceptance ratios for each gradient cutoff used on each configuration tested.

Gradient kcal mol^{-1} Å^{-1}	CR	Elongated CR	TM
1000	1.9%	5.8%	2.6%
100	2.0%	5.7%	2.6%
10	2.4%	5.2%	2.7%
1	2.5%	3.8%	4.1%
0.1	11.3%	11.0%	11.1%
0.01	22.1%	22.1%	24.2%
0.001	25.3%	22.6%	23.6%

Table 12: Simulation time (hours) for each gradient cutoff used on each configuration tested.

Gradient kcal mol^{-1} Å^{-1}	CR	Elongated CR	TM
1000	3	2.5	3
100	4	3	4
10	11	6	9
1	34	18	60
0.1	264	192	216
0.01	456	696	576
0.001	672	696	672

and only rotates about a bond, a smaller percentage is acceptable. Table 11 shows the acceptance ratios for each of the tested configurations for each of the gradients. For the most part, the acceptance ratio increases as the gradient decreases. This is expected, as minimizing to a lower gradient can generate conformations that are lower in energy and more likely to be accepted. The most interesting part of these results is that the smallest two gradients, 0.01 and 0.001 kcal mol^{-1} $Å^{-1}$, have comparable acceptance ratios. The simulation times for each of the TSMC simulations are listed in Table 12. These increase as the gradient decreases, but they increase exponentially as the gradient decreases. For the elongated CR configuration, the simulation time is the same for both the smallest gradients.

The real choice for the gradient cutoff value is between 0.01 and 0.001 kcal mol^{-1} $Å^{-1}$. From the TM configuration, it is easy to see that both are comparable, with the difference in potential energy deviating by 5-7 kcal mol^{-1}, or about 7%. For the two CR conformations, they both show almost overlapping performance between the two. Additionally, there is little to no gain in acceptance ratio for the smallest gradient, and the simulation time is generally shorter for the 0.01 kcal mol^{-1} $Å^{-1}$ gradient. This shows that the 0.01 kcal mol^{-1} $Å^{-1}$ gradient is adequate for the purpose of these simulations. In general, a difference of 10% or less between potential energies obtained from different gradients can easily be explained by the increased number of iterations resulting from the additional accuracy from the C-H bonds and the sidechain carbons. The backbone carbons should be similarly minimized for each gradient, as these would account for higher potential energy differences if they were to be different. In this case, sidechain and C-H minimization differences can be ignored, since the main focus is the conformation from the backbone.

Using these results, the TSMC simulations were run with gradient cutoffs of 10, 0.1, and 0.01 kcal mol^{-1} $Å^{-1}$ for steepest descent, conjugate gradient, and truncated Newton, respectively. The cutoff for the number of iterations was set to 4000. As can be seen in Figure 12(c), 10000 moves are not enough to achieve the low energy conformation region, so 20000 moves were used to ensure low energy conformation space. Equilibration will be

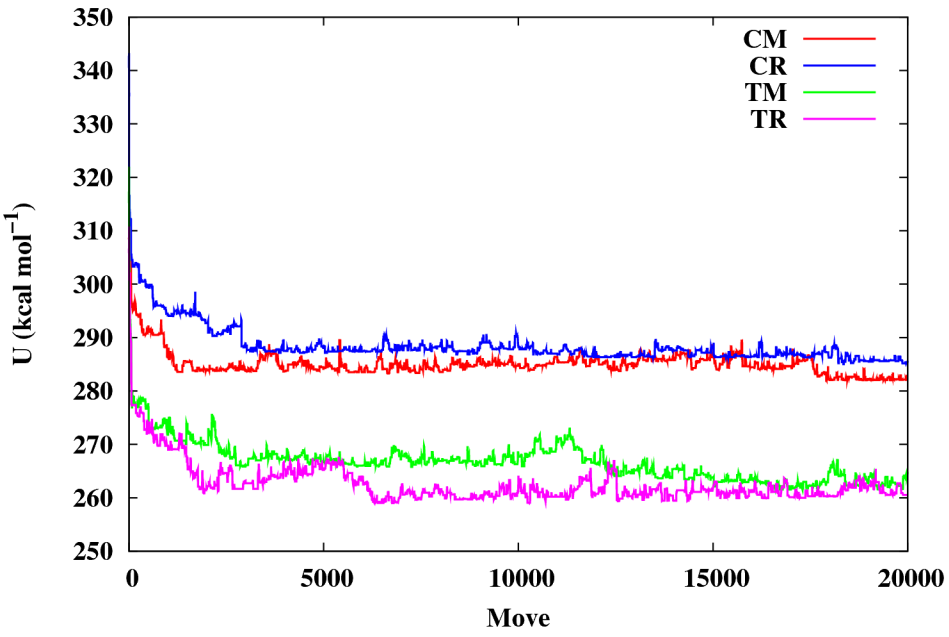


Figure 13: Potential energy as a function of TSMC move for all four configurations of ROMP PNB with no sidechains.

enforced by making sure that a stable potential energy value is maintained for at least 1000 consecutive moves.

3.3 Results and Discussion

Figure 13 shows the potential energy for each of the simulations of the ROMP PNB with no sidechains. All of the plots show that each configuration did indeed reach an energy plateau. The acceptance ratio was similar for each at 20-25% acceptance. However, there was definitely a difference between the final energy levels depending on the configuration. These results show that there are fundamental energetic differences between the configurations. Both the meso and racemic configurations have roughly the same final potential energy and are only offset by whether the polymer is cis or trans. Additionally, the separation of these two energy levels is 20 kcal mol^{-1}, which is a significant energy separation. Figure 14 shows snapshots of backbones of the final conformations of these polymers. While the trans configuration has a lower potential energy, all of the configurations had similar final

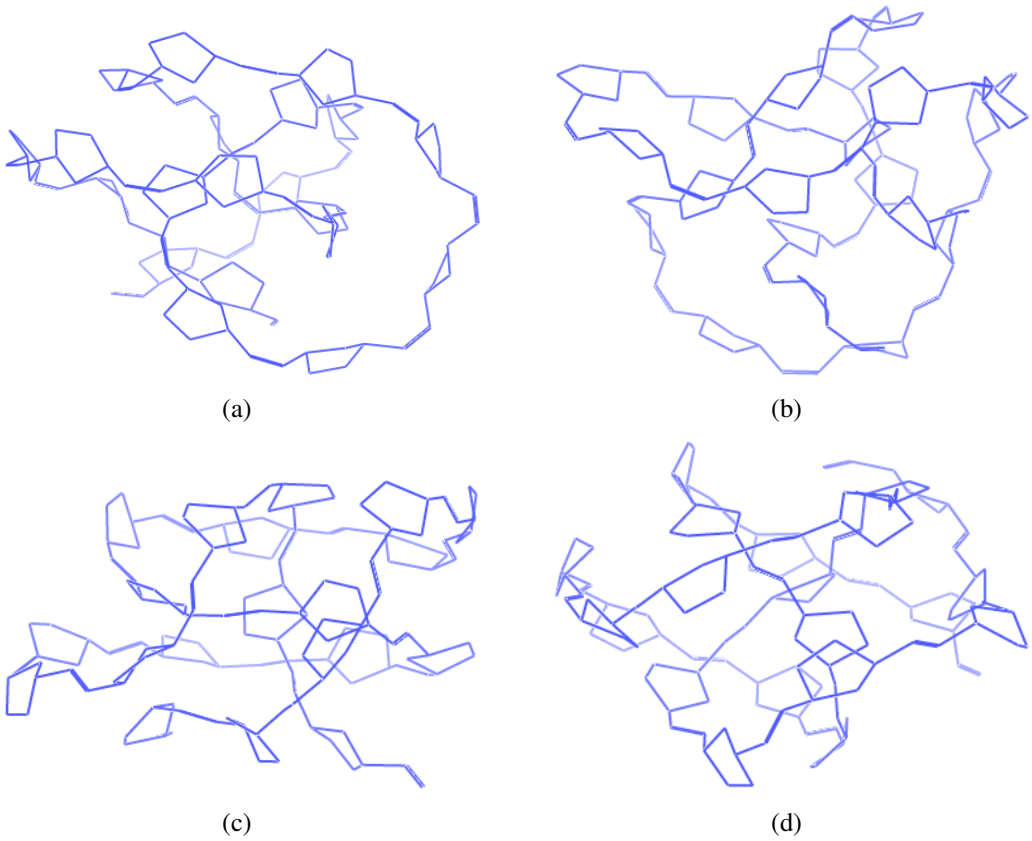


Figure 14: Final conformations obtained from TSMC simulations of ROMP PNB with no sidechains. The conformations were obtained for (a) CM, (b) CR, (c) TM, and (d) TR.

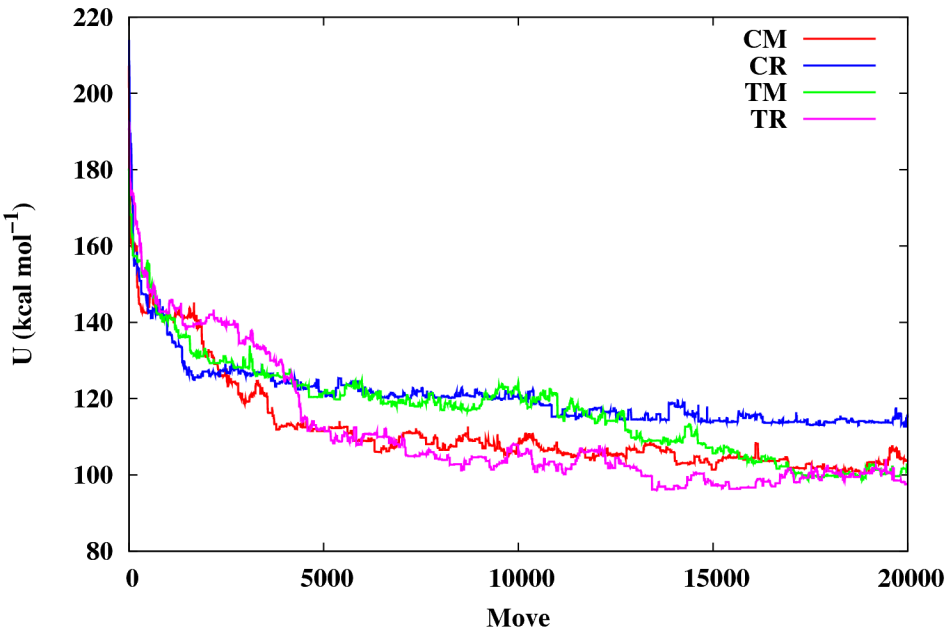


Figure 15: Potential energy as a function of TSMC move for all four configurations of ROMP PNB with sidechains on each monomer unit.

conformations of random coils. This occurred even though the starting conformation was that of a helix. This result backs the previous hypothesis that a bulky side group is required for a helical polymer.

Figure 15 shows the potential energy for all of the configurations with the side group attached to the cyclopentane ring. There are some interesting deviations from the results obtained without sidechains. The first is that all of the potential energy values are lower than that of the previous results. The second is that there still is a split in potential energy values for the final conformations, but it is smaller by half (10 kcal mol^{-1}), and it is now not as simple as cis and trans. For the polymers with sidechains, only the cis racemic polymer has a higher energy than the other configurations, and all of the other configurations have the same final potential energy. Figure 16 is also very different from the previous results. It is apparent from the backbone pictures that something changes when sidechains are added to the polymer. Both of the meso configurations are definitely helical. The trans racemic configuration *could* be a very loose helix, and the cis racemic configuration looks to

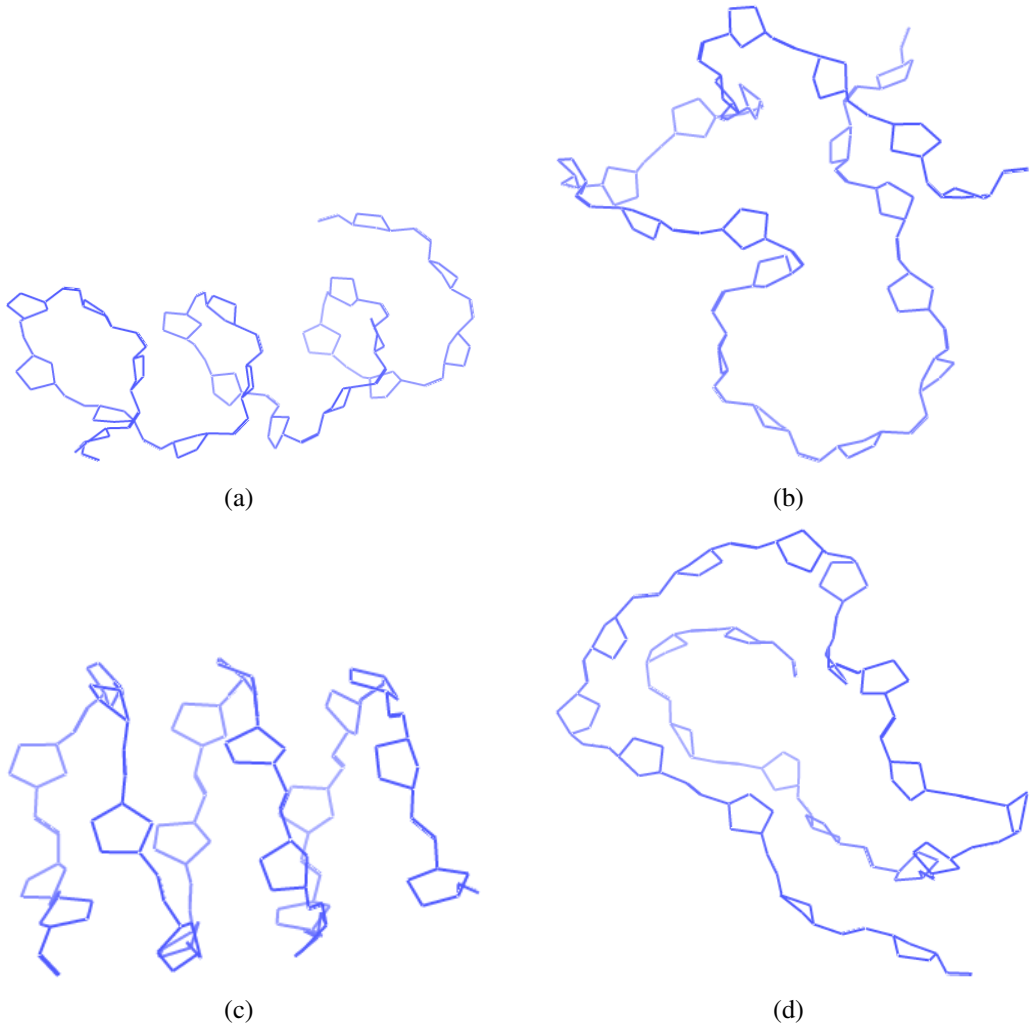


Figure 16: Final conformations obtained from TSMC simulations of ROMP PNB with sidechains on each monomer unit. The conformations were obtained for (a) CM, (b) CR, (c) TM, and (d) TR.

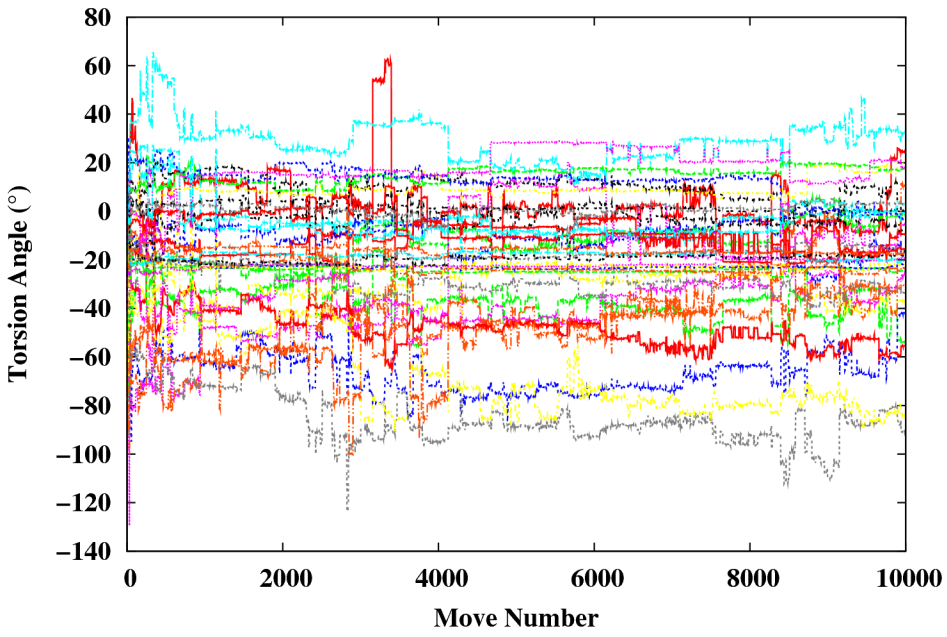


Figure 17: Torsion angles of rotatable single bond on the backbone of the CR configuration of a 20 repeat unit ROMP PNB chain with side groups from a TSMC simulation.

have no helical character whatsoever. However, these results are inconclusive from visual inspection.

Figure 17, which is a plot of the backbone rotatable bonds of a ROMP PNB chain, shows a very different result compared to the results from vinylic PNB. Unlike vinylic PNB, it is apparent that there is no repeatable torsion angle in the backbone of ROMP PNB. This lack of repeatable torsion angles very strongly suggests that ROMP PNB does not form a helix-kink conformation.

3.3.1 Disadvantages of TSMC

These results highlight both the strengths and weaknesses of torsional states Monte Carlo. This method was designed specifically for polymers to quickly determine the low energy conformation and sample that conformation space at a given kinetic energy (temperature). For the configurations without sidechains, it did this very well. These simulations were fast, finishing 20000 moves in less than 24 hours, and the results were independent of

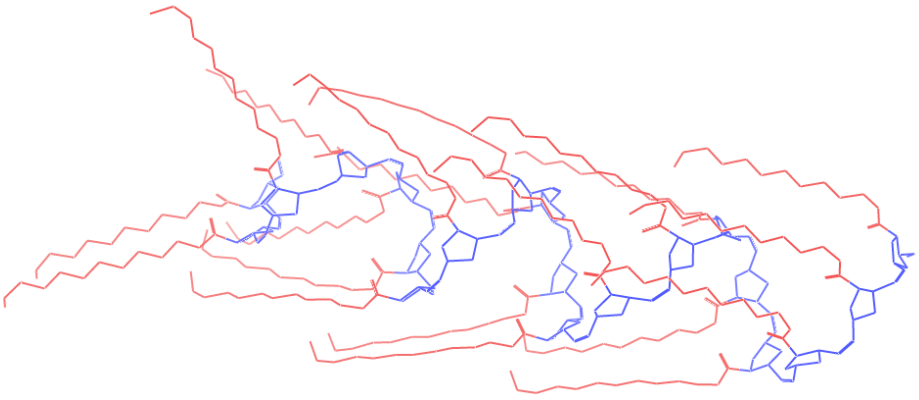


Figure 18: Starting conformation of CM configuration. Blue represents the backbone, and red represents the side chains. Hydrogens are not shown.

starting conformation. When sidechains were introduced, however, two major flaws in this approach arose.

The first is that the method was unmodified from that of the polymers without sidechains, i.e., only the backbone torsion angles were changed. This is acceptable when there were no sidechains, as the only interactions involved are those of the actual backbone. When sidechains were added, especially in this case due to their length and flexibility, this created a problem. Since they were not adjusted, the starting sidechain orientation was fixed with the initial conformation generation. As the simulation progressed, the sidechains were not affected, and so the conformation of the polymer now has interactions with both the the backbone and a long, fixed sidechain. In the case of the meso configurations, the starting conformation was such that all of the sidechains had the same orientation, as shown in Figure 18. The consequence of not perturbing the side chain atoms was that it was impossible for an acceptable move to exist that would change the torsion angle of the backbone, as this would create a high energy state that would never be accepted at the allowable acceptance energy. While the results for the meso configurations could be correct, these simulations may not represent the "correct" final conformation.

The second problem with this method is the amount of time necessary to run each simulation. For polymers of only 20 monomer units to make 20000 moves, or 20000

minimizations, the computational time went from less than 24 hours for a system with no sidechains to over two weeks for polymers with sidechains on machines that used 32-bit Intel Xeon processors. The speed decrease is wholly a function of the number of atoms in the system. In this case, sidechains add an additional 700 atoms, which is more than double the number of atoms in the polymer backbone. For this type of analysis, the time increases by roughly the number of atoms squared. This time increase for each simulation is significant enough to make repeated simulations to test for different possible conformations too costly to perform for multiple polymers.

The method could be modified in different ways in an attempt to fix these flaws, but they are all too disadvantageous to use. The gradient cutoff could be increased, but this risks the integrity of the simulation, as described in the previous section. The sidechain torsion angles could and should be included in the list of rotatable bonds that are used to determine the moves. However, this would increase the number of moves/minimizations that would be required before the simulation can be considered equilibrated, which would greatly increase the total required time for each simulation. The simulation could be run at a higher temperature, which would increase the acceptance ratio, but it would decrease the likelihood of obtaining a realistic conformation.

A few attempts were made to adjust the method parameters to try to produce faster simulations. The first was to change the gradient cutoffs for the steepest descent or the conjugate gradient methods, but this had little effect, as the majority of the computation time is required for the iterative Newton method. Minimizing the system every 10 moves was attempted, but this only served to keep the energy at the starting level. The energy of the system would increase with each move, and the minimization would be only adequate enough to bring the energy back to the starting level with a small amount of decrease. The energy decreased over the number of moves, but the same number of total minimizations would be required to achieve the same energy level.

The only change to the method that had a positive effect was decreasing the cutoff

number of iterations from 4000 to a more realistic value. However, for these systems, the number of iterations per minimization ranged from 200-400, with rejected moves generally having a higher number of iterations. This would decrease the acceptance ratio, but it would increase the speed of each simulation. This was not utilized, however, because with the addition of the sidechain torsion angles to the list of possible moves, the speed gained from the lower iteration cutoff would be offset by the increased required number of minimizations.

3.4 Summary

TSMC was used to find the low energy conformation of ROMP PNB. It was determined that MMFF94 would be the force field used, which was based on simulations of poly(*cis*-1,4-butadiene). An appropriate gradient cutoff for the minimization was found through trial simulations which varied the gradient cutoff. The results of these simulations showed that the polymers without side groups were random coils. However, the results of the simulations which added side groups were inconclusive due to the fact that there was little conformational change from the starting conformation. This most likely arises from not perturbing the side groups in the simulation. These results showed that TSMC is not applicable to this specific polymer system. This is mainly because both the number of atoms is very high and the addition of side groups will complicate the method.

CHAPTER IV

DETERMINING ROMP PNB ACCESSIBILITY

4.1 Introduction

4.1.1 Objectives

1. Determine a method to mathematically characterize the accessibility of a catalyst site attached to a polymer

2. Use that method to establish the accessibility of the various configurations of ROMP PNB

4.1.2 Motivation

As demonstrated with TSMC, it is difficult to fully establish visually whether or not a polymer is helical. Also, the torsion angle across the cyclopentane ring may or may not be the same for all monomer units, nor is it always the same between the cyclopentane ring and the double bond. With the lack of common repeat torsion angles, it is difficult to characterize the conformation of the backbone. For accessibility, it becomes even more difficult. Visually establishing accessibility will be difficult at best, and obtaining agreement on whether or not a polymer should be considered helical will be almost impossible for all but the most obvious case. Therefore, a systematic mathematical method needs to be utilized to determine a quantitative metric for the accessibility.

4.1.3 Scope

The overall goal is to develop a general method which quantitatively characterizes accessibility. This method will be applied solely to the various ROMP PNB configurations as the polymer support, and the accessibility determination method will be limited to simple catalytic systems that are as general as possible. This means that the actual tethered catalyst is

inconsequential, and it is assumed that the catalyst mechanism will only involve one active site and a single reagent. Additionally, it is assumed that the reaction chemistry does not affect the actual support.

An additional goal of this work is to examine the actual conformations of the different ROMP PNB configurations. Since this was inconclusive for the meso configurations with the TSMC simulations, it is desirable to see if a helical conformation is obtained with ROMP PNB using MD.

4.1.4 Background

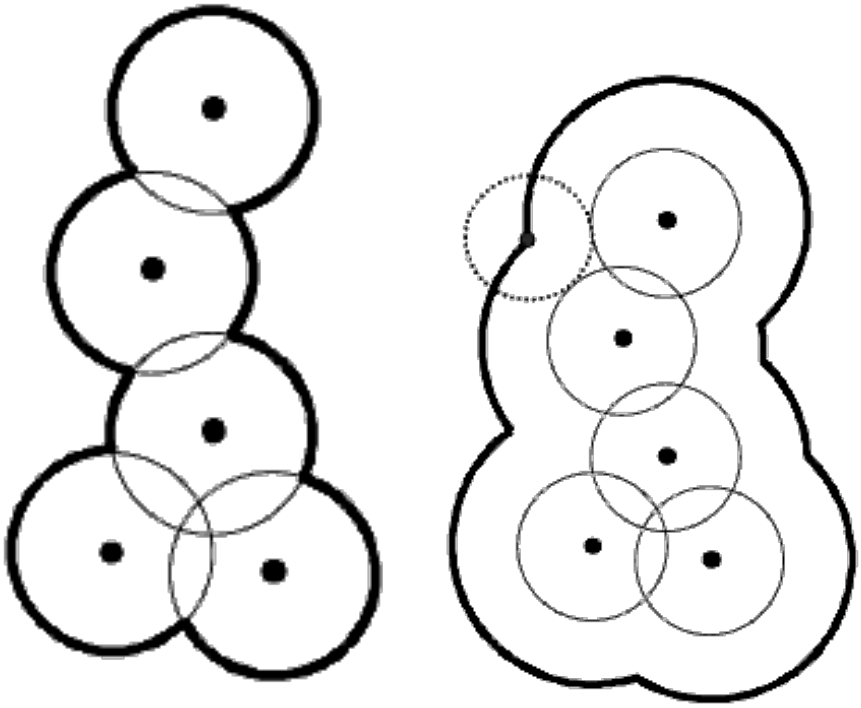
Traditional homogeneous catalysis is marked by reaction rate as the rate limiting step. Diffusion from the bulk is generally not a concern. With heterogeneous catalysis, diffusion of reactants from the bulk to the catalyst site becomes the rate limiting step. By immobilizing a homogeneous catalyst, the high reactivity and selectivity of the homogeneous catalyst remains. However, the accessibility of the catalyst to the reactants becomes an issue. Since the catalysts and reactants are still in the same phase, it is not so much diffusion through a phase boundary but diffusion through the support medium that becomes the limiting factor. One example of this is the case where a porous nanoparticle is used as a support. If the catalyst sites were attached in the pores rather than on the surface, the reactants would actually have to diffuse into the pore in order to react, which means that the closer the catalyst is to the pore entrance, the more accessible that catalyst site is. The optimal geometry, in this case, is the one where the catalyst site is always attached to the surface of the nanoparticle, and the system becomes less optimal the farther to the interior of a pore that the catalyst is attached. In the case of a polymer, a large random coil polymer could block catalyst sites due to the high flexibility of the backbone. The more flexible the backbone, the increased likelihood of the backbone cutting off a direct route to a catalyst site. This is not to say that the catalyst site is absolutely inaccessible, but now the situation occurs where the reactant must take a specific route to get to the catalyst, which is the least optimal situation. The

optimal polymer support would be a rigid rod polymer because all of the side groups would be accessible. The next best would be a polymer with a helical conformation if the backbone bonds are flexible, as this adds order to the backbone. Whether a polymer has a rigid rod, helical, or some other conformation, the most important aspect of the molecule as a catalyst support is whether or not the catalyst group is accessible to the reactant molecules for the desired reaction. Because of this, determining whether a catalyst group is on the external "surface" of the polymer support through molecular simulations would be helpful in choosing the most optimal polymer support.

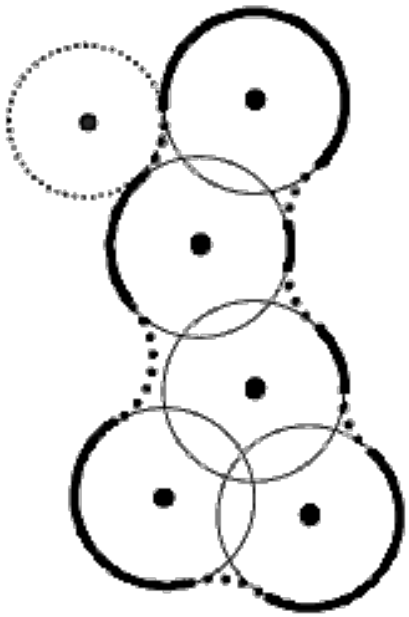
4.1.4.1 Accessibility

Accessibility has played a large role in biological systems. In the pharmaceutical industry, a major aspect of drug effectiveness is the accessibility of the drug to the active site of interest. An active site that is largely inaccessible to an effective drug is a less desirable target than an active site that is considered accessible. This is a somewhat different situation from immobilized homogeneous catalysis. In the case of drug design, the drug (catalyst) needs to be either targeted to a specific active site (reagent), or it must be designed such that it can access the active site. For immobilized homogeneous catalysis, the catalyst must be accessible to the reagent. In other words, the catalyst is now the molecule which must be accessible. Even though it is the catalyst which must be accessible, the tools which have been developed for biological systems can be applied to the catalysis case.

In 1971, Lee and Richards proposed the idea of and a method for determining the "solvent accessible surface" of a protein. This involved using a "probe sphere" that was "rolled" around the external molecules of a molecule. The rolling probe model starts with the van der Waals surface (the surface created by expanding the atom sizes to their van der Waals radii, shown in Figure 19(a)) and uses the van der Waals surface of the probe to create a map of the solvent accessible surface by manually rolling the probe around the van der Waals surface (Figure 19(b)). In this case, the probe radius is included in the solvent



(a) (b)



(c)

Figure 19: Solvent accessible surfaces[75]. (a) van der Waals surface, (b) Lee and Richards rolling probe surface, and (c) Connolly surface.

accessible surface. Given the right probe dimensions, this can generate a surface which accurately includes the solvent surface[76]. Connolly was the first to create a convenient algorithm to automate this process. The main difference between the "Connolly surface" and the rolling probe surface is that the solvent area is excluded from the Connolly surface, but overlapping regions of the van der Waals surface are smoothed out (Figure 19(c))[77, 78].

Edelsbrunner applied a different methodology to define the same surface for proteins, called alpha shapes[79]. Alpha shapes have already been used to measure the molecular surface and voids in proteins[80]. By applying this method to catalyst accessibility, accessible catalysts which are attached by a side group can be identified. An accessible catalyst would be one that is on the molecular alpha shape where the weight of the alpha shape includes the van der Waals radius of the potential reactant. The geometric equivalent would be a rolling probe surface (Figure 19(b)).

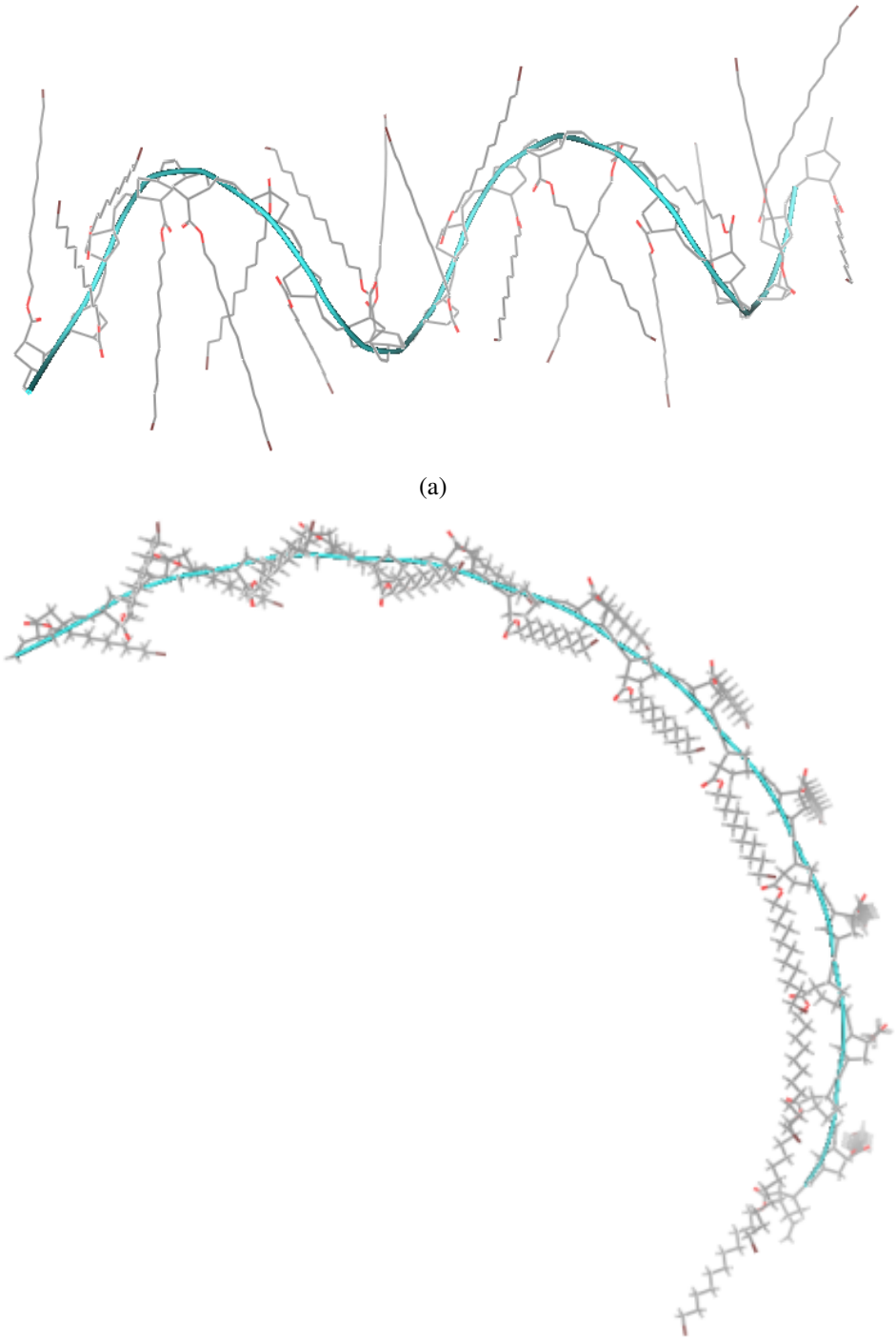
Rather than the previously used geometric algorithms, the alpha shapes method is an analytic method, i.e., it is an abstract mathematical method that can construct the solvent accessible surface without actually rolling a solvent around the molecule. The most general explanation of alpha shapes is that they are weighted Delaunay simplices. Delaunay tessellation normally involves constructing geometric shapes that resemble tiles with no overlap or empty space between the tiles. These tiles, called simplices, are simply triangles in two dimensions and tetrahedra in three dimensions. For a three-dimensional molecule, there would be four vertices forming the simplex, where each vertex was positioned at the center of an atom. The main contribution by alpha shapes is the addition of "weights" to these simplices[79]. If, for example, an atomic simplex were weighted by its van der Waals radius, the alpha shape would be a simplex that covered not just the atom centers, but also the van der Waals radius. Normal Delaunay tessellation can be used to generate the convex hull, where the convex hull is formed by the external atom centers. The convex hull can be described as the shape that is formed that contains the entire set of points with

no concave regions. More simply, it is very similar to the Connolly surface as it is shown in Figure 19(c). A true Connolly surface can include voids in side of the actual surface, whereas a convex hull would only include atoms which are in contact with the external system. The convex hull formed by alpha shapes on a molecule would include the atom centers of the convex hull formed by Delaunay tessellation, as well as their van der Waals radii. The molecular alpha shape would then be defined as the weighted convex hull, where the surface of the alpha shape may include more points than the convex hull, so long as those points are accessible to the external surface by a given distance with that particular weight[79].

4.2 Simulation Methodology

4.2.1 Starting Conformations

The simulations were created for each of the eight possible configurations of ROMP PNB that have side groups (Figure 11). These configurations are the same as those used for the TSMC simulations, except the case where there is no side group is ignored. Additionally, for each of the configurations, there were two starting conformations, which are shown in Figure 20. The first is a helix, and the second is an elongated straight chain. The curvature of the straight chain is due to the fact that the bonding across the cyclopentane ring is not completely linear. The purpose of having two starting conformations is so that there is increased sampled conformation space. Work performed by Lu and Kofke showed that to improve accuracy, a low entropy conformation has to be available to equilibrate into a high entropy conformation. This is because a high entropy conformation may not equilibrate to a low entropy conformation[81]. In this case, the helical conformation is the low entropy conformation and the random coil is the high entropy conformation. It is very unlikely for the random coil conformation to form a helical conformation in a reasonable amount of equilibration time, particularly if there is a large amount of steric crowding in the transitional conformations. A helical starting conformation can easily form a random coil, so



(a)

(b)

Figure 20: Representative backbone starting conformations of all of the ROMP PNB configurations. (a) Helical conformation (1) and (b) elongated conformation (2).

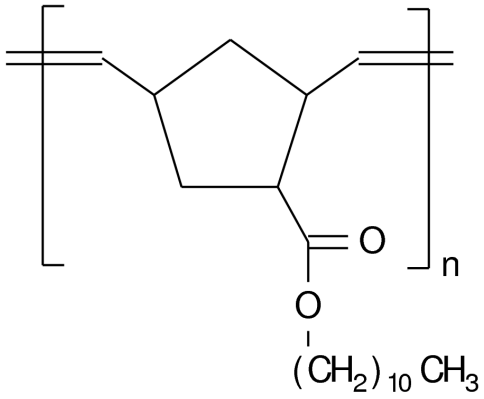


Figure 21: ROMP PNB monomer will L11 side chain. The number of repeat units (n) were 20, 50, and 100.

long as there is a pathway that does not involve unfavorable steric hindrance. A helix that becomes a random coil over the course of a simulation indicates that the conformation is definitely a random coil. However, because a random coil will not adopt a helical conformation under a reasonable simulation time, and it is possible for the helical conformation to remain a helix if it is in a local minimum, both structures need to be used. The likely conformation is determined by looking at the both the potential energy and the structure. The structure that has the lowest potential energy and is a realistic structure is the one that is most likely to be the global minimum conformation. However, in this case it is assumed that the entropy part of the free energy is negligible so that only the internal energy is used to determine the global minimum.. An elongated chain is used because the only easily available final conformation is that of a random coil. As with the isomeric designations, the starting conformation will be designated by the numbers *1* and *2*, where 1 will designate a helical starting conformation, and 2 will designated an elongated starting conformation.

In addition to the starting conformations, some other cases were tested. All polymers were simulated with three different polymer lengths. Specifically, polymer lengths of 20, 50, and 100 monomer units were simulated for all configurations and both starting conformations. This was done to determine the effect of polymer length on accessibility. All of these simulations use linkers that are an undecane chain attached to the cyclopentane ring

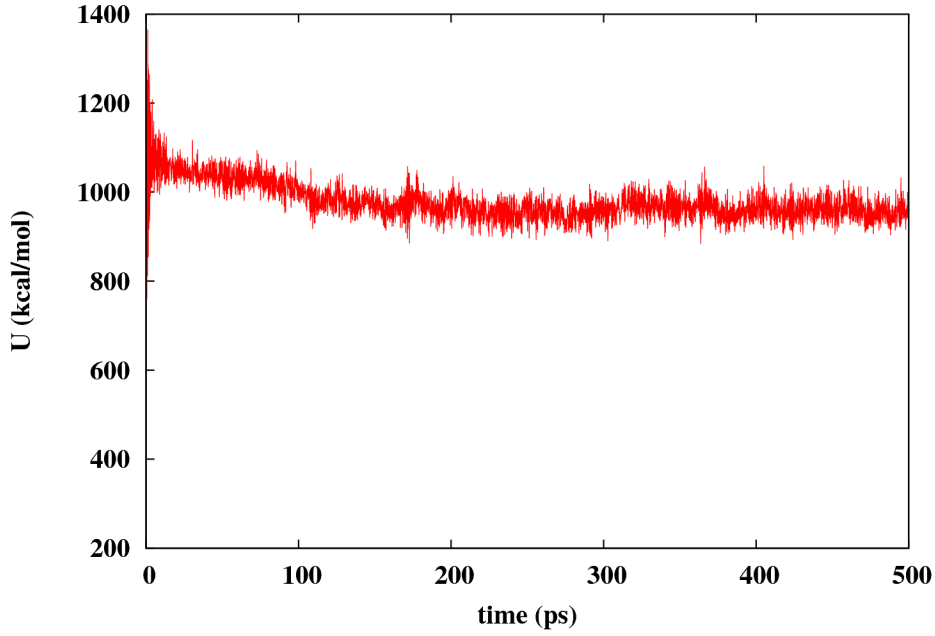
by an ester group (Figure 21). In order to determine the effect of linker length, the carbon chain was varied for the CR1 and CR2 polymers of 100 monomer units. Specifically, the side chain lengths used were 3, 5, 7, and 9 carbons long, designated as L3, L5, L7, and L9, respectively. The 11 carbon chain will be referred to as L11, which will be implicit unless specifically examining linker length.

Prior to running molecular dynamics, each starting conformation was minimized using gradient cutoffs of 10, 0.1, and 0.01 kcal mol^{-1} $Å^{-1}$ for the steepest descent, conjugate gradient, and truncated Newton minimization steps, respectively.

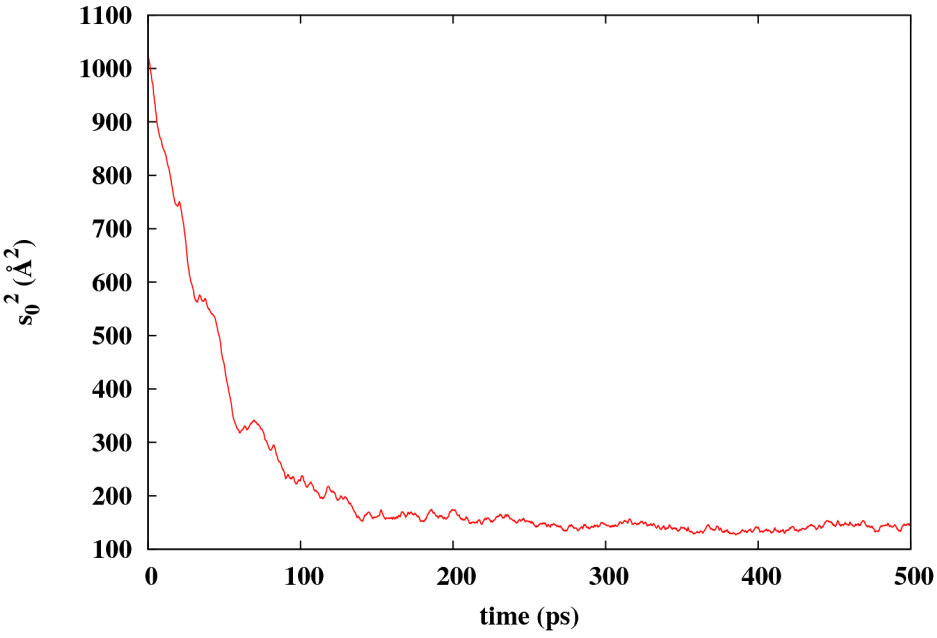
4.2.2 Molecular Dynamics

NVT MD was used to sample the conformation space at 300 K. The NHSV algorithm was the dynamics algorithm used[40], and the thermal coupling parameter was set to 0.5 kcal mol^{-1} ps $Å^{-1}$. The time step was 0.001 ps, and the hydrogens were constrained using the LINCS constraint algorithm[82]. Appendix C contains the custom SVL MD code which was used for these simulations, as well as a description of the differences between this code and the default MOE MD code. The simulation was sampled every 0.1 ps. The simulations were run for 500 ps, and the final 100 ps were used for the data analysis.

The simulations were repeated ten times for each starting conformation that was used. This was done for two reasons. The first is to have simulations which have different starting velocity distributions, and the second was to establish a confidence interval for that particular conformation. There were some cases in this work where there were two different final conformations generated for a given starting conformation. For these particular conformations, there were five simulations which generated one given conformation and five that generated the other. The one with the lowest potential energy is the one that is used, but this particular conformation will have a smaller sample size, which will be apparent by an enlarged confidence interval.

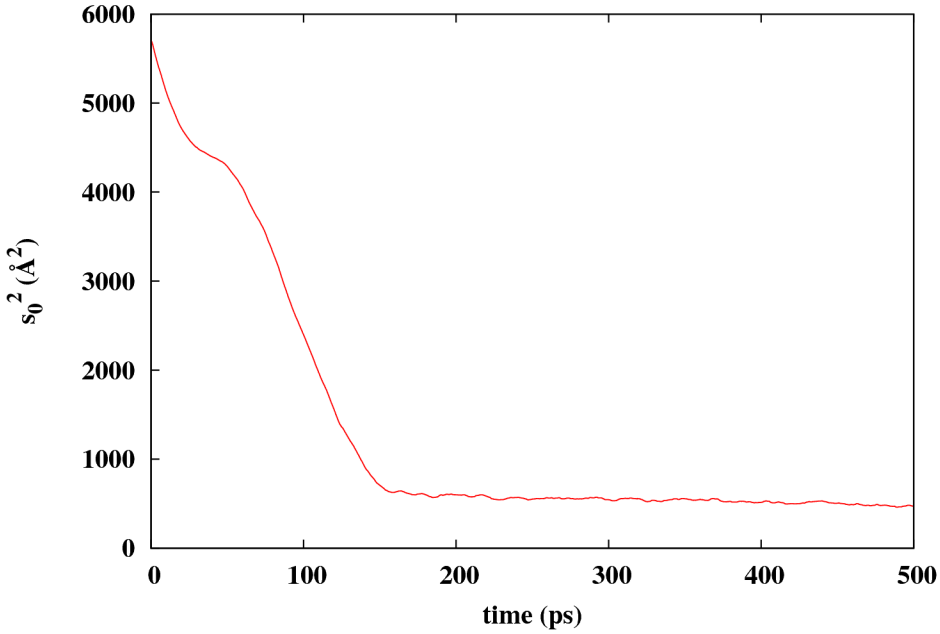


(a)

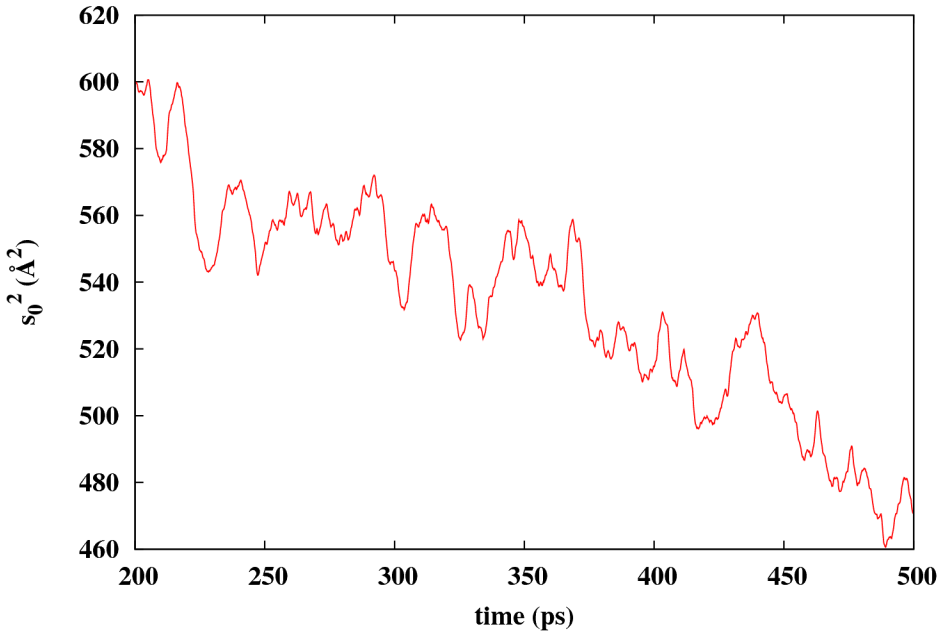


(b)

Figure 22: An equilibrated 500 ps simulation of ROMP PNB. The (a) potential energy and (b) unperturbed mean squared radius of gyration for the same 500 ps simulation are shown.



(a)



(b)

Figure 23: The unperturbed mean squared radius of gyration over (a) the entire 500 ps simulation time and (b) the last 300 ps of that 500 ps for an non-equilibrated simulation.

4.2.3 Equilibration

Figure 22(a) shows the potential energy for an equilibrated 500 ps simulation. While this plot does show equilibration, the variation in potential energy makes it difficult to accurately describe a simulation as equilibrated. This is particularly true for simulations that equilibrate slowly, and the potential energy decreases with an almost imperceptible slope. Therefore, equilibration was determined by the radius of gyration. Figure 22(b) shows that the radius of gyration has much less variation, and it is easy to see a plateau of the radius of gyration. However, care must be taken to determine equilibration. Figure 23(a) shows the squared radius of gyration over the length of the simulation. Because of the limits of the plot, it seems that the polymer has equilibrated, while Figure 23(b) is plotted on a finer scale and shows that the radius of gyration is constantly decreasing over the length of the simulation.

Each of the simulations was individually examined to determine equilibration from the radius of gyration. If the radius of gyration was fluctuating about a fixed average for the final 100 ps of the simulation, then the results from the final 100 ps of that simulation was used for analysis. Otherwise, the simulation was run for an additional 300 ps and the final 100 ps of that additional simulation time was used for analysis, so long as the radius of gyration showed a plateau.

4.2.4 Accessibility Determination

Alpha shapes were used to determine whether or not a catalyst site was considered accessible. The only drawback to using alpha shapes to determine accessibility is that the computation is expensive, particularly for the 100 repeat unit polymers. In order to decrease computational time, the hydrogens were removed from the calculation. For each monomer unit, there are 31 hydrogens, which corresponds to 3100 atoms removed from the calculation for the 100 repeat unit case, which substantially increased the calculation speed. Each sampled conformation was analyzed and given a “score” that had a range

between zero and one and was determined by

$$\text{Accessibility} = \frac{\text{Number of Accessible Catalyst Sites}}{\text{Number of Total Catalyst Sites}}. \tag{20}$$

The total score for a simulation was an average over each sample from the final 100 ps of the simulation. The only noticeable effect of not including the hydrogens in the determination of the alpha shapes was that there was a roughly 0.02 increase in the score for all cases. Because only relative changes are important for this method of tracking accessibility, this change in score from hydrogen removal is inconsequential.

The catalyst sites were modeled as LJ spheres that took the place of a hydrogen atom on each side group's terminating carbon. The default van der Waals radius used for the catalyst site was that of hydrogen, or 1.485 Å. However, to determine the effect of catalyst size on the accessibility, the CR1 and CR2 polymers were run for all of the polymer lengths with van der Waals radii of 2.30 Å, 3.13 Å, and 5.13 Å. These are designated as CS2, CS3, and CS5, respectively, and the default case is referred to as CSH. The simulation times had to be increased for CS3 and CS5, where the total simulation times were 1 ns and 1.5 ns, respectively. These van der Waals spheres needed additional parameterization to account for the fact that these radii are unrealistic for a single atom. In addition to the van der Waals radii, the atomic mass and well depth, ϵ, were chosen to reflect these larger van der Waals spheres. The well depth was 0.2, 0.4, and 0.5 kcal mol^{-1} for CS2, CS3, and CS5, respectively. These were increased from that of hydrogen to account for the lack of the additional van der Waals interactions that are no longer present by reducing a multi-atom catalyst into a single spherical approximation. The atomic masses of the spheres were also increased to 78, 369, and 606 g mol^{-1} for CS2, CS3, and CS5, respectively. By increasing all of the parameters, these catalyst sites are more realistic approximations of real catalyst systems without actually modeling all of the involved atoms separately, which greatly decreases the computational power required for each iteration of MD.

In addition to catalyst site size, the reactant size was also varied. The reactant sizes

used were 1.4 (water), 2, 3, 4, 5, 7, 10, 15, and 20 Å (designated RS1.4, RS2, etc.). This was done to see the effect of reactant size on accessibility. The accessibility scores were calculated using MOE, and the code is included in Appendix D.

4.2.5 Assumptions and Limitations

There are several major assumptions made for these simulations that must be noted to understand both the purpose of the setup of the simulations as well as the limitations of this work. These assumptions are very general to this method, and may not be applicable to specific applications of ROMP PNB-supported catalyst systems. The first major assumption is that the polymer that is being characterized is solvated using its θ-solvent. This allows the simulation of the polymer without using a solvent, as the θ-solvent, by definition, is the one in which the polymer chain is unperturbed by the solvent. This allows for either faster simulations or simulations of a longer chain length, compared to a simulation with an explicit solvent. This is the assumption used in this work, but it should be noted that this was done as a computational expedient. A simulation could be performed if a specific solvent that is known not to be a θ-solvent is used. The drawback, of course, is that a simulation with an explicit solvent takes substantially more computational time.

The second major assumption is that the polymer is being used as an immobilized homogeneous catalyst in dilute amounts, i.e., the polymer is not in a bulk solution where the bulk is mostly polymer chains. This allows for single polymer simulations, which, again, allows the use of larger polymer sizes while keeping the simulation time reasonable. If there were a significant amount of polymer added to the solution (i.e., non-dilute), the simulation would have to include additional polymers to account for polymer-polymer interactions. In general, immobilized homogeneous catalysis is performed with a dilute amount of catalyst, but it is important to note this in cases where this accessibility analysis might be used for other applications.

Another major assumption was that the diffusion of a reactant to a "blocked" catalyst

site through the polymer matrix will not affect the accessibility. Since the diffusion of a reactant will be slower than the diffusion of a reactant from the bulk to a catalyst site on the exterior of the polymer matrix, only these sites on the molecular alpha shape need to be considered as accessible. For example, in the comparison of two polymers where the first polymer had all of the catalyst sites on the exterior surface and the second had all of its sites only accessible through a winding channel through the polymer structure, the first polymer would be considered more accessible by this method. That is not to say that the catalyst sites on the second polymer are completely inaccessible, but this method's main purpose is to look at accessibility as a bulk property that can be used to compare multiple types of polymers. This method uses the idea of accessibility as a rate limiting step determinant. If the catalyst sites are the same between two polymer chains, then the accessibility measurement will determine which polymer will make the optimal support.

It is also assumed that the reactant is perfectly spherical. This is a valid approximation because, if done correctly, it can be used to model a reactant which has multiple orientations. By choosing a van der Waals radius of the reactant that starts at the center of mass and extends to the average van der Waals radius of the reactant compound at its surface, the reactant can then exist in the simulation as a sphere that takes into account the average orientation. This approximation is valid for reactants which are either completely symmetric or mildly asymmetric in three dimensions.

The last major assumption that was employed is that the catalyst site itself is spherical and equally reactive on all points on the surface. This is similar to the assumption of a spherical reactant in that highly asymmetric catalyst sites would not be scored accurately. Additionally, catalysts whose reaction mechanisms involve specific regions of the catalyst surface cannot be accurately characterized as it is done in this work.

Some limitations to the use of alpha shapes for accessibility characterization must also be described. The first such limitation is in the simplicity of the system. This method can only be used for single reactant and single catalyst reactions. For a multiple reactant

reaction, this method would probably still work as long as the van der Waals radius for the largest reactant is used. However, it would be difficult to just use alpha shapes to determine the accessibility for multiple catalyst reactions. A possible solution would be to cross-reference all accessible catalysts with other catalysts that lie within a certain separation distance, but if the catalyst site is specific in orientation or what part of the site is active, this method would be difficult to apply in the manner that it is applied for this work.

By far, the largest limitation of this method is that there is no concrete meaning of accessibility with regard to a physical experiment. More specifically, the number that is used to score accessibility only has significance when compared to another system that has had the identical method for characterizing accessibility performed. The results of this method can only indicate that a particular support is more accessible than another, so the more accessible one would be a better candidate for a catalyst support. However, the accessibility value of a single support does not indicate that it is absolutely accessible or inaccessible. Additionally, this is a very coarse-grain method. Accessibility values would need to be substantially different to be a concrete indicator of support performance. This indicates that this method should be a starting point when picking a catalyst support, and that its primary function is to narrow down a list of possible support candidates, even though it may not find an absolute optimum.

4.2.6 Simulation Averages

For each individual simulation, there are 1000 data points for the 100 ps of sampled data that is further analyzed. These data points are averaged for all of the properties, e.g., energy, radius of gyration, and accessibility. Since each simulation was repeated a number of times, each of these values is averaged over the repeated simulations, and the real average is determined through the use of the 90% confidence interval.

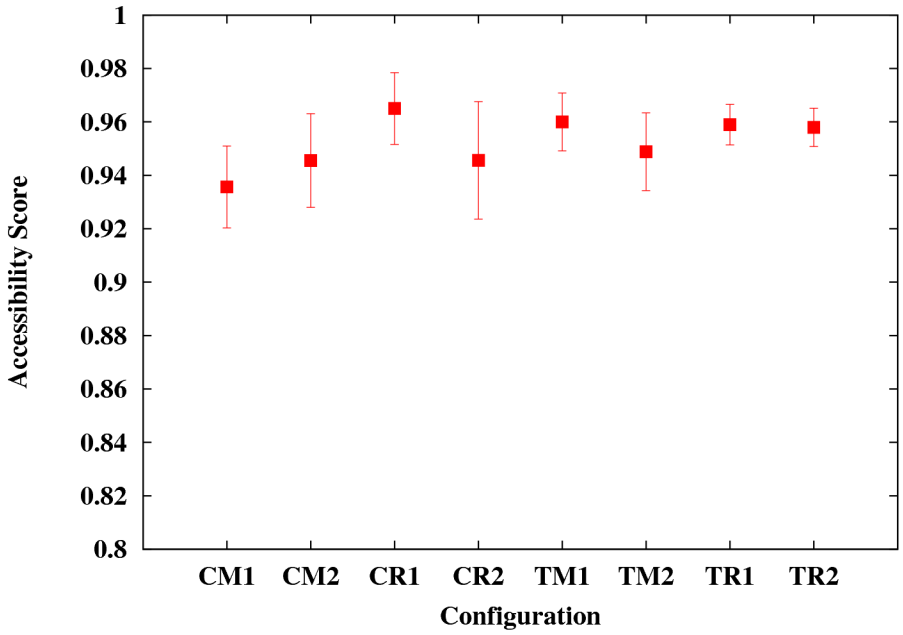


Figure 24: Accessibility scores for all ROMP PNB configurations with 20 repeat units with RS1.4 and CSH.

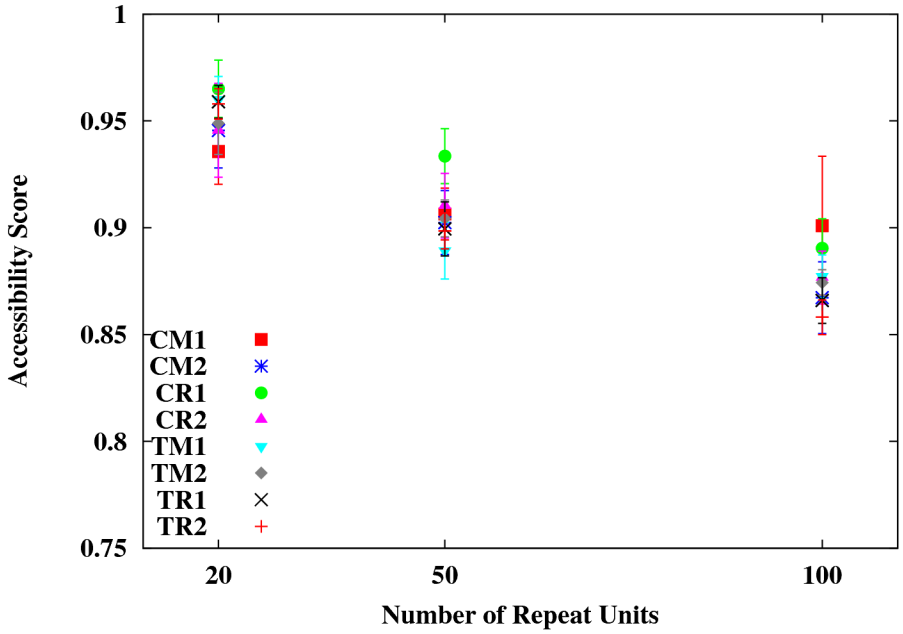


Figure 25: Accessibility scores for all ROMP PNB configurations with 20, 50, and 100 repeat units with RS1.4 and CSH.

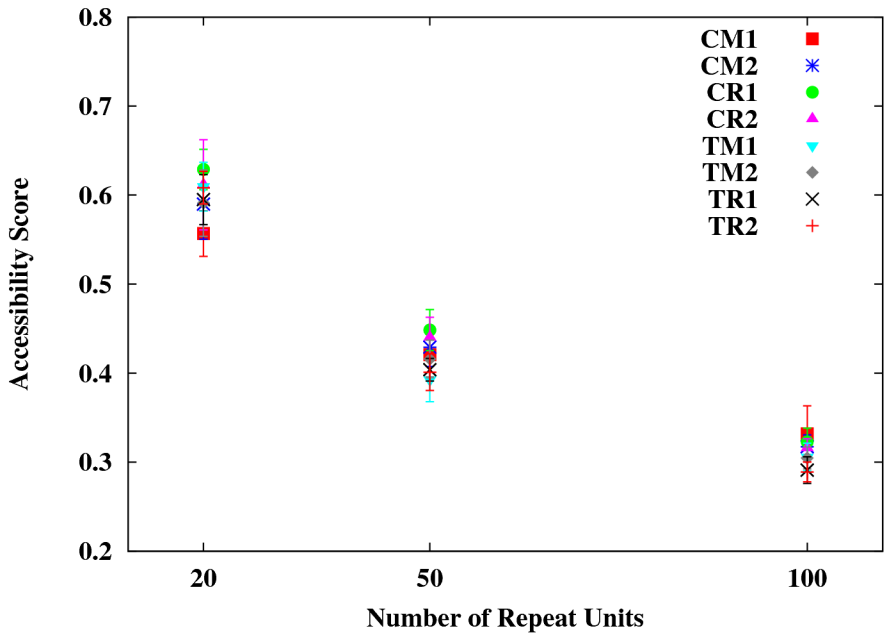


Figure 26: Accessibility scores for all ROMP PNB configurations with 20, 50, and 100 repeat units with RS20 and CSH.

4.3 Results and Discussion

4.3.1 Accessibility Scores

Figure 24 shows the accessibility for each configuration for the case of 20 monomer units with RS1.4 and CSH. It is obvious from this graph that the accessibility score for each configuration is statistically the same. Figure 25 shows the accessibility for each configuration as a function of polymer size. The expected result is that the accessibility will substantially decrease as the polymer size increases for random coils. For rigid rod structures, it is expected that the accessibility may decrease, but not by a large amount. Figure 25 clearly shows two important details about the accessibility score. First, the accessibility does decrease as the polymer size increases, but the decrease is not as much between 50 and 100 repeat units as it is between 20 and 50 repeat units. The second is that the accessibility score is statistically the same for each configuration. While the range seems large enough to differentiate configurations, each point is statistically the same as its nearest neighbor(s). This would suggest that all of the configurations develop the same final conformation, that

of a random coil. This would seem to confirm the result from the TSMC simulations. Additionally, these results are independent of reactant size. All of the configurations had statistically similar accessibility scores when comparing each configuration at a given reactant size. However, the major change is that the average difference in accessibility score between chain sizes increases as the reactant size increases, which is shown in Figure 26.

It is important to note here that the CM1-100 configuration produced two significantly different final conformations. This is the only configuration and size that yielded different final conformations. However, only the conformation with the lowest energy is included in Figure 25. The average energy of the conformation which was used was 4720 kcal mol^{-1}, while the other conformation had an average energy of 5710 kcal mol^{-1}. Since these are identical polymer simulations, a potential energy difference of 90 kcal mol^{-1} is indicative of two distinct conformations. The difference in $\langle s^2 \rangle_0$ is 40 $Å^2$, which also indicates differing conformations. Because of these differences, only the set of conformations with the lowest average energy were used in the configuration average. This highlights the need to run multiple simulations when using MD to sample conformation space. In this case, randomly generated initial velocities can yield different conformations.

Figure 27 shows the relationship of reactant size to the accessibility score. This is definitely the expected result for reactant size. As the reactants get larger, the more important it is that the catalyst site is part of the most exterior van der Waals surface. An extreme example would be the difference between a water molecule and a buckminsterfullerene (C_{60}). The water molecule can access a catalyst site that, while on the external surface of the molecular alpha shape, may be sandwiched between two more external atoms. This site could be considered accessible to water, but something the size of a buckminsterfullerene would not be able to access the catalyst site. This is seen exactly in Figure 27, and while this is only based on values for one configuration, this same relationship is seen for all of the configuration. This relationship decays exponentially, and it levels off at a certain accessibility score. This would be expected because as the reactant size approaches infinity,

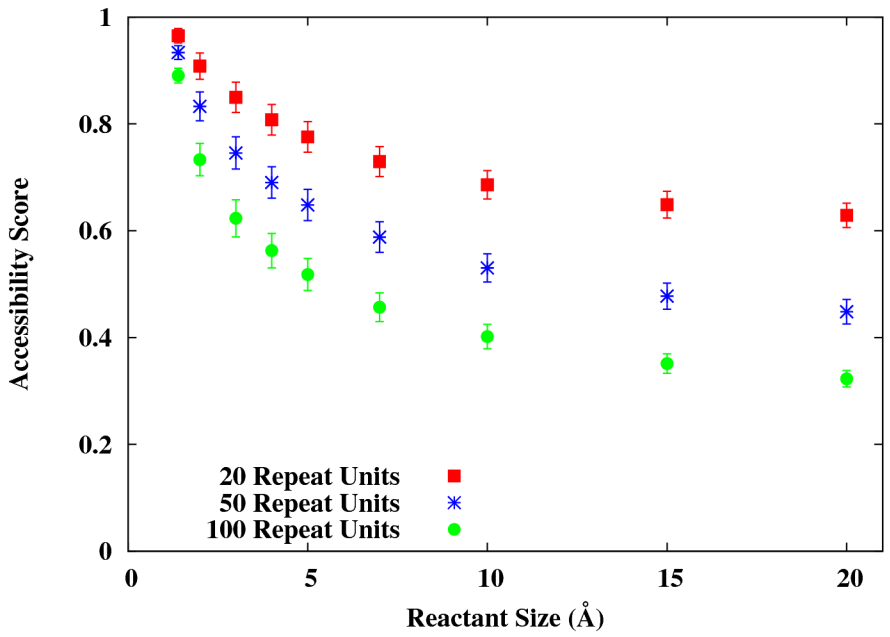


Figure 27: Accessibility scores for the CR1 configuration with 20, 50, and 100 repeat units for all reactant sizes.

only the catalyst sites on the absolute polymer surface would be accessible. As the reactant size increases, these sites become isolated in the accessibility score. This figure also shows the impact that polymer size has on the relationship between accessibility and reactant size. As the polymer size becomes larger, the accessibility decreases to a greater extent for larger reactant sizes. This supports the hypothesis that random coil polymers would block catalyst sites with the polymer backbone. A helical polymer should have a more linear relationship with a negligible downward slope.

The expected behavior for the effect of linker length is that the accessibility should increase as the linker length increases. This is due to the increased linker length allowing the catalyst site to be farther from the polymer backbone. As the catalyst site moves farther away from the backbone, it has more of a potential to be on the external surface of the polymer and in the bulk solution. This should hold true regardless of the conformation of the backbone, however, the effect should be smaller for a helix-kink conformation because the catalyst sites should already be more accessible. This relationship between linker length

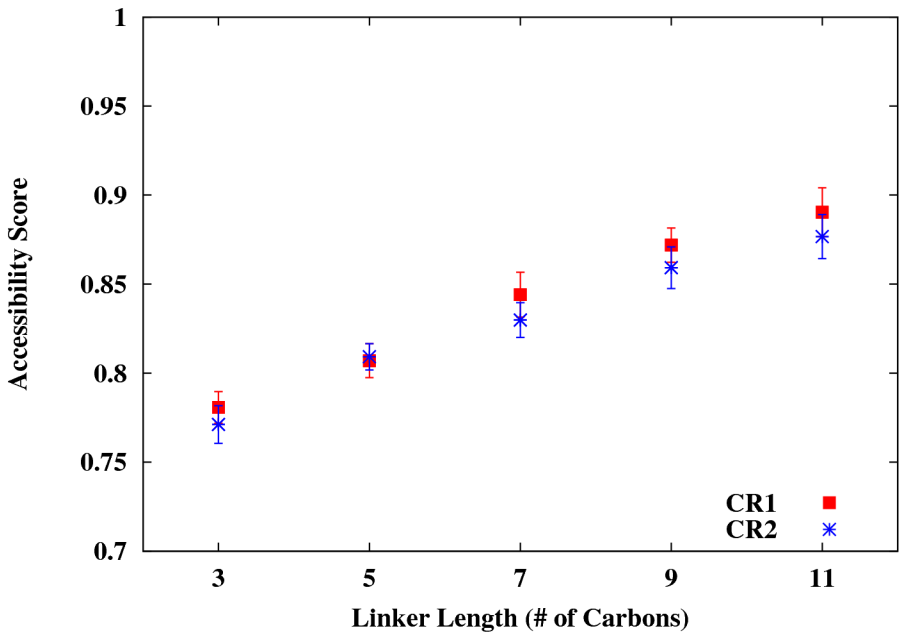


Figure 28: Accessibility scores for the CR1 and CR2 configurations with 100 repeat units for all linker lengths.

and accessibility should hold true up to the point at which the linker length gets to such a length that it could be considered a random coil in itself. There would be two ways to alleviate this. The first is to use a linker that is not flexible, and the second would be to keep the linker length a reasonable length. The expected linker length behavior can be seen in Figure 28. For both CR conformations, the accessibility score increases from L3 through L11.

The effect of the catalyst itself is somewhat harder to understand, but it has an expected behavior as well. As the catalyst site gets larger, van der Waals forces will keep it more towards the external surface of the molecule. This is done by two different variables - the polymer backbone and the linker. For flexible linkers, the van der Waals forces will keep a larger distance between the backbone and the catalyst site because of the larger van der Waals radius of the catalyst site. A linker that is rigid will show little difference in this regard. However, a larger catalyst site on a rigid linker attached to a random coil polymer will still keep the backbone at a greater distance from the catalyst site for the same reason.

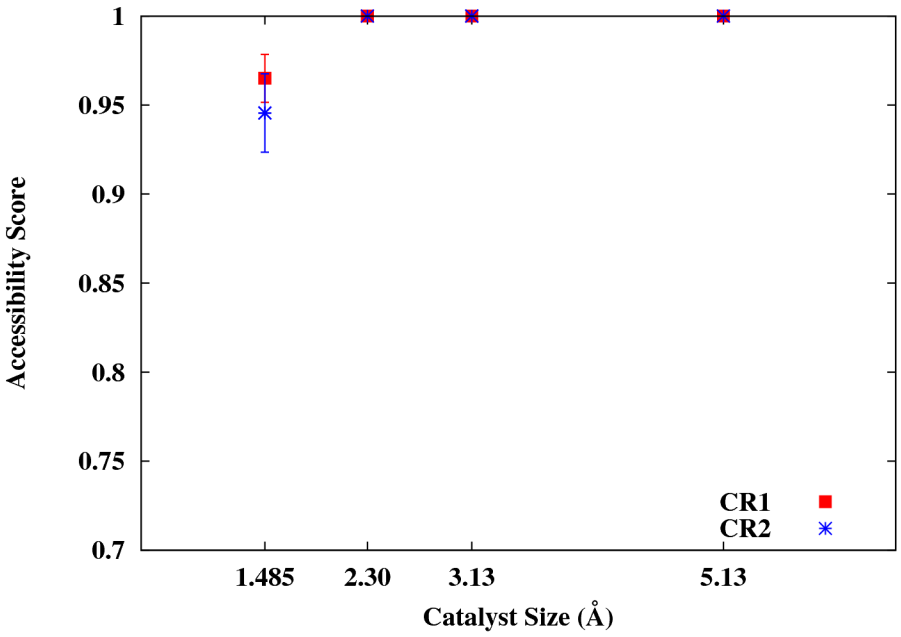


Figure 29: Accessibility scores for the CR1 and CR2 configurations with 20 repeat units for all catalyst sizes and RS1.4.

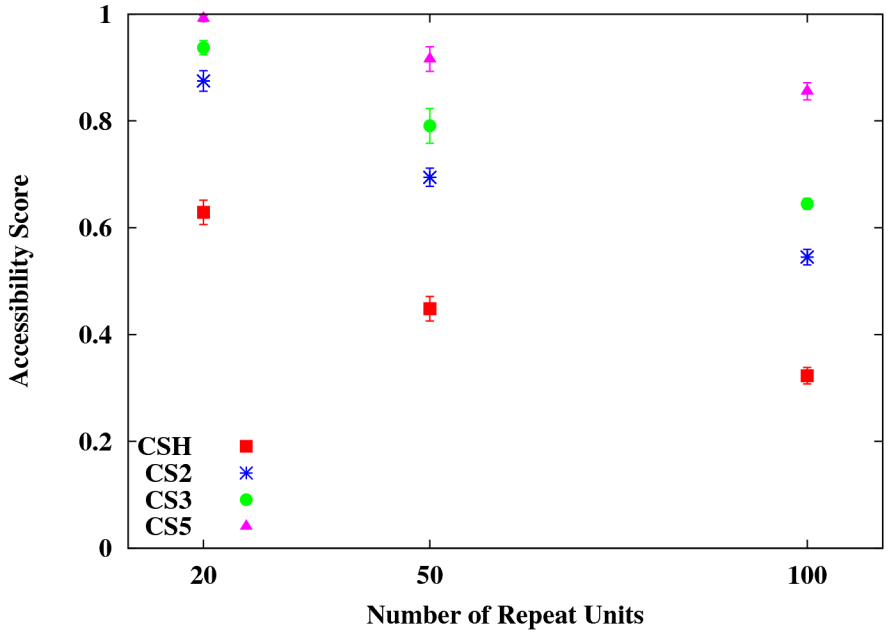


Figure 30: Accessibility scores for the CR1 configuration with 20, 50, and 100 repeat units for all catalyst sizes and RS20.

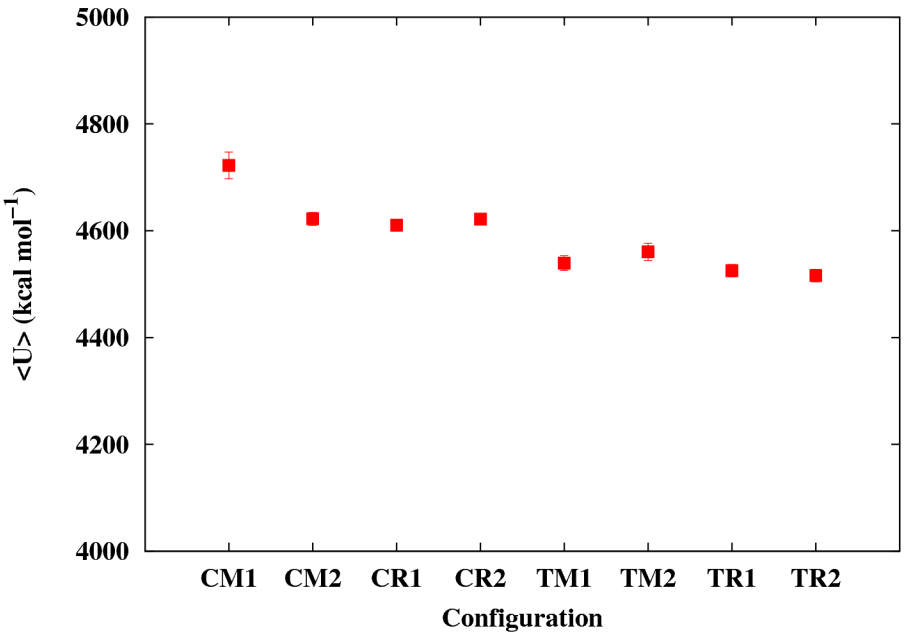


Figure 31: Average potential energy for all ROMP PNB configurations with 100 repeat units and CSH.

For the CR1 and CR2 polymers, the expected result would be for the CSH catalyst size to have a lower accessibility compared to the other catalyst site sizes. Additionally, the accessibility should increase for the increasing catalyst size. This is exactly the obtained result, shown in Figure 29. However, it would seem that all catalyst sites become accessible for CS2 and greater. This trend is actually the same for all chain lengths simulated for the smaller reactant sizes. Figure 30 increases the resolution of the effect of the catalyst site size by showing the results for RS20. It shows the expected trend of the accessibility increasing as catalyst size increases, and this effect is coupled with the change in accessibility due to chain length.

4.3.2 Energies and Conformations

One way to determine which conformation is the most energetically desirable is to look at the average potential energy for each conformation. Plots of the energy for the 100 repeat unit case and for all polymer sizes can be found in Figure 31 and Figure 32, respectively.

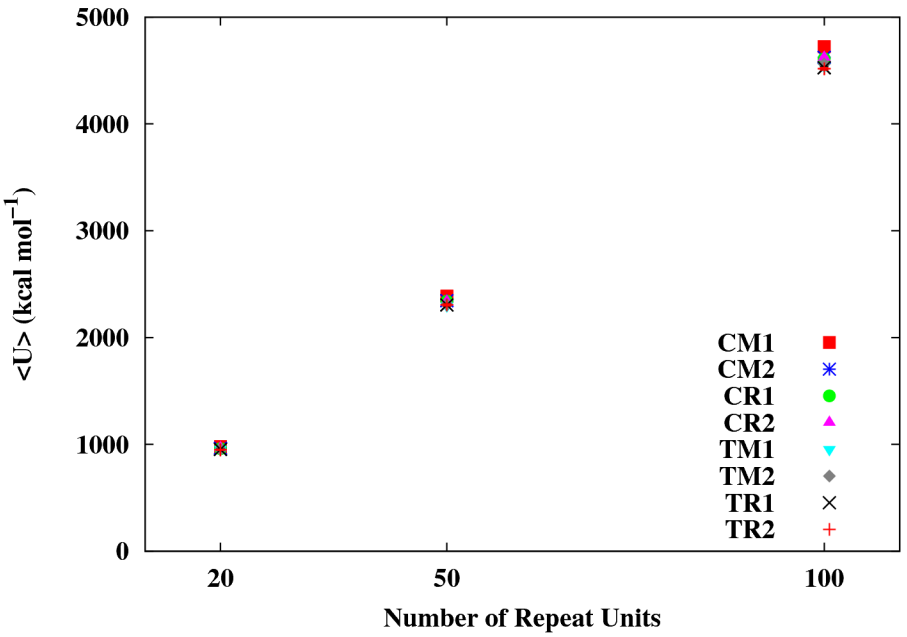


Figure 32: Average potential energy for all ROMP PNB configurations with 20, 50, and 100 repeat units and CSH.

The most important message from these plots is that there is no difference in energy between any of the configurations. It may seem at first glance that there is some spread in the 100 repeat unit case, however, the difference between the highest energy value and the lowest energy value is less than 5% of the total energy, which is an almost negligible energy difference. Energetically, it would seem that the trans configurations are more preferential, but only by a small amount.

The easiest way to determine whether or not a conformation itself is similar using a polymer property is by looking at the mean square radius of gyration. For rigid rod polymers, the scaling of the radius of gyration with polymer size will be dissimilar to that of a random coil polymer. This is useful to determine whether or not conformation really has an effect on the accessibility. If two polymers have identical accessibility scores and different radii of gyration, then the effect of the conformation is negligible on accessibility. Figure 33 shows $\langle s^2 \rangle_0$ for all of the configurations for polymers of 100 repeat units. Figure 34 shows $\langle s^2 \rangle_0$ for all configurations for all of the polymer sizes. It is obvious from these plots

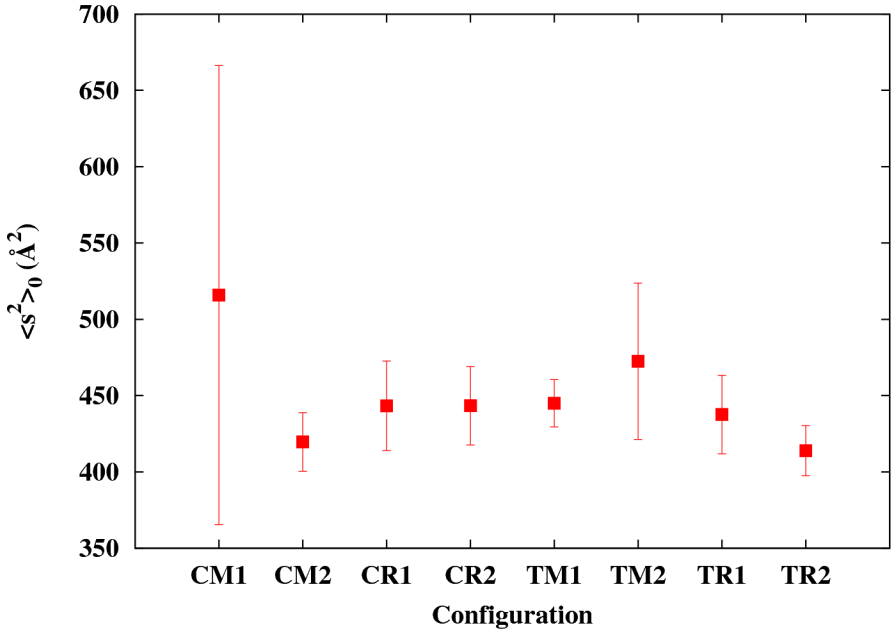


Figure 33: Average $\langle s^2 \rangle_0$ for all ROMP PNB configurations with 100 repeat units and CSH.

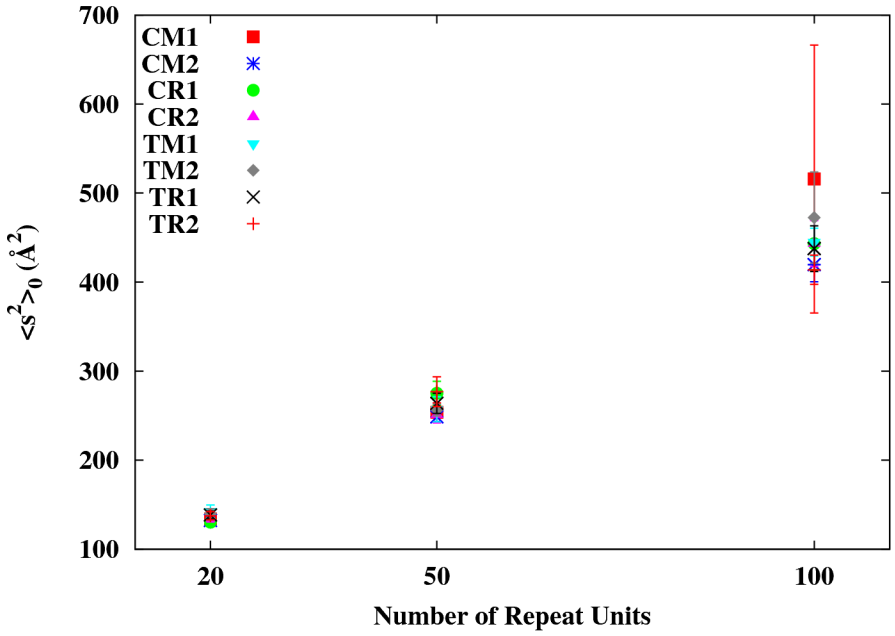
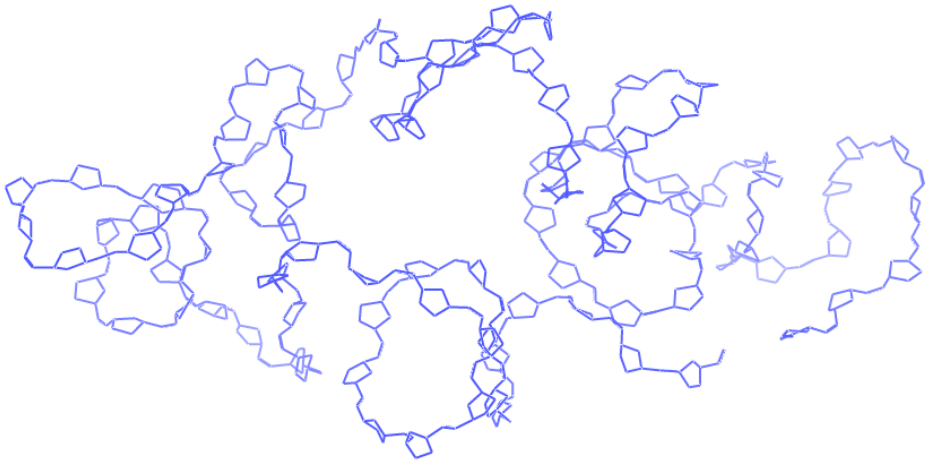


Figure 34: Average $\langle s^2 \rangle_0$ for all ROMP PNB configurations with 20, 50, and 100 repeat units and CSH.

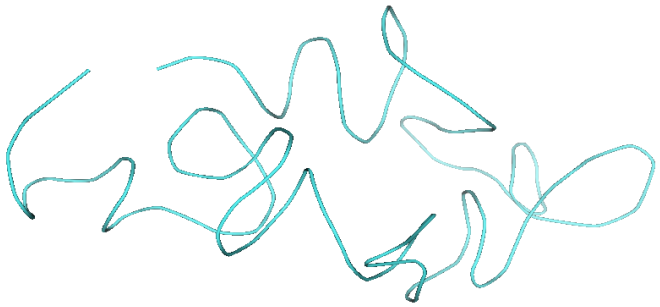
that the mean square radius of gyration is statistically identical across all configurations. While this does not show a relationship between the accessibility and the conformation, it does show that all of the configurations have similar final conformations.

The only configuration that might seem to have a higher $\langle s^2 \rangle_0$ is CM1-100. This particular configuration also has the largest confidence interval because it had two separate final conformations, and only one was used for the analysis. However, the size of the confidence interval causes that particular configuration to be statistically similar to the other configurations. The fact that it is a different conformation means that it is worth investigating. Figure 35 shows the final conformation of CM1-100. Comparing that to Figure 36 (CM2-100) shows that while the backbone of CM1-100 seems more expanded, there is nothing that would suggest increased order in CM1-100. Additionally, the energy of the CM1-100 configuration is 100 kcal mol^{-1} higher than that of CM2-100, which means that the most likely source of the difference between CM1-100 and all of the other configurations is that there is an effect from the starting conformation of this particular configuration.

Figure 37 shows the effect of linker size on the average potential energy and $\langle s^2 \rangle_0$. As the linker length increases, there is also an increase in potential energy. This is expected because the main force field terms that are added to the system are always positive for additional bonded atoms (see Appendix A for equations). What is very interesting is that the linearity of the figures shows that essentially there is the same amount of energy per atom. This means that there is energetically no difference between the polymer backbones for each linker length. The increased $\langle s^2 \rangle_0$ is most likely due to the longer linker lengths expanding the polymer by not allowing the backbone to fold over itself like it would if there were no side groups on the backbone. However, even though $\langle s^2 \rangle_0$ increases, the chain with the largest linker length (previously discussed) does not become more ordered, the expansion of the polymer conformation is purely steric. It is also interesting to note that both of these quantities increase linearly. This can be useful in predicting the potential energy or $\langle s^2 \rangle_0$ for flexible linkers up to the point where the linker size has a similar order

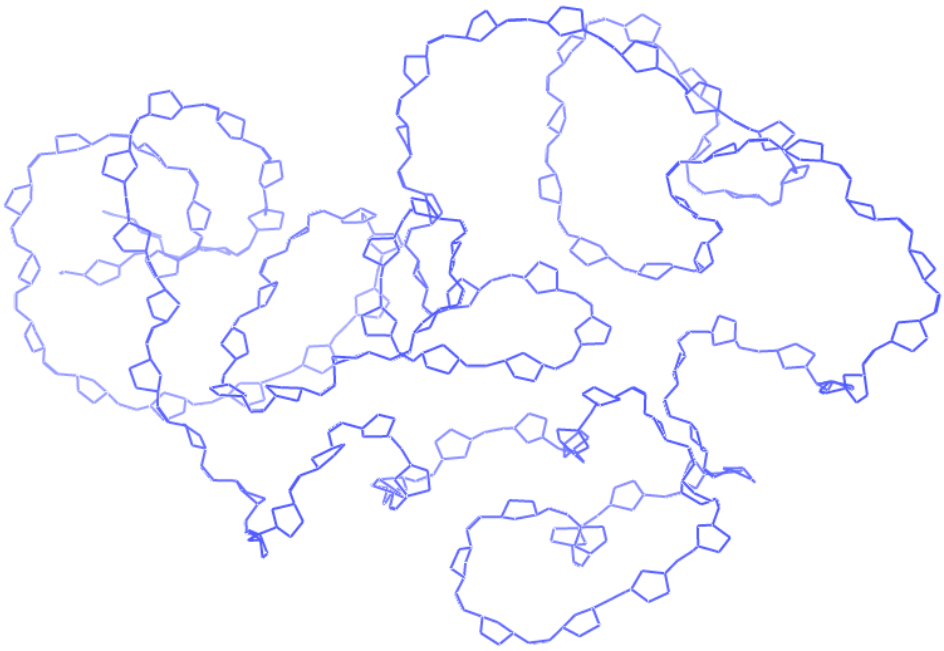


(a)

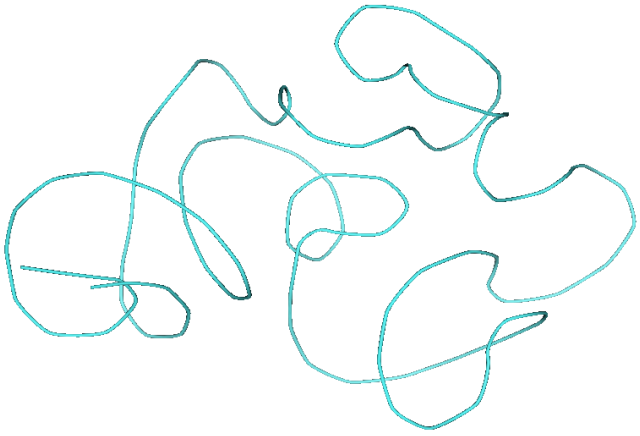


(b)

Figure 35: Final conformation of CM1-100 showing (a) the backbone and (b) a rendering of the backbone.

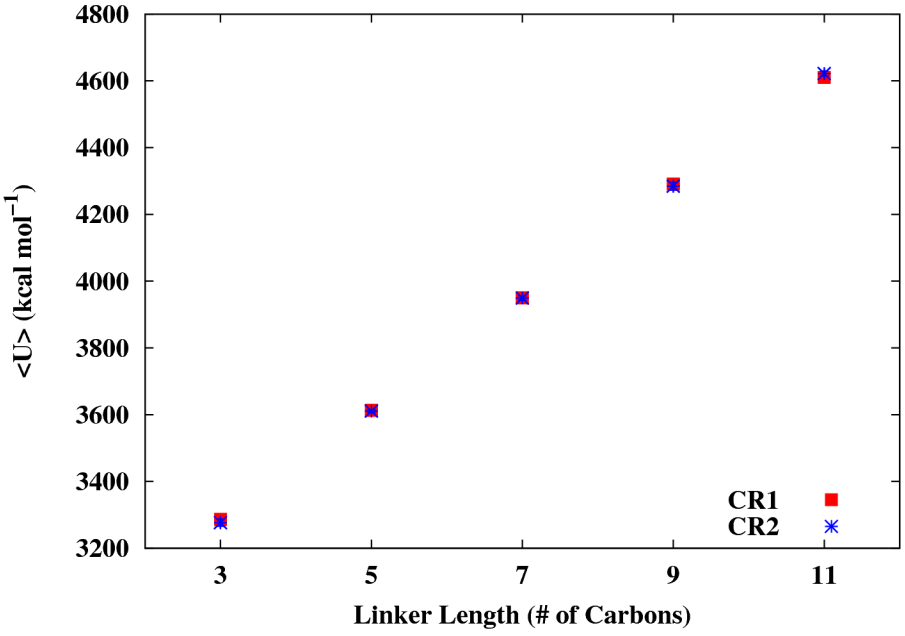


(a)

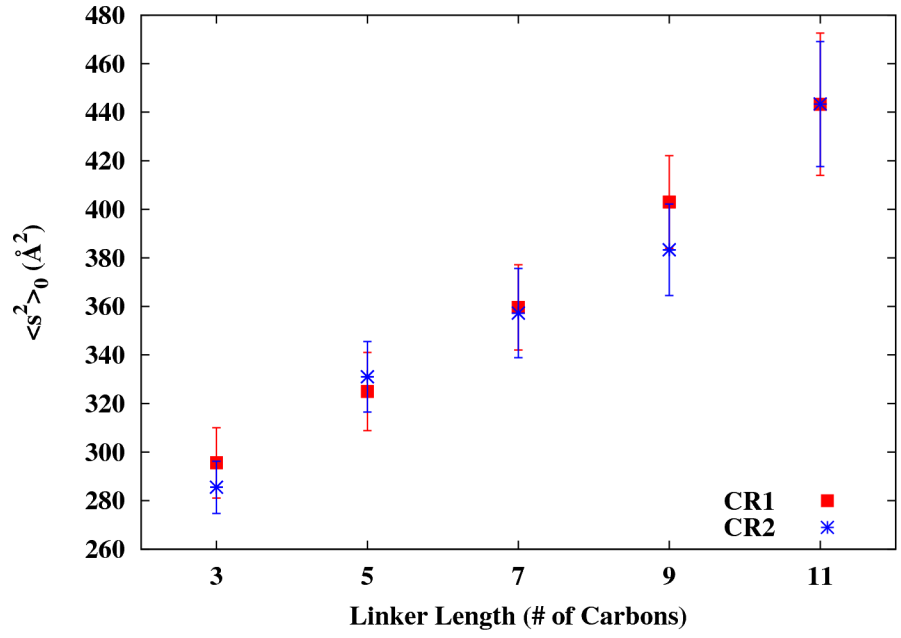


(b)

Figure 36: Final conformation of CM2-100 showing (a) the backbone and (b) a rendering of the backbone.



(a)



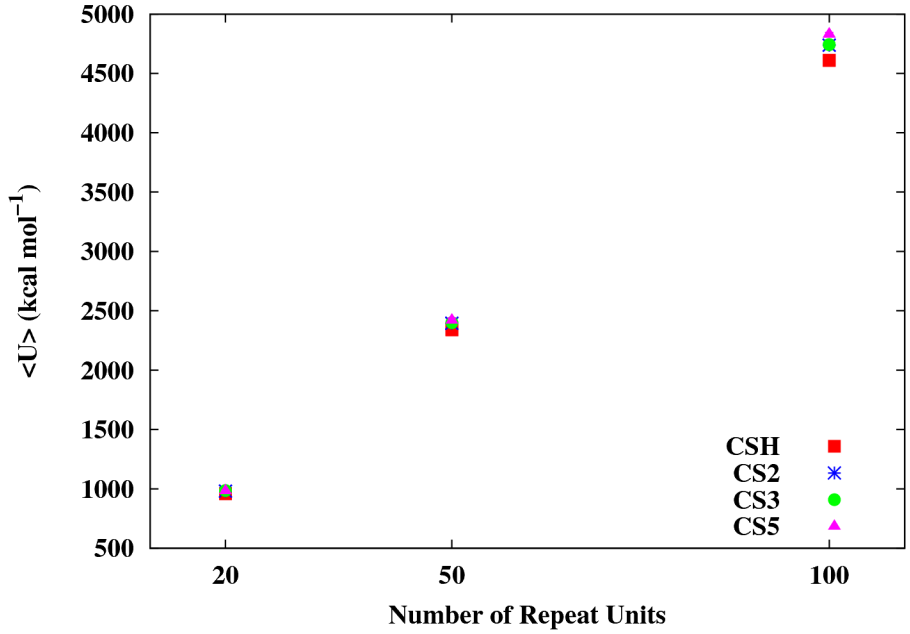
(b)

Figure 37: (a) Average potential energy and (b) $\langle s^2 \rangle_0$ for CR1 and CR2 with 100 repeat units for all linker lengths.

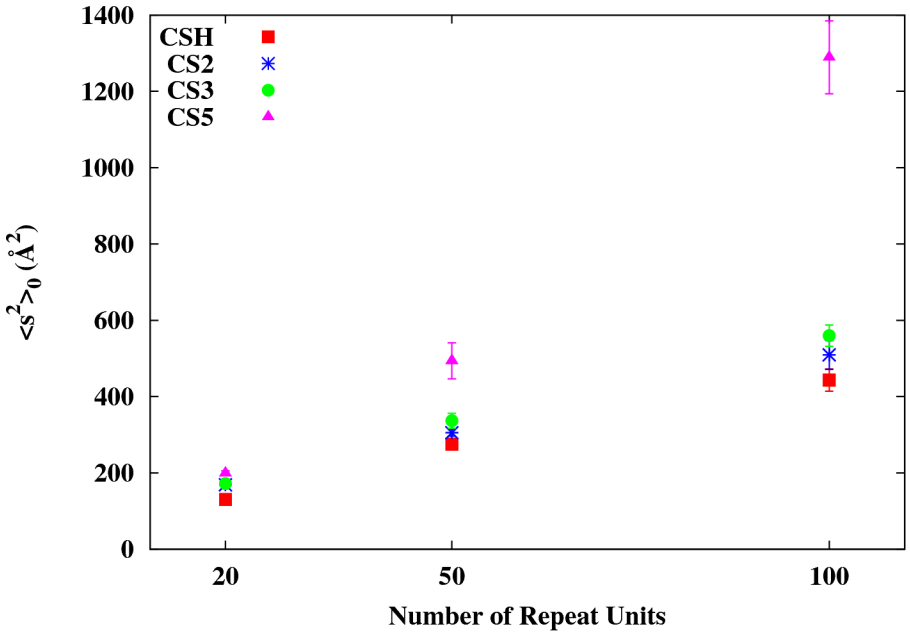
of magnitude compared with the backbone.

Figure 38 shows the effect of catalyst size on the average potential energy and $\langle s^2 \rangle_0$. Unlike with the linker length, the energy does not change significantly for the various catalyst sizes. This would be expected, as van der Waals interactions are now the only part of the potential that is affected by changing the van der Waals radius of the catalyst site. Even at 100 repeat units, the energy for CS5 is only slightly higher than that of the other catalyst sizes, and this is most likely due to increased van der Waals energy due to steric crowding on the large catalyst site. Additionally, the same linearity with energy is present as it was for the different linker lengths. This would show that energetically, there is a constant amount of energy per atom. The most interesting effect of catalyst size, however, is in $\langle s^2 \rangle_0$.

Figure 38(b) shows that CSH, CS2, and CS3 all have a similar behavior for $\langle s^2 \rangle_0$, whereas CS5 shows substantially different behavior. For random coil polymers, the mean square radius of gyration should be linear with molecular weight. This behavior was found for each catalyst size, even though $\langle s^2 \rangle_0$ was much higher for CS5 for 100 repeat units. This increased $\langle s^2 \rangle_0$ could actually be indicative of a conformation that falls between a random coil and a rigid rod. However, an interesting conformation emerged for the CS5 system, which is shown in Figure 39 for both CR1 and CR2. What is interesting about this conformation is that not only is it unique compared to all of the other conformations of this polymer, it seems that there is some order to the backbone where it forms large helical loops. This was not seen in any of the other configurations that were simulated. Additionally, the CR2 CS5 simulation started as an elongated chain and formed an almost identical structure as CR1 CS5. This is very interesting because it shows that making the side group large enough will cause the backbone to start to generate ordered sections of the backbone. This indicates that the hypothesis for a bulky side group is very important for the formation of a helix, and it would seem that a bulky enough side group will form regional helical structures. However, this is still a random coil polymer based on the relationship between

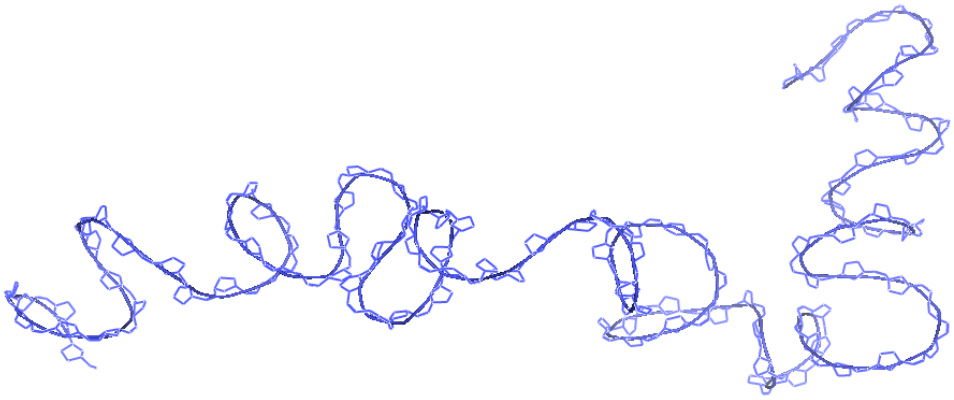


(a)



(b)

Figure 38: (a) Average potential energy and (b) $\langle s^2 \rangle_0$ for CR1 with 100 repeat units for all catalyst sizes.



(a)

(b)

Figure 39: Snapshots of the backbone conformations at the end of the (a) CR1 and (b) CR2 configurations with CS5 and 100 repeat units. The backbones have been traced with a cylindrical line to see them more easily.

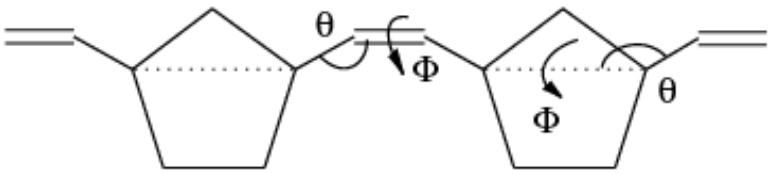


Figure 40: The torsion angles (Φ) and bond angles (θ) on the assumed non-rotatable bonds. The dotted line represents the "pseudo-bond" across the backbone cyclopentane ring.

$\langle s^2 \rangle_0$ and MW, even though there is some order in the backbone.

4.3.3 Non-rotatable Bonds

The results of the accessibility scores, energies, and radii of gyration definitively show that there is no significant difference in conformation between all of the simulations. Additionally, it seems to disprove the hypotheses of what is required for a helical polymer, and it confirms that ROMP PNB does not form a helical conformation. However, upon closer examination, it may be that one of the assumptions for these configurations is invalid. That assumption is the idea that the pseudo-bond that lies across the cyclopentane ring on the polymer backbone is non-rotatable. The original assumption was that the cyclopentane ring is constrained enough to be considered non-rotatable. For an isolated cyclopentane ring, there is an inherent pseudorotation of the molecule, i.e., the ring is flexible. It was thought that by coupling a cyclopentane ring to a large polymer backbone would remove the flexibility of the ring, but this may not be the case.

An examination into the flexibility of the cyclopentane ring in the polymer backbone was necessary to ensure that this can be considered a non-rotatable bond. The best way to do this is to compare the torsion angle of a backbone double bond, which is definitely non-rotatable, with that of a backbone cyclopentane ring (Figure 40). The results are shown in Table 13. The double bond torsion angle has a mean value that is roughly 0° with a standard deviation of 7.5°, which corresponds to roughly ±3°. Contrast that to the torsion angle about the cyclopentane ring which has a mean that is not 0°, and it has a higher standard deviation. In reality, the reason for this is that the cyclopentane ring is not in a planar

Table 13: Simulation average ($\bar{x}$) and standard deviation (s) for the torsion angle (Φ) and the bond angle (θ) for a double bond in the ROMP PNB backbone, an approximated pseudo-bond across a cyclopentane ring in the ROMP PNB backbone, and an approximated pseudo-bond across an isolated cyclopentane ring (Figure 40).

Bond	$\bar{x}$ (°)	s (°)
Double Bond Φ	2	7.5
ROMP PNB Cyclopentane Pseudo-bond Φ	17	11
Cyclopentane Pseudo-bond Φ	1	19
Double Bond θ	127	3
ROMP PNB Cyclopentane Pseudo-bond θ	145	7
Cyclopentane Pseudo-bond θ	123	19

conformation, so the mean torsion angle will differ depending on its actual conformation. However, the standard deviation is higher because it actually still pseudorotates, albeit slower than that of an isolated molecule, which causes it to change the equilibrium value of the torsion angle. This equilibrium value has a standard deviation that is closer to that of the double bond, but this flipping between conformational states is seen as a higher standard deviation. It is still not as great as an isolated molecule, which has a much higher standard deviation due to the greater degree of the pseudorotation.

Additionally, the actual bond angle could play a part in the flexibility of the bond. For the double bond, the bond angle is constant with a small degree of flexibility. However, the bond angle between the pseudobond across the cyclopentane ring and the single bond of the backbone has a much greater degree of variability. It is unknown as to whether or not this bond angle stiffness is necessary in obtaining a helical conformation, but the cyclopentane pseudo-bond cannot be considered non-rotatable, which was one of the key hypotheses for achieving a helix-kink conformation.

4.4 Summary

The accessibility of ROMP PNB was examined. MD was used to sample conformation space, and alpha shapes were used to characterize accessibility. This was done by determining whether a catalyst site dummy atom is accessible to the external surface of the polymer support. It was found that all of the conformations of ROMP PNB form a random coil. It was also found that a very bulky catalyst site causes the polymer to adopt a helical secondary structure. It was determined that the presence of the random coil conformations was due to the fact that the cyclopentane ring on the backbone is not as rigid as it was originally assumed.

CHAPTER V

ROMP

POLY(N-UNDECYL-*EXO*-NORBORNENE-5,6-DICARBOXIMIDE)

5.1 Introduction

5.1.1 Objectives

1. Using the tools and results from ROMP PNB, model a more functionalized ROMP PNB in an attempt to improve accessibility by increasing polymer dimensions

5.1.2 Motivation

One of the major hypotheses for a helix-kink polymer is that it must have alternating rotatable/non-rotatable bonds. It appears that ROMP PNB did not conform to this requirement due to the lack of rigidity in the cyclopentane ring in the backbone. However, it may be that it is possible to functionalize the cyclopentane ring in such a manner as to remove the flexibility of the backbone cyclopentane ring. Recent experimental studies have shown results that may suggest that the polymerization of a repeat unit with the addition of a second ring fused to the cyclopentane ring could produce a polymer that has a helix-kink conformation[83, 84, 85, 86, 87, 88]. In order to test this, ROMP poly(N-undecyl-*exo*-norbornene-5,6-dicarboximide) (PNBDC) was simulated to determine if the results could provide insight into these promising experimental studies.

5.1.3 Scope

This work will be limited to ROMP PNBDC as the polymer under investigation. For this study, only the case with the undecyl side group with CSH will be examined. This is done to allow for a direct comparison to the simulations performed in the previous chapters of this work. Other side groups will be discussed, however, they are not simulated because

there would be no way to compare the results to the work on ROMP PNB.

5.1.4 Background

ROMP PNB, as it was modeled in the previous chapter, is the simplest form of this broad polymer family. ROMP PNB has been created with a number of functional groups in various positions about the cyclopentane ring. One such repeat unit contains an additional fused ring on the fourth and fifth carbons of the cyclopentane ring. Such strained ring monomers have been successfully polymerized using the ruthenium-based Grubbs catalyst[89], and there have been some detailed studies in recent years of the properties of such polymers. These studies have produced results that show properties that were indicative of helix-kink conformers in vinylic PNB.

The first property that indicates a possible helical conformation is variance in T_g based on backbone structure. Hino and coworkers have reported the T_g of various ROMP PNB polymers[87, 88]. The first study reported properties of ROMP PNB derivatives with cyclic carbonate units as side groups co-polymerized with cyclooctene. One of the polymers had the cyclic carbonate attached to one carbon on the cyclopentane ring, and the other attached to two carbons on the cyclopentane ring, increasing the cyclopentane ring strain. The study examined the effect on T_g and volume expansion. The most interesting aspect of this data is that both the T_g and the partial molar volume increased with increasing amounts of the functionalized PNB, with the doubly-linked carbonate showing increased T_g and volume expansion. Additionally, the T_g was higher by a factor of two for the lowest level of cyclooctene co-polymer. The second study increased the number of copolymers that were examined and served to support the first study[88]. Note that the volume expansion as it pertains to this study was compared to the monomer, but individual polymers could be compared based on the differences in expansion between polymers.

Another piece of interesting data that would support a structure with intermolecular order is slightly more empirical, as it cannot be quantified as a numerical value. However, Lin

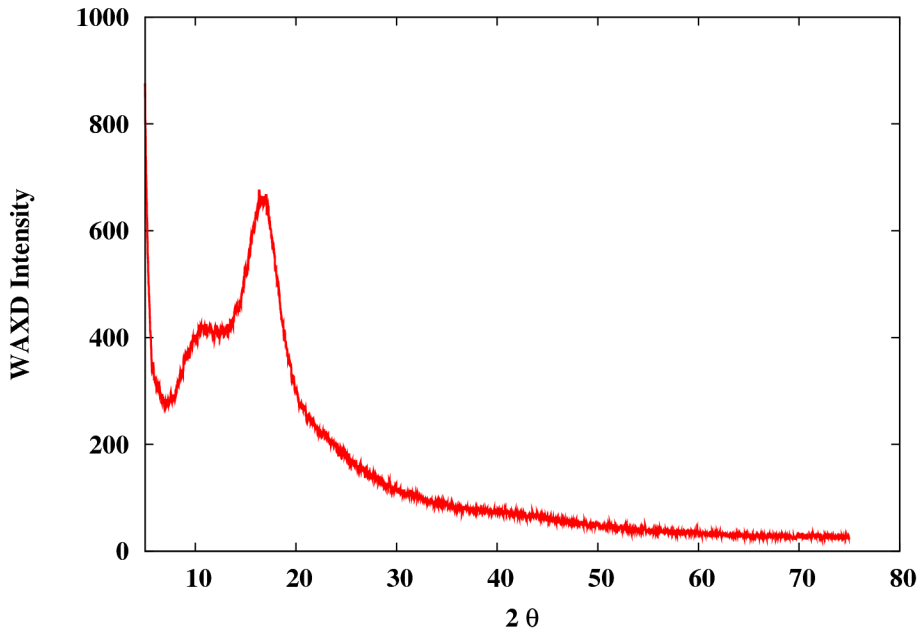


Figure 41: WAXD spectra of ROMP poly(*exo*-N-cyclohexyl norbornene-5,6-dicarboximide).

and coworkers showed that a pyrrolidine ring fused to the backbone cyclopentane ring exhibited rod-like structures in atomic force microscopy images[84]. Additionally, the TGA and DSC experiments showed stability up to 280 °C with no apparent T_g in the examined temperature range. This particular polymer had a dipolar chromophore pendant attached to the backbone of this molecule for use in a second-order nonlinear optical device. There is an enhancement of the optical properties when such pendant groups are attached to rigid backbones. In an earlier study, Sattigeri and coworkers showed that this same polymer exhibited this enhanced nonlinearity for a dipolar pendant group attached to the pyrrolidine ring, indicating that the backbone could have a rigid structure[83]. This same study used MD in an attempt to show a probable conformation, however, these modeling results are inconclusive, as the simulations were only run for 20 ps, which means that the structures did not vary much from their starting conformations.

One final aspect of ROMP PNB with a fused ring attached to the backbone cyclopentane rings is the WAXD spectra of such polymers. As previously mentioned, vinylic PNB

showed a WAXD spectra (Figures 4 and 5) that is unique to polymers, as the amorphous halo is split into two distinct peaks, one for intermolecular interactions and the other for intramolecular interactions. Norbornene monomers have been functionalized with a dicarboximide attached to the cyclopentane ring[86, 85]. The polymers that were produced via ROMP polymerization show high values for T_g, again by a significant margin over an unfunctionalized ROMP PNB, as well as higher fractional free volume, both of which are known properties of helical polymers. In addition, WAXD was performed on ROMP poly(*exo,endo*-N-cyclohexyl norbornene-5,6-dicarboximide)[90]. This WAXD pattern is identical to that obtained for the same polymer without endo configurations by the Jones and Weck groups at the Georgia Institute of Technology (Figure 41). This WAXD spectra shows the same peak splitting as was seen for vinylic PNB, though the peaks are not as sharply defined, as compared to Figures 4 and 5.

These results suggest the possibility of a helical polymer. For T_g, vinylic PNB showed a similar trend, in that it had a high T_g. The fact that T_g can increase so drastically based on decreasing the presence of a flexible co-monomer suggests that there could be increased intermolecular order. Additionally, the AFM and optical results indicate a rigid conformation. On top of this, the WAXD data shows a split peak rather than an amorphous halo, which is indicative of a helix-kink polymer. Any of these results by themselves does not guarantee a helical polymer, however, all of these results together very strongly suggests potential for an ordered conformation of a ROMP PNB polymer with ring-functionalized cyclopentane groups.

5.2 Simulation Methodology

5.2.1 Polymer and Starting Conformations

ROMP poly(N-undecyl-*exo*-norbornene-5,6-dicarboximide), shown in Figure 42, was used to test the hypothesis that the added strain from a second ring fused to the backbone cyclopentane ring of ROMP PNB will make the cyclopentane ring behave as a non-rotatable

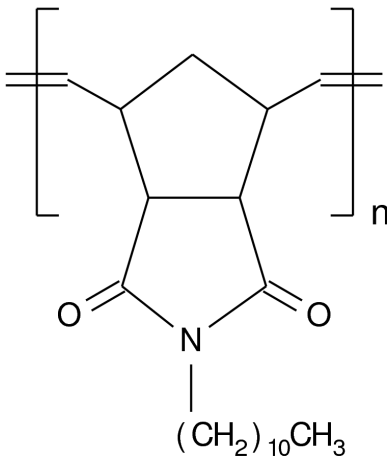


Figure 42: Structure of ROMP poly(N-undecyl-*exo*-norbornene-5,6-dicarboximide).

bond, thus producing a helical conformation. This particular repeat unit was chosen in order to be able to have a comparison to experimental work that was previously performed by other groups[86, 85]. Additionally, the N-undecyl side group was chosen over the cyclohexane ring used in experiment because it will offer a more direct comparison to the other ROMP PNB simulations performed in this work, particularly for accessibility, as changing the length of the side group was previously shown to have a significant effect on the accessibility.

The simulations were created for each of the eight possible configurations of ROMP PNBDC. These configurations are the same as those used for the ROMP PNB simulations (Figure 11). Additionally, each configuration was modeled with the same two starting conformations, that of a helix and that of an elongated chain. These starting conformations will again be designated by the numbers *1* and *2*, where 1 will designate a helical starting conformation, and 2 will designated an elongated starting conformation. Additionally, polymers with 20, 50, and 100 repeat units were simulated. Only the CSH case was considered for these polymers.

5.2.2 Molecular Dynamics and Simulation Analysis

The simulation details are the same as those for ROMP PNB. The simulations were again run for 500 ps, and the final 100 ps were used for the data analysis. As before, the equilibration was determined by the mean squared radius of gyration, and when necessary, the simulations were run for an additional 300 ps if the original 500 ps did not fully equilibrate. The simulations were repeated 10 times for each starting conformation that was used to establish the 90% confidence interval for that particular conformation. There were some cases where there were two different final conformations generated for a given starting conformations, and for these configurations, there were 5 repeated simulations for each final conformation.

The accessibility of ROMP PNBDC was characterized through the use of alpha shapes. The similar side group on each repeat unit allows for a direct comparison of this polymer to the original ROMP PNB. The assumptions used for the previous simulations also hold true for this case as well. When comparing the configurations for ROMP PNB to ROMP PNBDC, the configurations will be prefixed with *T1-* and *T2-*, respectively. Because the ROMP PNB configurations all had statistically similar results for the radius of gyration, potential energy, and accessibility, these will be compared to ROMP PNBDC as an average of each property over all of the configurations and conformations.

5.3 Results and Discussion

5.3.1 Backbone Cyclopentane Ring Flexibility

The first important property to examine is the flexibility of the cyclopentane ring. If the addition of another ring fused to the cyclopentane ring does not reduce the flexibility of the cyclopentane ring, there would be no reason to suspect a conformational change from ROMP PNB. Table 14 is an expansion of Table 13 to include the results of a ROMP PNBDC simulation. The first thing that is apparent from these results is that the only major change is the decrease in the standard deviation of the bond angle for the cyclopentane ring for

Table 14: Simulation average ($\bar{x}$) and standard deviation (s) for the torsion angle (Φ) and the bond angle (θ) for a double bond in the ROMP PNB backbone, an approximated pseudo-bond across a cyclopentane ring in the ROMP PNB and ROMP PNBDC backbones, and an approximated pseudo-bond across an isolated cyclopentane ring.

Bond	$\bar{x}$ (°)	s (°)
Double Bond Φ	2	7.5
ROMP PNB Cyclopentane Pseudo-bond Φ	17	11
ROMP PNBDC Cyclopentane Pseudo-bond Φ	15	11
Cyclopentane Pseudo-bond Φ	1	19
Double Bond θ	127	3
ROMP PNB Cyclopentane Pseudo-bond θ	145	7
ROMP PNBDC Cyclopentane Pseudo-bond θ	145	4
Cyclopentane Pseudo-bond θ	123	19

ROMP PNBDC. However, while the addition of the fused ring does not decrease the standard deviation of the torsion angle, the cyclopentane groups on the backbone of ROMP PNBDC no longer "flip" between cyclopentane conformations. There is still some torsional motion, but once the simulation has equilibrated, the cyclopentane conformation is going to be constant for the remainder of the simulation. This definitely shows a reduction in cyclopentane ring flexibility.

5.3.2 ROMP PNBDC Simulation Results

5.3.2.1 Multiple Final Conformations

For ROMP PNBDC, there was one configuration that had two separate final conformations. For all polymer sizes, the elongated TR conformation had a final conformation that was more expanded than the other, however, this conformation had a higher energy by 50-100 kcal mol^{-1}. This is likely due to the elongated starting conformation sampling a local minimum that did not allow it to contract to the lower energy conformation. Therefore, these were removed from the average, so all TR2 results are those for only five simulations, rather than ten. The only effect this will have is that the confidence interval will be broader due to the reduced sample size.

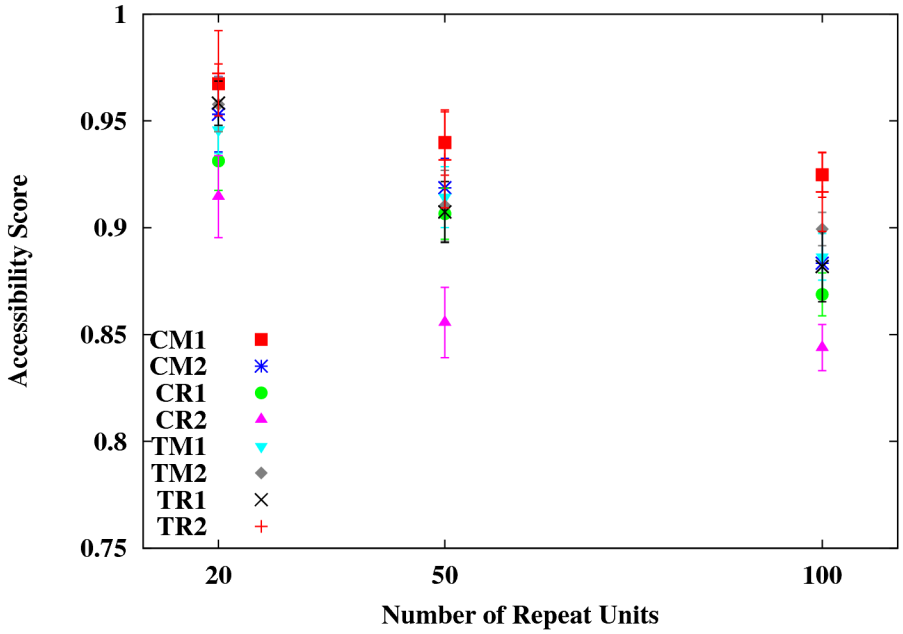
5.3.2.2 *Accessibility*

ROMP PNBDC accessibility scores are shown in Figure 43. In contrast to Figure 25, it is immediately apparent that there is a wider range of accessibility scores among the different ROMP PNBDC configurations. Figure 43(b) shows this more closely, where the range of values is definitely dependent on both configuration and starting conformation. From these results, CM has a higher accessibility of the cis configurations, while the trans configurations are less dependent on starting conformation. In this case, the average CM accessibility is on par with TM and TR, and CR is the lowest.

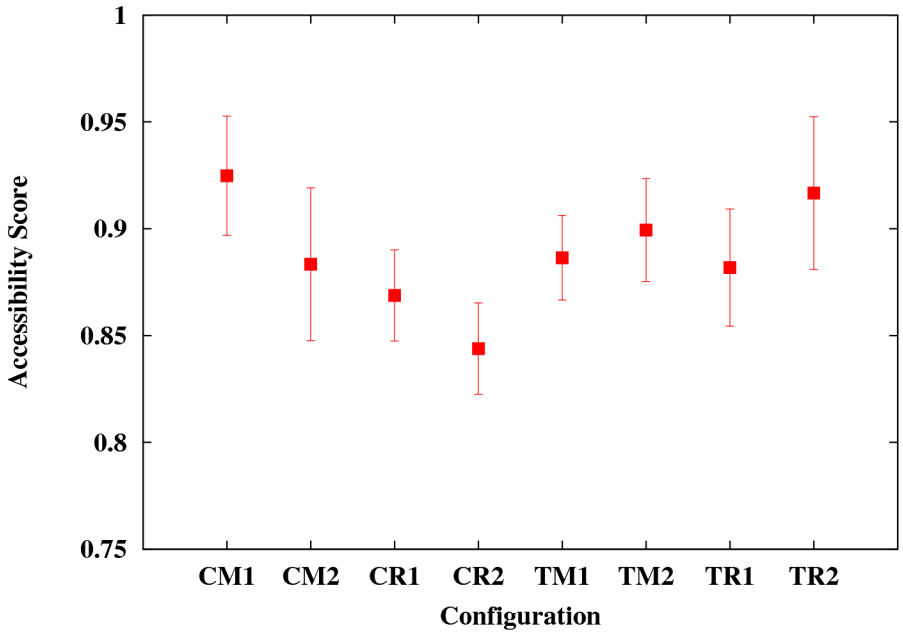
5.3.2.3 *Energies and Conformations*

Figure 44 shows the average potential energies for the ROMP PNBDC configurations. From Figure 44(a), the potential energies seem to be the same, and in fact, they are statistically similar for the 20 and 50 repeat unit polymers. However, as shown by Figure 44(b), there is an interesting trend for the potential energies. Unlike accessibility, the starting conformation has no impact on the potential energy. Additionally, the trans configurations are more energetically favorable than the cis configurations by roughly 100 kcal mol^{-1}. In this case, the separation for statistical significance is split into the cis and trans configurations, i.e., the average of the cis configurations is statistically different than that of the trans.

Even with the results of the potential energies, conformational analysis is necessary to determine the actual differences between the different final conformations. As with ROMP PNB, $\langle s^2 \rangle_0$ will be used to analyze the conformational dimensions for ROMP PNBDC. These results are very different from ROMP PNB. While ROMP PNB had statistically similar $\langle s^2 \rangle_0$, there are significant differences between the configurations for ROMP PNBDC, as shown in Figure 45(a). Figure 45(b) shows $\langle s^2 \rangle_0$ for the 100 repeat unit case. The first thing to note is that there is a dependence on starting conformation for CM, TM, and TR. This is a warning sign that the starting conformation is stuck in a local minima, which indicates that the final conformations with higher potential energies are most likely not

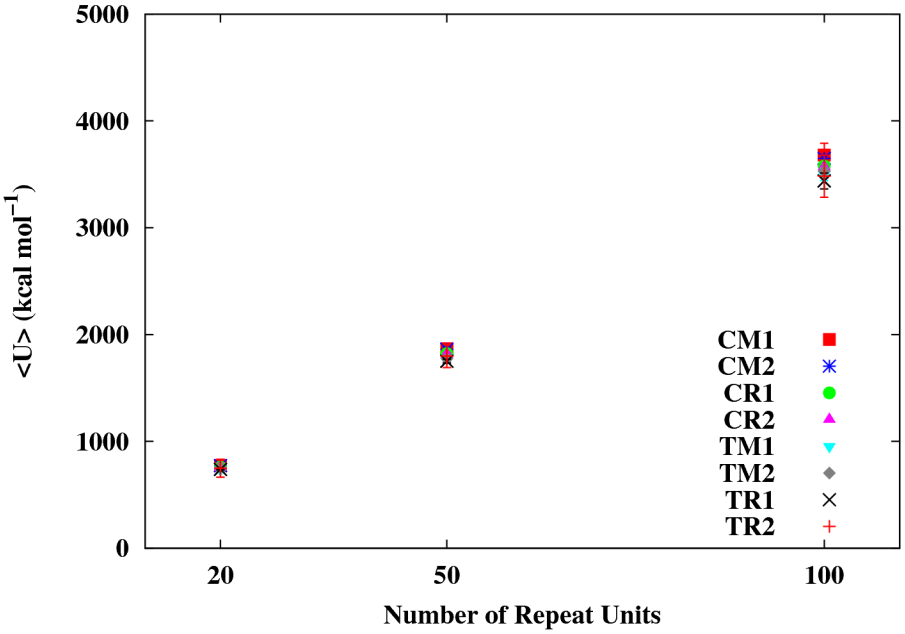


(a)

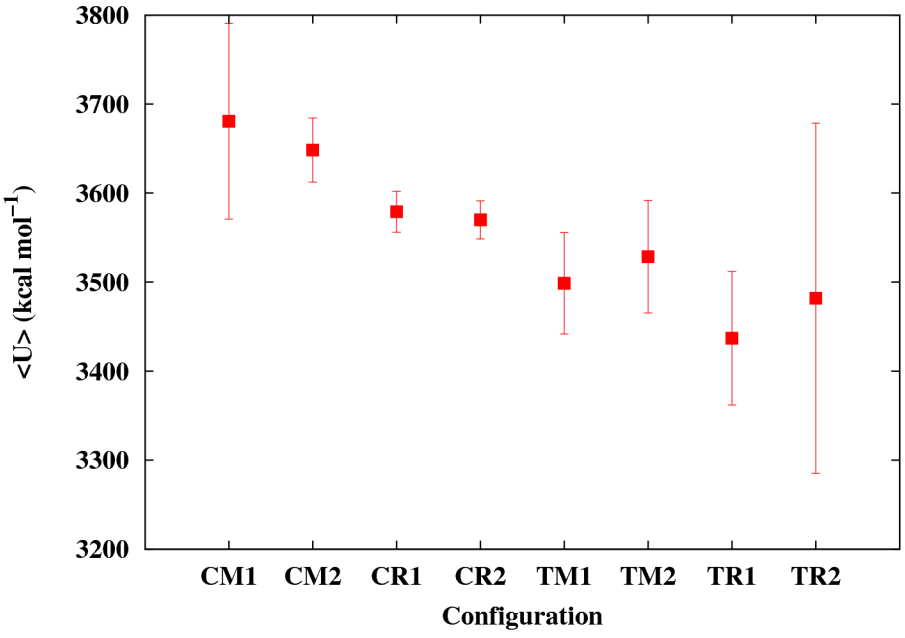


(b)

Figure 43: Accessibility scores for all ROMP PNBDC configurations with (a) 20, 50, and 100 repeat units and (b) for 100 repeat units with RS1.4.

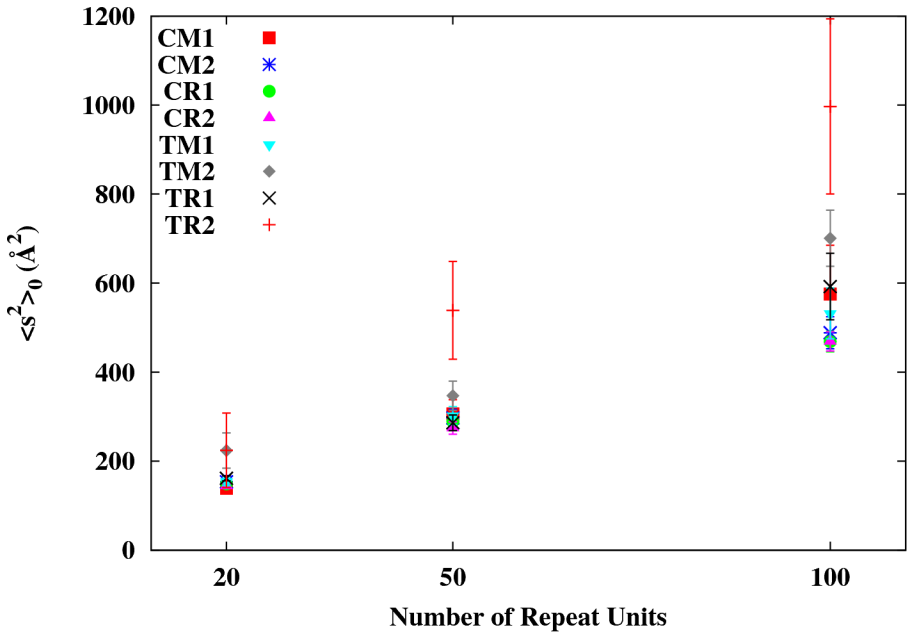


(a)

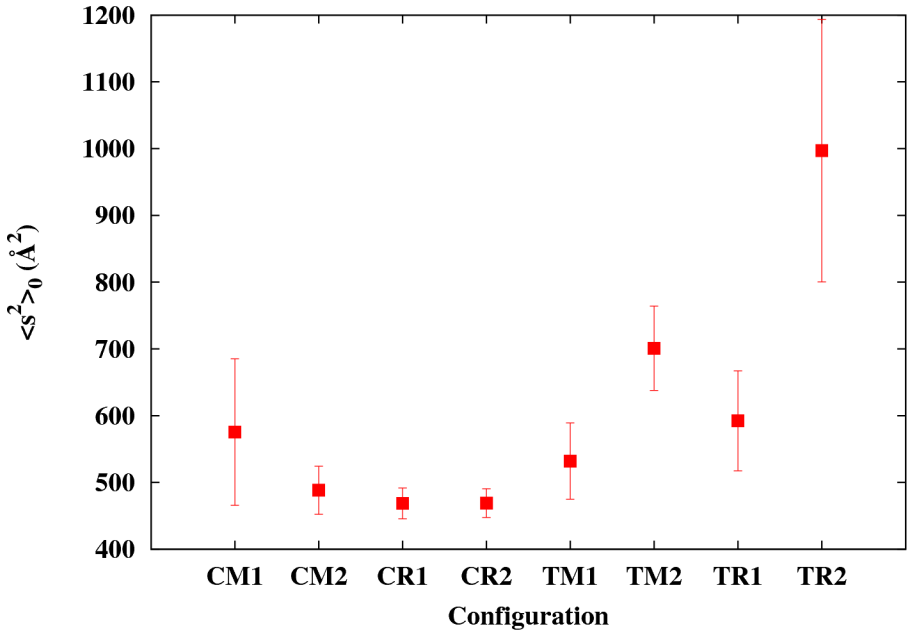


(b)

Figure 44: Average potential energies for all ROMP PNBDC configurations with (a) 20, 50, and 100 repeat units and (b) for 100 repeat units.



(a)



(b)

Figure 45: Mean square radius of gyration for all ROMP PNBDC configurations with (a) 20, 50, and 100 repeat units and (b) for 100 repeat units.

converged. This can happen with MD, where convergence of an unrealistic starting conformation will take an inordinate amount of simulation time to actually converge. This underscores the importance of multiple starting conformations. In this case, one conformation will generate a realistic final conformation, while the other will just not converge in the same amount of time. While the energies are statistically similar, they are averages of 5-10 simulations of 1000 data points for each simulation, which means that it is essentially an average over 5000-10000 data points. If the actual points themselves are taken as the energy value, it can shed some light on which conformation is the correct one. In this case, CM1, TM2, and TR2 all have slightly higher potential energies than their counterparts, so it is likely that the actual conformations are the ones with the lower energies. One thing that the dependence on starting conformation indicates is that ROMP PNBDC is much more stiff than its ROMP PNB counterparts, which is an indication that the fused dicarboximide ring is adding rigidity to the polymer.

Coupling the potential energy and $\langle s^2 \rangle_0$ results, it is likely that CM2, TM1, and TM2 are the low energy conformations. Additionally, there seems to be no difference between CR1 and CR2 conformations. However, the only way to truly see the full picture is to actually examine the final conformations, as shown in Figure 46. It should be noted that all of the conformations shown here are representative of all of the obtained final conformations. The first thing that can be noticed is that there is little conformational difference between CM2, CR1, CR2, TM1, and TM2. This would show that even though these conformations have different $\langle s^2 \rangle_0$, it is only by a slight degree of polymer expansion, but they are all random coils. The CM1 conformation should be considered an artifact of the simulation, as it has a much higher energy than CM2. Although CM1 is a helix, it is most likely that its starting conformation was such that it locked the polymer into the helical conformation, even though this is not the correct final conformation. This was very similar to what was found with the previous TSMC simulations. In this case, it is most likely due to the decreased flexibility in the cyclopentane ring due to the fused dicarboximide ring.

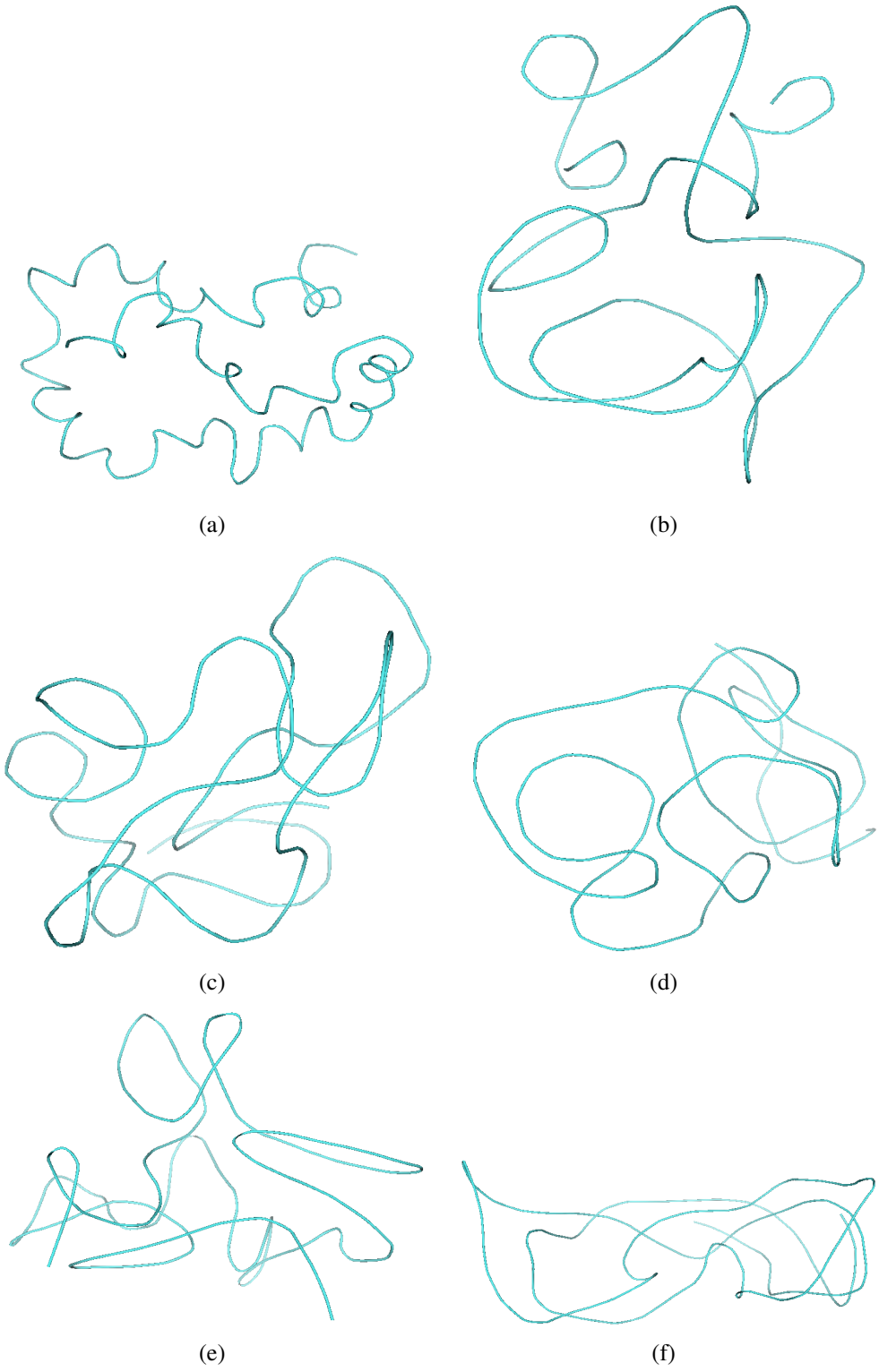
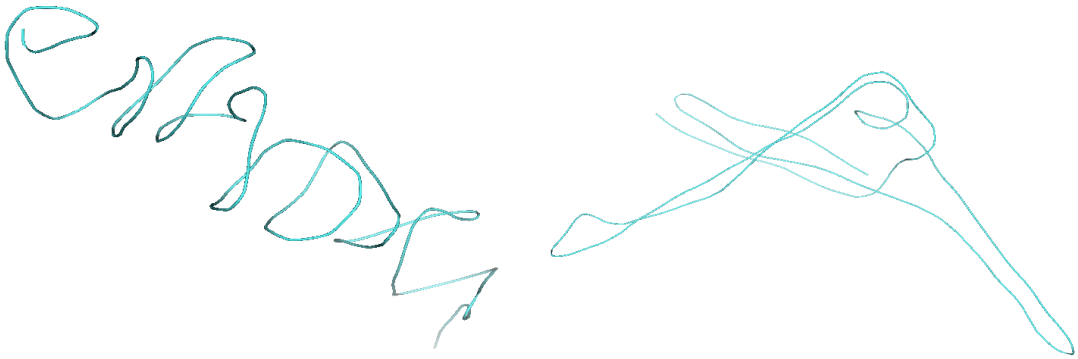


Figure 46: Renderings of the ROMP PNBDC backbones of final conformations of 100 repeat units for (a) CM1, (b) CM2, (c) CR1, (d) CR2, (e) TM1, (f) TM2, (g) TR1, and (h) TR2.



(g) (h)

Figure 46: Renderings of the ROMP PNBDC backbones of final conformations of 100 repeat units for (a) CM1, (b) CM2, (c) CR1, (d) CR2, (e) TM1, (f) TM2, (g) TR1, and (h) TR2, *cont.*

The most interesting result from these conformations is the TR configuration. TR1 shows a loose helix, while TR2 shows a seemingly unrealistic structure with long U-bends in the backbone. This highlights the purpose of multiple starting conformations. TR2 was unable to form a helix in the simulation, but it also did not become a random coil, which should have been the case if the helix was constrained, as with CM1. However, TR1 does not match the starting conformation, as the loops are much more broad, unlike CM1. In CM1, the polymer could not "unwind" itself, but the helix loops are exactly the size of those from the starting conformation. In TR1, the starting conformation was able to relax in its low energy conformation, and interestingly enough, this conformation had the lowest average potential energy of all of those studied. Also, this polymer "relaxed" from its starting conformation. Both the starting and final conformations had eight helical loops. These can be seen in Figure 46(g) for the final conformation. What is interesting is that the helical starting conformation collapsed into the final structure. The eight regular loops of the starting conformation collapsed the helix into half the full length, and the loops lost their regularity.

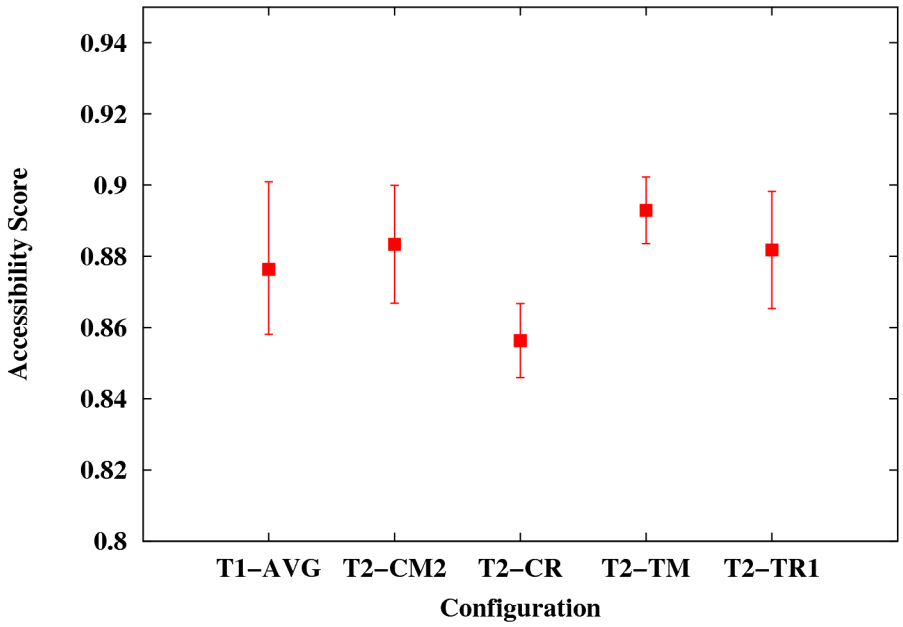
This loose helical conformation can explain a number of the interesting properties for the ROMP PNB polymers with rings doubly-attached to the backbone cyclopentane ring.

All of the studies that were discussed above for the T_g, AFM results, and WAXD spectra of ROMP PNB polymers with fused rings all used a generation of the Grubbs catalyst to polymerize their PNB monomer. These catalysts have been characterized as producing PNB polymers with low cis content (less than 30% cis, depending on the monomer and catalyst generation), and they have been characterized as having predominantly racemic diad for the trans content[72]. This means that all of these studies most likely were performed on predominantly TR structures, which would show this unique behavior.

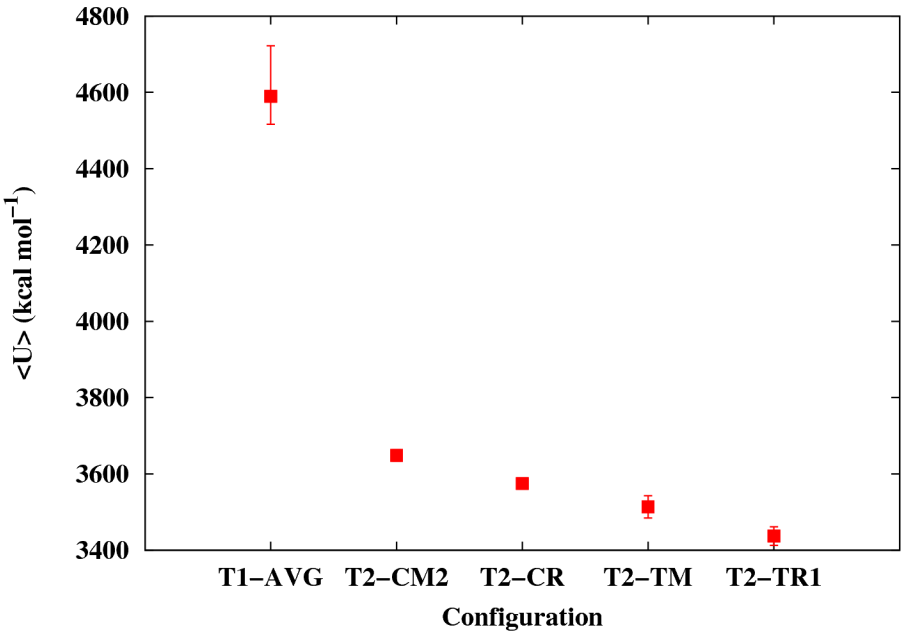
5.3.3 Comparison of ROMP PNBDC and ROMP PNB

To compare the properties and performance as a catalyst support for ROMP PNB and ROMP PNBDC, the proper conformations must be compared. Because the ROMP PNB simulations showed no conformational difference based on configuration, it is presented as an average, labeled T1-AVG. For ROMP PNBDC, the CR and TM results will be averaged, while only CM2 and TR1 will be shown, as these were the realistic conformations for these configurations. Additionally, only the 100 repeat unit polymers will be compared.

Figure 47(a) shows the accessibility scores for T1 and T2 polymers. Unfortunately, there is no statistical difference between any of the shown values. However, this is not true for the average potential energy or the mean square radius of gyration. The average potential energy (Figure 47(b)) is a very interesting property in this case, as T1 is higher than all of the T2 energies by about 1000 kcal mol^{-1}. This is the exact opposite of the expected value. Generally, adding more atoms will show an increase in potential energy, as the bond stretch and angle bend energies are additive positive quantities. For both of these energy terms, the potential energy of a molecule will be higher than one with fewer atoms. This means that there are three possible (non-additive) contributions to the energy which could be causing a decrease in the potential energy for the T2 polymers, which are the electrostatic and van der Waals pair interactions and the bond torsion energies. An examination of the potential energy components revealed that for the bond stretch, angle



(a)



(b)

Figure 47: Comparison between ROMP PNB (T1) and ROMP PNBDC (T2) for 100 repeat units with CSH and L11 for (a) accessibility for RS1.4, (b) average potential energy, and (c) mean square radius of gyration.

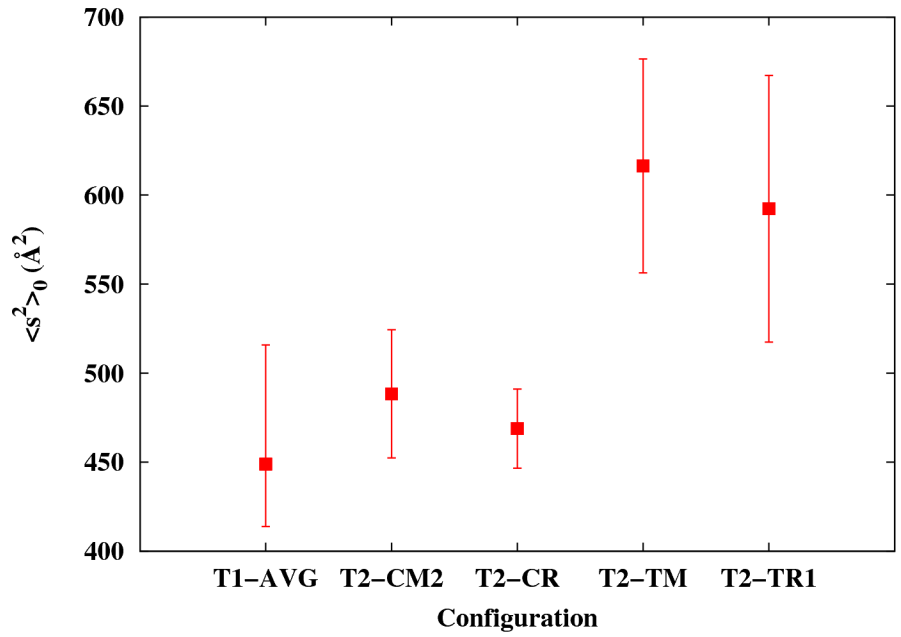


Figure 47: Comparison between ROMP PNB (T1) and ROMP PNBDC (T2) for 100 repeat units with CSH and L11 for (a) accessibility for RS1.4, (b) average potential energy, and (c) mean square radius of gyration, *cont.*

bend, torsion, stretch-bend, out-of-plane, and van der Waals energies are all higher for T2 compared to T1, which means that the individual energy terms are all behaving as expected. The only exception to this is the electrostatic interaction energy. The electrostatics for T2 are more favorable than that of T1, most likely due to the addition of the polar nitrogen and additional carbonyl oxygen. Since these atoms have partial charges with higher magnitudes than the surrounding carbon/hydrogen atoms, these groups will have a greater effect on the electrostatics. If these atoms are in a favorable geometry relative to each other, it would cause the electrostatic energy to decrease, which would be seen as a large decrease in total potential energy.

Figure 47 shows $\langle s^2 \rangle_0$ for the T1 and T2 polymers. The trans T2 polymers both have much higher $\langle s^2 \rangle_0$ values than the cis T2 and average T1 values, which means that the trans T2 polymers are more expanded than the others. It would seem that for this case, the

trans configuration would be more desirable as an immobilized catalyst support. Additionally, these results show that the helix formed by the T2-TR1 configuration is too broad to differentiate itself from a random coil at these polymer dimensions.

5.3.4 Limitations of the Accessibility Analysis

It would seem that the accessibility analysis using alpha shapes to determine an accessibility score is not fine enough to account for the conformational differences between polymers. However, it is more likely that this method is valid, but it was applied to make comparisons that were fundamentally insensitive to accessibility. Trends for polymer size, linker length, catalyst size, and reactant size were found that successfully showed the expected relationship for each case with accessibility. However, there was little sensitivity to actual conformation, as was shown with the comparison of the ROMP PNB and ROMP PNBDC accessibility. Even T2-TR1, which showed a loose helical conformation, had similar accessibility with entirely random coil conformations.

The main reason for this lack of resolution in accessibility is that the systems simulated for these comparisons were unintentionally pre-optimized for accessibility. In a dilute system, there will be little interaction with other polymer chains in the bulk. A helical polymer as a bulk system will have a greater accessibility than a random coil in the bulk. However, there will be less of a difference in a dilute system because only intermolecular chain order will affect accessibility. Additionally, the size of the polymers were such that they were not large enough for accessibility to become as much of an issue. A random coil in a dilute system will not be as different from a helical polymer at smaller dimensions. In this case, 100 repeat units is probably not enough to show a large enough difference in accessibility without overlapping confidence intervals. The repeat units would probably have to be increased by an order of magnitude in order to increase the magnitude of this difference, particularly when the helical polymer is such a loose helix. However, the use of 100 repeat units is a useful comparison to experimental data, where immobilized homogeneous catalysts have

been used with less than 100 repeat units as the largest polymer size utilized[69]. Finally, it was shown that linker length has an effect on accessibility, and longer linker lengths have higher accessibility. At smaller linker lengths, again, the conformation differences should be much more apparent in accessibility. However, the simulations that more closely resembled experimental systems were compared, so the linker length used was one which would have a higher accessibility. By having a dilute system with relatively small polymer dimensions and using the longest linker length of those simulated, the difference in accessibility has effectively been smoothed out. However, if these aspects of the simulation are taken into account for future use of this property, it will be much more useful as a tool for comparison.

5.4 Summary

In an effort to change the backbone structure of ROMP PNB to make it more favorable to form a helical conformation, an additional ring was attached to the cyclopentane ring in the ROMP PNB backbone in order to make it less flexible. The results showed that this polymer had a lower average potential energy that ROMP PNB due to electrostatic interactions. Additionally, it was found that the TR configuration of ROMP PNBDC has a very broad helical conformation, which would explain some of the unique properties that have been reported for this particular polymer. Additionally, it was found that accessibility is fairly insensitive for these particular systems, though it is believed that this measurement would be useful under less optimal conditions.

CHAPTER VI

SUPPORTS FOR CO-SALEN CATALYSTS

6.1 Introduction

6.1.1 Objectives

1. Determine an accurate force field for a Co-salen catalyst

2. Determine the optimal support size and configuration of a poly(cyclooctene) polymer support for a Co-salen catalyst

6.1.2 Motivation

The motivation for this work comes from work done by Zheng, Jones, and Weck on an immobilized homogeneous catalyst system where the catalyst that was used was a Co-salen complex, and the support was poly(cyclooctene) (PCO)[91]. Their work found that if PCO was polymerized into a macrocycle, the reactivity was increased compared to PCO that was polymerized as a straight chain. Additionally, only very small amounts of catalysts were required for full reaction yields. They hypothesized that the reason for this is that flexibility of the backbone of the macrocycle coupled with the constrained Co-salen complexes produces a superior geometry which allows for the proper geometric reaction conditions over the polymer system[91].

Because of this previous work, it is desirable to use simulations to provide a recommendation for the optimal macrocycle dimensions for this reaction system. If modeling can support the superiority of the macrocycle geometry over the straight chain, then it should also be able to be used to determine whether, for example, a 3 repeat unit macrocycle would have a more favorable geometry over a 5 repeat unit macrocycle. Once an optimal geometry is chosen, macrocycle production can be targeted to a dimension that

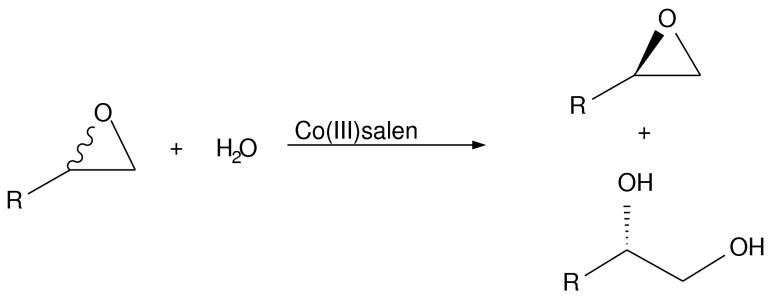


Figure 48: Hydrolytic kinetic resolution of terminal epoxides.

would increase reaction performance even further.

6.1.3 Scope

This work will be limited to three supports. The first is the cyclic oligomeric system proposed by Ready and Jacobsen[92]. The second and third are both poly(cyclooctene) derivatives, where one is polymerized as a straight chain, and the other is polymerized as a macrocycle[91]. Additionally, this work will only be focused on current knowledge of the Co-salen reaction mechanism, i.e., the current theoretical reaction geometry and mechanism is accepted as true.

6.1.4 Background

Producing enantiospecific products of chemical reactions is becoming more important in the chemical industry. This is paramount for the pharmaceutical industry, where one enantiomer produces the desired effect, while the other is ineffective or causes unwanted side-effects. This is important even in the specialty chemical industries for reactants like epichlorohydrin. Therefore, enantiospecific catalysis are becoming more and more desirable for these types of applications. One such catalyst is the chiral M-salen complex, where M is a metal atom that was originally manganese[93], chromium[94], or cobalt[95], though others have been reported in recent years. This work will focus on the Co-salen, though many of the results may hold true for other metal-salen complexes. This catalyst has been

used in the preparation of enantiospecific terminal epoxides via both the asymmetric epoxidation of olefins and the hydrolytic kinetic resolution of racemic terminal epoxides (Figure 48). One of the advantages of this catalyst is that it requires no solvent. A mixture of reactants with a slight excess of water is all that is required to perform this reaction[95].

These reactions are characterized by the enantiomeric excess (*ee*) of one enantiomer over the other. The *ee* for a racemic mixture is 0% and is 100% for a solution with a single enantiomer. As these reactions progress, the *ee* increases because only one enantiomer is involved in the reaction, and it plateaus at a certain value. This value is mainly dependent on the catalyst loading. The catalyst loading for a Co(III)salen is normalized to the amount of cobalt that is in the system. The effect of catalyst loading depends on a number of factors. In this work, *ee* will be used to compare catalyst loading. Generally, at low catalyst loadings, the time for the reaction to reach an *ee* plateau will increase compared to the same catalyst at higher catalyst loading. Additionally, the value of the *ee* plateau can decrease at lower catalyst loading. However, in comparing supports for a catalyst, the support that would allow for greater reactivity at lower catalyst loadings would be ideal. This does not mean that the catalyst itself would become more reactive. The optimal support would be one which has a greater propensity for the proper geometry for the reaction mechanism, be it for the mechanism itself or, as was the issue previously, for the accessibility of the catalyst to the reactants. For one support to be considered better than another as a function of catalyst loading, the support would produce a superior reaction profile at an equivalent catalyst loading.

One of the interesting properties of this catalyst is that the reaction mechanism is bimetallic, i.e., two metal centers of the metal-salen complexes are involved in the reaction[96]. Additionally, it has been found that there is a quadratic dependence of the cobalt species concentration on the reaction rate[97]. A step from the mechanism showing the role played by the two cobalt centers is shown in Figure 49. This brings an extra degree of complexity to the design of systems which use this catalyst for immobilized homogeneous catalysis.

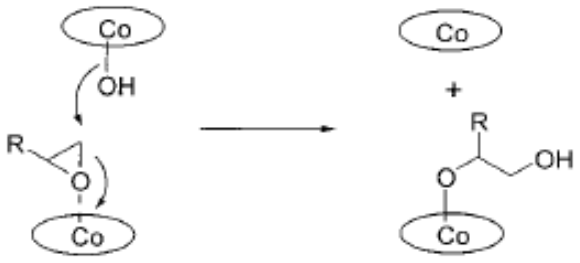


Figure 49: A step of the HKR mechanism that shows the role played by two cobalt centers of two separate Co(III)salen complexes[96].

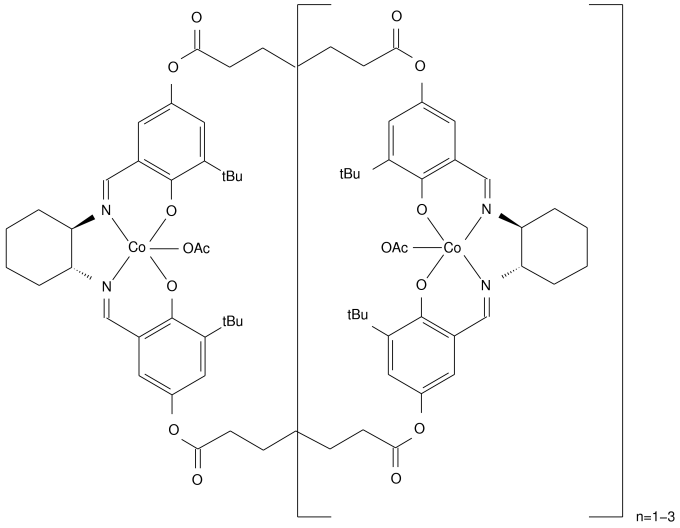
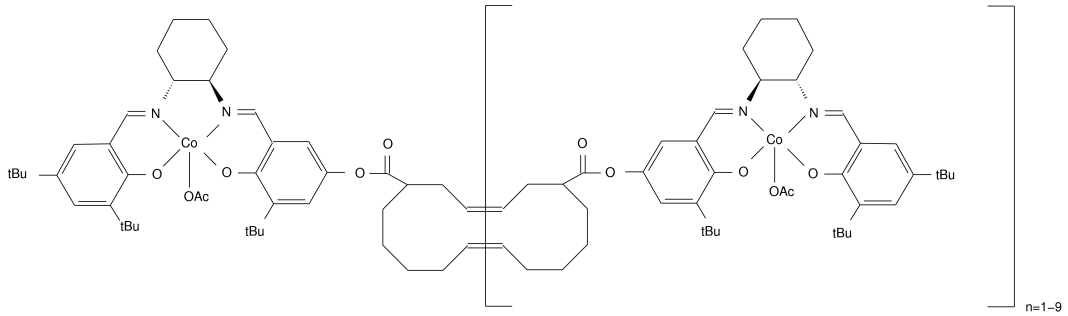


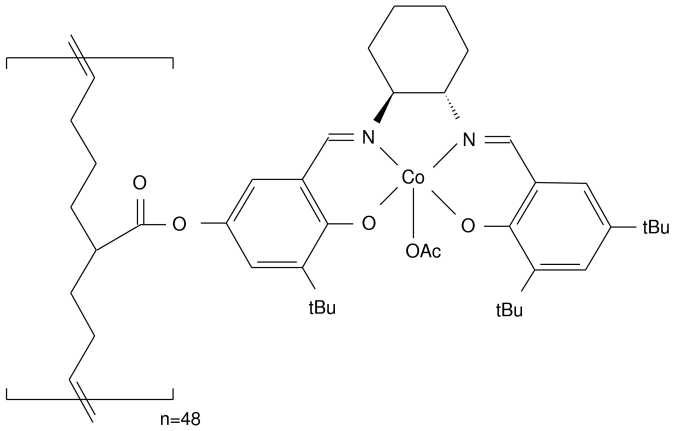
Figure 50: Oligomer used by Ready and Jacobsen to improve the reactivity of the Co(III)salen over the isolated complex.

Ultimately, the desire for any immobilized homogeneous catalyst is for both the ability to recycle the catalyst and the ability to use low catalyst loadings. Both of these properties are desirable because they decrease the catalyst cost. This particular family of catalysts has been shown to be reusable[95]. Unfortunately, the rational design of a system that requires low catalyst loading becomes much more difficult due to the bimetallic nature of the reaction mechanism.

One of the factors in determining the catalyst loading is the amount of catalyst that is required to perform the reaction in a reasonable amount of time. Originally, this catalyst was used as isolated salen complexes that required a large amount of the catalyst to be added to the reactant system. The reason for this is that with a bimetallic reaction mechanism,



(a)



(b)

Figure 51: Poly(cyclooctene)-supported Co(III)salen systems. Cyclooct-4-en-1-yl was functionalized with a Co(III)salen catalyst and polymerized into (a) oligomeric macrocycles and (b) a straight chain polymer.

isolated metal-salen complexes would randomly be in a position to react. As the catalyst loading increased, the probability for these random reaction events increases. Ready and Jacobsen improved upon the original system by constraining the metal-salen complex in a simple oligomer, shown in Figure 50[92, 98]. This allowed a reduction in the loading of the catalyst for the same *ee*, and the reaction time even improved over the isolated Co-salen complex for the same reaction.

This highlights the importance of specifically designing the overall catalyst system to accommodate the mechanistic properties of the Co-salen catalyst. Isolated Co-salen complexes are undesirable because the probability of a reaction is random and requires a high catalyst loading in order to perform the reaction in a reasonable time frame. However, as demonstrated by Zheng, Jones, and Weck, the choice of support cannot be a random choice

for this particular catalyst. The catalyst must be able to take advantage of a support's geometry in order to increase the probability of a bimetallic interaction on a reactant model. This was experimentally proven through the use of PCO as a support for a Co(III)salen complex. PCO was polymerized into two separate configurations. The first was a macrocycle with 2-10 repeat units (Figure 51(a)) and the second was a polymer chain with 48 repeat units (Figure 51(b))[91]. These will be referred to as cPCO for the macrocycles and lPCO for the linear polymer chain for the remainder of this work. One additional advantage that the cPCO system has is that it is more reusable than the monomer, though this is more due to counter-ion effects than geometry[99].

The results from these two supports showed that even though the repeat unit was the same, the reaction kinetics depended on the overall structure. The reactivity of lPCO was less than that of cPCO by 11% *ee* with equivalent catalyst loading and reaction time. This result suggests that the cPCO structure increases the probability of the Co-salen complexes to be in a geometrically superior position so that a reaction event was more likely. From a simulation standpoint, a reaction event is defined as a reaction that stems from the proper reaction mechanism occurring for only the species involved in given reaction. For this system, a reaction event involves two Co-salen complexes in the proper geometries with water and a terminal epoxide. A possible reaction event simply involves two Co-salen complexes in the proper geometries.

Ready and Jacobsen hypothesized that a specific geometry is required for a reaction to occur. The Co-salen complexes have to be in roughly parallel planes, and they require a "head-to-tail" orientation[98], where the head is in the direction of the nitrogen atoms bonded to the cobalt center, and the tail is in the direction of the oxygen atoms bonded to the cobalt center. For a possible reaction event, these are considered the proper geometries required for a reaction event to occur.

One final aspect of this catalyst system that will be ignored for this work is the counter-ion. Several counter-ions have been used for this system, including chlorine, iodine, an

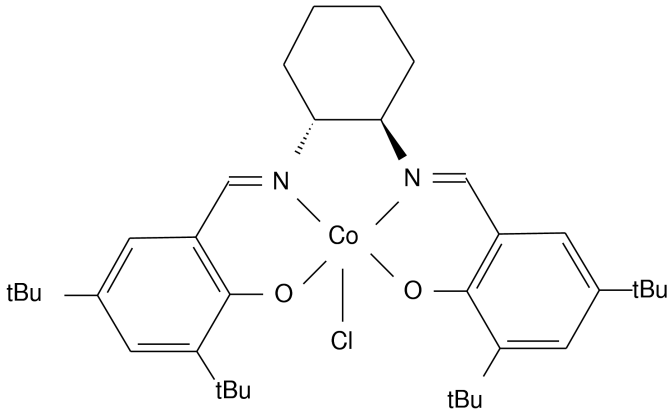


Figure 52: Co(III)salen complex with chlorine counter-ion.

acetoxy group, and a tosylate group. To compare directly to the work done by Zheng, Jones, and Weck, the acetoxy (OAc) counter-ion will be used. For the most part, the counter-ion should not affect the simulation results for the purpose of this study, however, they do have significant effects on the reactivity of the Co-salen complex[99].

6.2 Simulation Methodology

6.2.1 Force Field

6.2.1.1 Extended Force Field Development

From the previously described work with poly(*cis*-1,4-butadiene), MMFF94 was chosen for the force field for the PCO backbone. Since the PCO backbone is simply a random coil chain of seven single bonds with a lone double bond, this is a reasonable choice, as poly(*cis*-1,4-butadiene) has a similar backbone with four fewer single bonds. However, the issue now is that there is no atom type defined in MMFF94 to describe cobalt. Additionally, all of the bonds that are connected to the cobalt atom in the Co-salen are very specific to that geometry. Therefore, it is necessary to define the cobalt atom and all adjacent atom types for this specific geometry.

To do this, there were two sources used for comparison structures. The first is the results of X-ray diffraction on a Co-salen complex, and the second is *ab initio* quantum

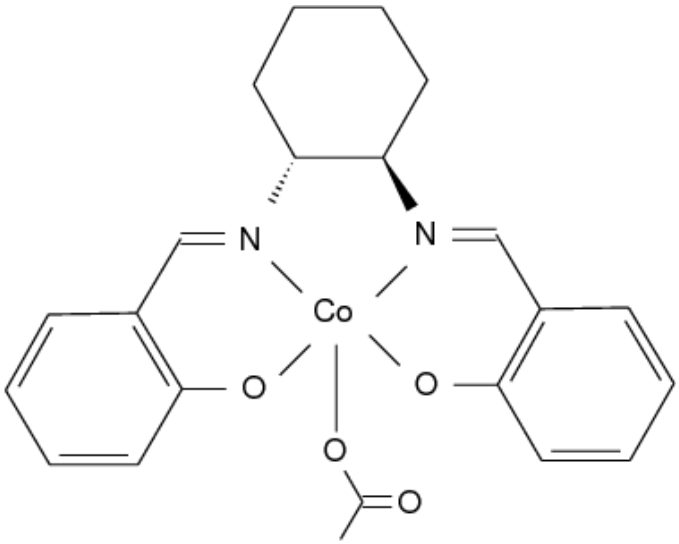


Figure 53: Reduced Co(III)salen complex with OAc counter-ion used in *ab initio* quantum calculations.

calculations on a reduced Co-salen complex. There have been a number of Co-salen X-ray diffraction structures that have been published in recent years[100, 101, 102]. Unfortunately, none of them are performed on a structure which uses the OAc counter-ion. Therefore, the structure reported by Cohen and coworkers on the Co(III)salen complex with a chlorine counter-ion was used as the comparison X-ray structure (Figure 52)[102]. Any comparison between this structure and the structure developed here will require the removal of the counter-ion.

The *ab initio* structure that was used as the basis for the force field parameterization was developed by Takatani and Sherrill, and the details for this calculation can be found in Appendix E. The structure used for this calculation used the OAc counter-ion, though it was a reduced Co-salen structure(Figure 53). One advantage to using an *ab initio* structure to extend the MMFF94 force field is that its parameterization basis was *ab initio* structures[56], and it has been found that this force field agrees very well with *ab initio* structures that were not used in its parameterization[103].

The parameterization of a force field is non-trivial, even in a case such as this, where

only one atom type needs to be parameterized. Therefore, another force field was employed to obtain all of the force constants, while the X-ray diffraction structure was used to determine the equilibrium bond lengths and angles. The force field parameters that were used were from the extensible systematic force field, which is abbreviated as the ESFF force field[104]. This force field has been previously used to model Mn-salen complexes, where the simulations used the GROMOS96 force field for the majority of the Mn-salen complex with ESFF parameters for manganese[105]. The advantage of using ESFF for the cobalt atom in the Co-salen is that ESFF was parameterized with the expectation that it might be used for organometallic molecules. Therefore, geometries that are common in organometallic systems, such as a cobalt atom bonded to four atoms in a plane, will have parameters available for that specific geometry. This would not be true in, for example, force fields parameterized for biological molecules, where such an orientation is uncommon.

MMFF94 requires terms for bond stretch, angle bend, torsion, stretch-bend, out-of-plane, van der Waals, and electrostatic interactions. Fortunately, it already has defined atom types for almost every atom and combination that would be involved in the Co-salen ligand except for those involving cobalt and a couple of uncommon atom type pairings. Therefore, only these parameters need to be determined. By using the ESFF force field as a starting point for cobalt parameterization, simplifications that were made in parameterizing ESFF can be applied to MMFF94. One of these simplifications is that torsion energies are zero for torsion angles that have central atoms with coordination numbers that are four or greater. Additionally, the torsion energies are zero that involve any metal atom that has bonding that involves its delocalized π orbital. Therefore, all torsion energies that involve cobalt for the Co-salen ligand can be approximated as zero, which means that all of the Fourier components of the MMFF94 torsion equation will be set to zero. Additionally, the ESFF force field does not have defined cross terms or out-of-plane energies for metal atoms, so these can also be set to zero. This means that the only parameterization required

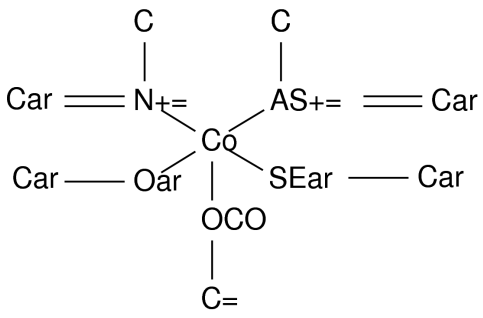


Figure 54: Reduced Co(III)salen complex with OAc counter-ion that shows the atom type that was used in the force field for the atoms directly involved in parameterization to extend the MMFF94 force field.

is for bond stretch, angle bend, van der Waals, and electrostatic interactions. Of these, the ESFF force field will be used to parameterize the MMFF94 form of all of these interaction energies except for electrostatics. The electrostatic interactions will be parameterized using *ab initio* quantum calculations.

To properly enforce the new force field parameterization, new atom types had to be used to describe the atoms involved in the parameterization. The names of these atom types can be seen in Figure 54. It is shown in this figure that each atom that is bonded to the cobalt center has its own distinct type. This was very important in order to properly identify differences in angle bend parameters. The linear O-Co-N angle is very different from the perpendicular angle. By giving each atom its own type, this distinction can be made. When the parameter values are reported, they will only be reported for the N+= and Oar atom types when these parameters are the same for the AS+= and SEar atom types. If the AS+= and SEar parameters are not explicitly listed, then the AS+= parameters are identical to those of N+=, and the SEar parameters are identical to the Oar parameters. The only other note from the following tables that show the parameters for the new atom types is that only parameters that are not already present in MMFF94 will be listed, and only the parameters that are changed (when applicable) from the standard MMFF94 value will be displayed.

In order to determine the bond stretch and angle bend force constants, the ESFF form

Table 15: Fit MMFF94 parameters for initial charges (q_0), van der Waals radius (R_i), and van der Waals well depth (ϵ) for Co-salen.

Atom	R_i (Å)	ϵ (kcal mol^{-1})	q_0
N+=	-	-	-0.100
CO	2.009	0.298	3.000
Oar	1.825	-	-0.250
OCO	-	-	-1.000

Table 16: Fit MMFF94 parameters for bond stretch and bond charge increment parameters for Co-salen. The units for k_x are kcal mol^{-1} Å^{-x}.

Atom i	Atom j	r_0 (Å)	k_2	k_3	k_4	w_{ik}
CO	N+=	1.885	155.726	-311.452	363.361	0.3125
CO	Oar	1.845	222.353	-444.706	518.824	0.3125
CO	OCO	1.877	279.851	-559.702	652.986	0.7500
N+=	Car	1.296	320.335	-640.670	747.448	0.3000
Car	Oar	1.306	528.606	-1057.212	1233.414	-0.3500
C	N+=	1.486	-	-	-	-

of the force field equation with the ESFF parameters was fit to the MMFF94 form. These equations and parameter descriptions can be found in Appendix A. Gnuplot (http://www.gnuplot.info) was used to determine the force constants that fit the MMFF94 equations to the ESFF values by using its implementation of the nonlinear least-squares Marquardt-Levenberg algorithm. The parameters that were fit were the bond stretch and angle bend force constants. This method was also used to find the van der Waals radius and potential energy well depth (Table 15).

Table 17: Fit MMFF94 parameters for linear angle bend parameters for Co-salen.

Atom i	Atom j	Atom k	θ_0 (°)	k_θ (kcal mol^{-1})
Oar	CO	AS+=	180	72.2528
SEar	CO	N+=	180	72.2528

Table 18: Fit MMFF94 parameters for normal angle bend parameters for Co-salen. The units for $k_{\theta x}$ are kcal mol^{-1} rad^{-x}.

Atom i	Atom j	Atom k	θ_0 (°)	$k_{\theta 2}$	$k_{\theta 3}$
Oar	CO	OCO	90.000	32.3020	-12.9208
N+=	CO	OCO	90.000	33.0406	-13.2162
AS+=	CO	N+=	90.000	34.8000	-13.9200
N+=	CO	Oar	90.000	33.7495	-13.4998
Oar	CO	SEar	90.000	32.8852	-13.1541
HC	Car	Car	120.571	40.5167	-16.2500
C	C	N+=	106.424	84.4157	-33.8567
HC	C	N+=	106.973	62.8980	-25.2265
CO	Oar	Car	128.965	61.5483	-24.6193
CO	N+=	C	111.553	47.2114	-18.8846
CO	AS+=	C	123.395	47.2114	-18.8846
CO	N+=	Car	125.243	46.9740	-18.7896
Oar	Car	Car	121.185	52.1477	-20.8591
C	N+=	Car	122.374	50.4427	-20.1771
N+=	Car	Car	125.505	57.2041	-22.8816
N+=	Car	HC	117.202	48.3113	-19.3245
CO	OCO	C1=	119.679	48.9986	-19.5994

As previously mentioned, the X-ray diffraction structure was used to determine the values for the equilibrium bond lengths and angles. The only exception to this is that any bond lengths or angles involving the OAc counter-ion had to be found using the optimized *ab initio* geometry (Appendix E). Once these were chosen, all of the bonded interactions that involved cobalt were fully parameterized (Tables 16, 17, and 18), except for the electrostatic interactions. These interactions were parameterized using fundamental knowledge about the system to assign initial full charges combined with the proper value of the bond charge increment method used by MMFF94 to match the partial charges provided by the *ab initio* calculations (Appendix E). The value of the initial charges can be found in Table 15, while the bond charge increment values can be found in Table 16.

Once all of these parameters were added to the original MMFF94 force field, the Co-salen with a OAc counter-ion is fully parameterized. This extended MMFF94 force field is used for the rest of this work in all of the simulations involving Co-salen ligands.

6.2.1.2 Root Mean Square Deviation

The accuracy of the force field was determined from the root mean square deviation (RMSD) of the simulated structure compared to a standard. This value is calculated by taking the square root of the average of the squares of the difference in each atomic coordinate between two superimposed structures. One structure is the "ideal" structure, which is the structure that is determined to be accurate. The other is the structure that, for this analysis, is the one that depends on the extended MMFF94 force field. In this case, the ideal structure was chosen to be the experimental structure obtained through X-ray diffraction, but the RMSD was also compared to the structure obtained from the *ab initio* calculations. A number of isolated Co-salen molecules were tested to truly determine the accuracy of this extended force field.

Because the X-ray diffraction structure used Cl as the counter-ion, it was necessary to remove it before finding the RMSD. Additionally, the t-butyl groups add an additional

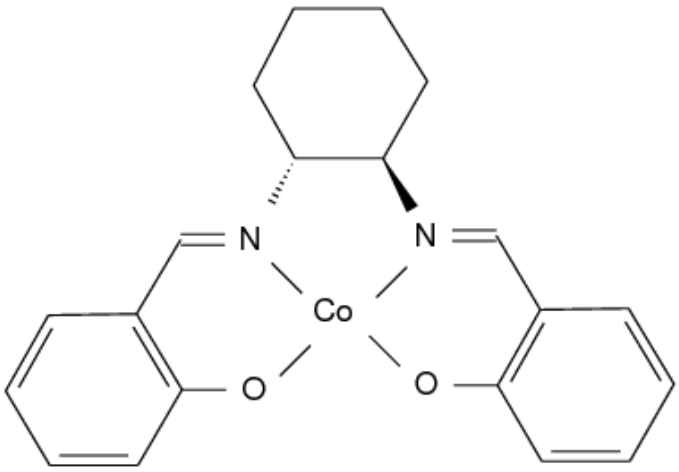


Figure 55: Reduced Co(III)salen complex that was used to compare both the X-ray diffraction structure and the *ab initio* structure.

complication, as there is no reason for the hydrogens on the t-butyl groups to align themselves in any given order compared to the ideal structure. Therefore, all four t-butyl groups were removed before the RMSD was determined. This also helps to compare the modeled structures to the *ab initio* structure, as this structure was not optimized with t-butyl groups. This reduced structure that is used for comparison can be seen in Figure 55. All minimizations were performed with counter-ion and t-butyl groups, and they were only removed to determine the RMSD.

For comparison purposes, the *ab initio* structure was compared to the X-ray diffraction pattern, and it had an RMSD of 0.27 Å. The first simulated compound was the X-ray structure as the starting structure. The Cl counter-ion was replaced with the OAc counter-ion, and this structure was minimized with the extended force field. This minimized structure had a RMSD of 0.42 Å. Considering that each structure is comparing 45 atoms, the fact that the simulated structure had a RMSD that was relatively close to that of the *ab initio* structure means that this force field adequately describes this molecule.

However, this still is not the best comparison for this molecule, as the hydrogens were not optimized for this systems, and the hydrogens may not always have a specific global minimum (similar to the t-butyl groups). Therefore, these structures were again compared,

but this time all hydrogens were removed after minimization. The *ab initio* structure decreased its RMSD to 0.20 Å, and the structure with the Cl counter-ion that was minimized with the extended force field also had an RMSD of 0.20 Å. The minimized structure with the OAc counter-ion achieved a RMSD of 0.27 Å once the hydrogens were removed.

These RMSD values show a very high level of agreement with the X-ray diffraction structure, but this does not give any indication of performance compared to the *ab initio* structure. To determine this RMSD, the ideal structure is now the *ab initio* structure, and this is compared to the same minimized structure three times, each with one less degree of complexity. The t-butyl groups were again removed after minimization due to the lack of t-butyl groups in the *ab initio* structure. This structure had a RMSD of 0.83 Å. Removal of the hydrogens decreased the RMSD to 0.52 Å, and the removal of the rest of the OAc counter-ion decreased the RMSD even further to 0.26 Å. This trend in RMSD should be expected. The reason the first value is so high is that in a molecular mechanics simulation, all of the hydrogens that are of the same type will have the same behavior. Because a quantum calculation allows each bond length to be what it is based on the atomic orbitals, it is not going to be the case that an *ab initio* structure would have hydrogen bonds that are all exactly the same length or orientation. Therefore, it is likely that all of the hydrogens will contribute non-trivial deviation to the RMSD value. Additionally, if the torsion angle of the terminal C-C bond on the OAc group is rotated 90°, this will cause the deviation of the hydrogens on the terminal methyl group to be higher because now, not only do they have different lengths, there is a complete shift in position. This is a very real possibility, as there is nothing to hinder that particular torsion angle, so if the torsion angle is changed during minimization, the RMSD will increase. The RMSD for the case without hydrogens should also be larger than that without the OAc group. This is because, similar to the terminal methyl group, there is no energetically preferred orientation for the doubly bonded oxygen atom. If a rotation occurs in the O-C torsion angle, this can substantially change the coordinate of the doubly bonded oxygen.

Even with the addition of all of the hydrogens and the counter-ion, the fact that the RMSD values are all less than 1 Å means that this structure will adequately model the Co-salen for minimization. However, MD will also be performed on this molecule for the analysis of the entire system, so it is necessary to check the effect of MD on the structure with the extended force field. MD was performed on the molecule with no t-butyl groups with the OAc counter-ion and explicit hydrogens, but the hydrogens were constrained using the LINCS algorithm. A 100 ps simulation was performed with the NH-SV algorithm at 300 K with a sampling period of 0.5 ps. The 200 structures which were recorded were compared to the *ab initio* structure to determine the RMSD. Nothing was removed from the MD structures. The average RMSD for the simulation was 0.93 Å, which is very good agreement with the *ab initio* structure, considering that there was added molecular strain that comes from the addition of kinetic energy. It is worth noting that while this shows good structural agreement, this parameterization may to stiff for this molecule, as 0.93 Å is a very low RMSD value for a structure with kinetic energy at 300 K.

6.2.2 Simulated Systems

Three different systems were studied. The first is the Jacobsen oligomer (Figure 50), the second is the linear PCO chain (Figure 51(b)), and the third is the PCO macrocycle (Figure 51(a)). To compare with the experimental data, the Jacobsen oligomer was simulated as a dimer, trimer, and tetramer. The PCO macrocyles were simulated in lengths ranging from 2-10 repeat units, and the linear PCO polymer was simulated with 48 repeat units. While the Grubbs catalyst, which was used in the polymerization of the PCO supports[91], produces a predominantly trans isomer[89], both the cis and trans isomers were simulated.

6.2.3 Molecular Dynamics

MD was used to sample the conformation space at 300 K. The NH-SV algorithm was the dynamics algorithm used[40], and the thermal coupling parameter was set to 0.5 kcal ps $Å^{-1}$. The time step was 0.001 ps, and the hydrogens were constrained using the LINCS

constraint algorithm[82]. The code used to run the molecular dynamics differed from the default MOE code and is included in Appendix C. The simulation was sampled every 0.1 ps. The simulations were run for 500 ps, and the final 300 ps were used for the data analysis. No pair interaction cutoff was employed for these simulations.

An implicit solvent model was used to approximate solvent interactions on the system. The solvent model that was used for all supports was the generalized Born implicit solvent model[106], and the dielectric constant was set to that of ethylene oxide, which is 14[62]. Ethylene oxide was used because it is the smallest terminal epoxide. In order to determine the effect of the solvent model, the Jacobsen oligomer system was repeated a total of four times with different solvent model parameters. The generalized Born model was repeated with the dielectric constant of water and with the average dielectric constant of water and epoxide. The final set of simulations used no solvent model.

Because all of these simulations were performed on oligomers with very flexible backbones, only one starting conformation was used. This conformation is irrelevant because after equilibration, the lPCO chains should be in a random coil conformation due to the backbone flexibility. Because the cyclic compounds are small, they will be sampling the proper conformation space because there should be no barrier to the realistic conformation as there is for more stiff molecules. The simulations were repeated three times for each compound that was used. This was done in an attempt to sample the conformations at different starting conditions (randomly generated initial velocities) in order to sample the different possible conformation spaces for each starting conformation. Since it is unlikely that there are different final conformations, the only thing that the different starting velocities can achieve is to produce simulations with alternate possible reaction event profiles. By adding the results of the three simulations, a reaction event profile can be found that should roughly encompass the majority of the conformation space corresponding to possible reaction events.

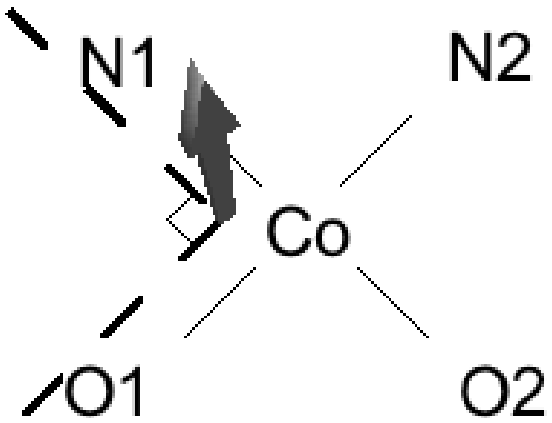


Figure 56: Rendering of the determination of the normal vector for the N1-Co-O1 atom point set. The dashed line is the angle defined by the Co-N1 and Co-O1 bonds.

6.2.4 Analysis of Reaction Events

6.2.4.1 Geometric Analysis

Once the simulations had generated a large number of MD conformations, it became necessary to determine reaction events. These reaction events should follow the general geometric constraints, i.e., the cobalt planes should be roughly aligned, and there should be a head-tail alignment. Additionally, the analysis of these events should be systematic and should only include cobalt centers which are within the proper distance for a possible reaction. Unfortunately, this distance has not been determined experimentally or theoretically, so a number of distances were investigated to act as a comparison. For this work, 7, 8, 9, and 10 Å separation cutoff distances were used for the Co-Co distance. If two cobalt atoms are within this distance, the geometry is checked to determine a reactive event.

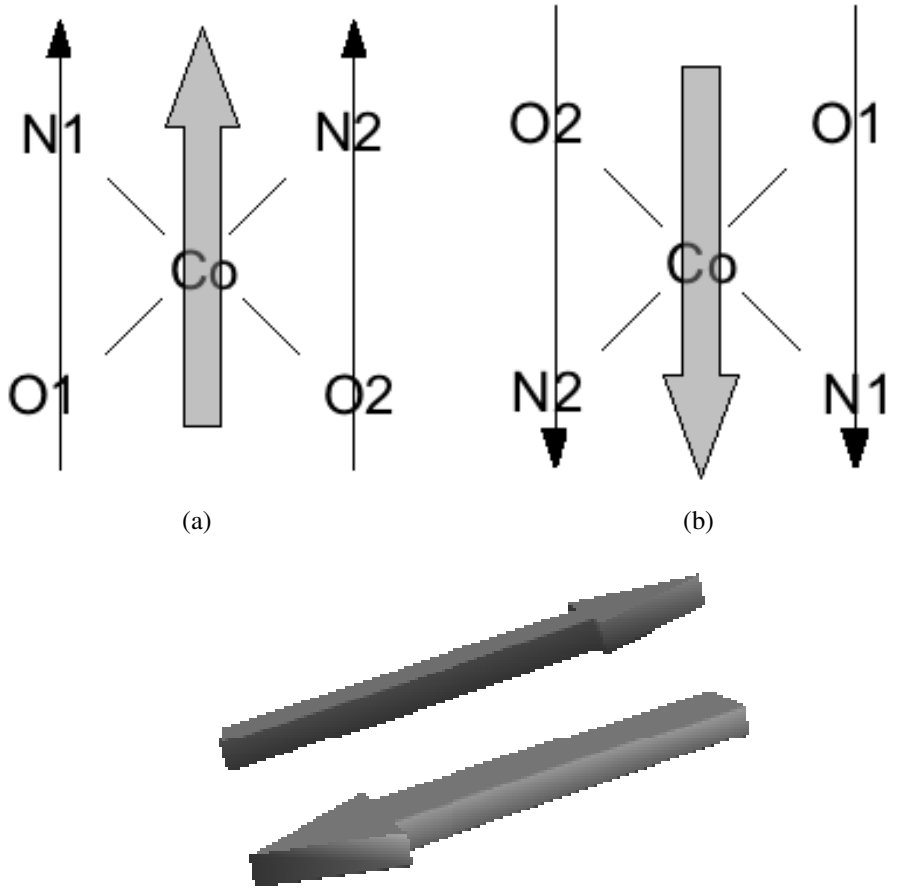
Because of the molecular geometry of the cobalt center, the only feasible arrangement for two cobalt centers to catalyze a reactant is for the two Co-salen structures to be in an arrangement with aligned planes, where the plane is defined by the cobalt and the bonded nitrogen and oxygen atoms. However, particularly with MD, this atom set may not be in a perfect plane. Therefore, the plane must be defined first.

Geometrically, a plane is defined by the normal vector, which is a vector that is perpendicular to the plane. In this case, where the plane is not perfectly defined by any three of the five atoms involved, the plane can be defined by determining the average of the normals of each angle set about the cobalt center. This is shown in Figure 56, where the N1-Co-O1 normal is shown. By determining the cross product of the Co-N1 and Co-O1 vectors, the N1-Co-O1 normal vector is obtained. The normal vector is found for each of the remaining perpendicular angles about the cobalt center (O1-Co-O2, O2-Co-N2, and N2-Co-N1). To calculate the average cobalt plane normal, all four normals are averaged, and this average vector is normalized to obtain the unit cobalt plane normal vector.

To determine the plane alignment, the application of the dot product is necessary. If two planes are parallel to each other, their normals are also parallel. For parallel vectors, or vectors that have angles of 0° or 180°, the dot product takes a value of either -1 or 1 and approaches 0 as the vectors become perpendicular. For this particular application of determining whether two approximate planes are parallel, some latitude should be allowed. This will be shown later through a sensitivity analysis.

The other constraint on geometry is that it was theorized by Ready and Jacobsen that the two Co-salen groups must be in a head-to-tail arrangement[98]. Figure 57(a) shows a cobalt center in a "head" orientation, Figure 57(b) shows a cobalt center in a "tail" orientation, and Figure 57(c) is the three dimensional rendering of the approximate directions in a head-to-tail geometry. In order to determine the direction vector, a similar method was used as the average plane determination. In this case, the O1-N1 and O2-N2 vectors are individually calculated, and these are averaged and normalized to produced an average direction vector that points in the direction of the nitrogen atoms from the cobalt center.

To determine the head-to-tail geometry, the dot product is used once again. Here, the only acceptable values are ones in which the directions are roughly parallel and in opposite directions, or a dot product of -1. This most likely does not need to be as stringent as the parallel plane geometries, so a higher degree of latitude should be allowed for this



(a) (b)

(c)

Figure 57: Geometric approximations for direction vectors on Co-salen for the (a) head and (b) tail orientations. The thin black arrows represent the individual direction vectors, and the large gray arrows represent the average direction vector. Also shown is a (c) three-dimensional rendering of the head-to-tail direction vectors.

orientation. This will also be discussed later in Section 6.3.1.

6.2.4.2 Limitations of the Model

This model and subsequent analysis has one major limitation, which is that each potential reaction event is not necessarily going to result in a reaction in a real experiment. These models should be adequate for the initial system which contains mostly the racemic epoxide and water, where there is enough of the reactant epoxide enantiomer to not require diffusion of the remaining reactant through the bulk to the supported catalyst. This is not as much of a concern, since these reactions are typically stirred, however, at some point there will be too small of an amount of the reactant epoxide enantiomer to account for all possible reaction events. Therefore, this analysis will only provide insight into the portion of the reaction kinetics that is linear with time. However, relative differences between structures will be meaningful, since the purpose of this analysis is to compare the likelihood of a reaction-favorable geometry.

One additional important limitation of this model is the counter-ion behavior. In this case, the counter-ion is explicitly bonded to the cobalt center. In reality, this counter-ion is not explicitly bonded. For the purposes of this model, an explicit bond will adequately model reaction events, as the assumption is that a possible reaction can be modeled as a Co-salen with counter-ion side will approach another on the side without the counter-ion. However, it should be noted that this explicit bond removes any effect that the counter-ion has on the system. One alternative for this would be to replace the counter-ion with a hydroxy group to more accurately capture the actual reaction mechanism.

This model could be extended to account for all possible kinetic regions as a function of time if the solvent and reactant were explicitly modeled with the concentration that corresponded with the current reactant conditions, i.e., the amount of reactant that is left in the system. The advantage of this is that each reactant event would be more accurate because it would only be counted if a reactant was present between two cobalt centers that

are separated by a specific amount and in a specific geometry. However, this would involve a number of additional atoms as well as a different simulation for each amount of leftover reactant, and it would have to be run far longer due to ensuring a good description of the reactant motion in the solution. Simplifying the system by removing this constraint is imperative for reasonable simulation times and repeatability, and the results without an explicit solution should be adequate enough to determine the trends which should be expected from the experiment.

6.2.4.3 Normalization of Possible Reaction Events

In order to adequately compare the relative reaction events with experiment, the reaction events must be normalized. This normalization involves individual catalyst site efficiency and overall supported catalyst efficiency. Both of these come from one property of these systems – catalyst cost. For any of immobilized homogeneous catalyst, unless the catalyst is an acid based or equivalently simple or cheap catalysts, the catalyst itself is often the major process cost. This is especially true for precious metal catalysts, but in the case of a Co-salen, where cobalt is not a precious metal, the ligand is sufficiently complex that the manufacturing process of such a ligand is the major cost. This is true for a number of homogeneous catalysts, so there is a motivation to use as few catalysts as possible to perform a given reaction. This manifests itself experimentally in two ways. It drives the need to design the most efficient catalyst, i.e., the one that performs the most number of reactions per supported catalyst. In addition, it is desirable to use the lowest possible catalyst loading in the reaction volume that will achieve the desirable reaction yield. These are not mutually exclusive, in that a more efficient catalyst should require less loading, but this brings another complication that will be explained shortly.

There is an advantage to immobilizing a Co-salen to a support due to the nature of the bimetallic reaction mechanism. By immobilizing these catalysts, random interactions of isolated Co-salen catalysts is no longer the bottleneck for performing this reaction. The

reaction is now isolated to a single supported catalyst, but because the Co-salen catalysts are constrained to a support, there is now a higher probability of catalyst interactions. Not only does this drastically improve the number of possible reaction events, it also allows the catalyst to be simulated as an isolated supported system, particularly at the highly dilute experimental conditions. Additionally, it allows for the individual catalyst efficiency to be based on a single support. Individual catalyst site efficiency is simply the number of possible reaction events divided by the number of catalysts attached to a given support. This measure will allow a comparison of supports based on intermolecular interactions. Using the Co-salen supported on PCO as an example, if a trimer has the same number of possible reaction events as a dimer, the dimer is more efficient, as it has a higher individual catalyst efficiency. Each Co-salen would be involved in more reactions for the dimer than the trimer.

Overall catalyst efficiency takes into account the fact that these systems should be compared at a constant catalyst loading. By supporting the catalyst, the local catalyst concentration is increased, but in comparing supports with a varying number of tethered catalysts at a constant loading, the global catalyst concentration is actually decreased. By tethering the catalyst to a support, the entire catalyst-support unit becomes the active species under constant catalyst loading. As an example, again using the dimer and trimer, a trimer may have a greater individual catalyst efficiency. However, under constant loading, the global concentration of catalysts would be two-thirds of the global concentration for the dimer. The trimer would have to have 1.5 times the number of possible reaction events to catalyze the same number of reactants as the dimer at equivalent catalyst loadings. Since there is no good way to compare the possible reaction events for a supported Co-salen with the random nature of the possible reaction events for the monomer, the dimer case will be used for comparison. By normalizing all of the supported catalysts to a dimer case under constant loading, the different sizes of the supports can be directly compared.

By integrating both the individual catalyst efficiency and the overall catalyst efficiency

into a "reaction score" all of these systems can be directly compared, and an optimum can be determined. Therefore, the reaction score will be defined by

$$\text{Reaction Score} = \frac{\text{\# of Possible Rxn Events}}{\text{\# of Catalysts per Support}} \frac{\text{\# of Catalysts per Reference Support}}{\text{\# of Catalysts per Support}}, \quad (21)$$

where the reference support is the dimer, or two catalysts per support. The reaction score is simply the individual catalyst efficiency (first term) multiplied by the overall catalyst efficiency (second term). If either term is increased, the overall reaction score is increased.

6.3 Results and Discussion

6.3.1 Geometric Sensitivity Analysis

There are three main parameters important for this analysis. They are the required catalyst separation distance, the allowed plane angle deviation, and the allowed head-to-tail angle deviation. These were all varied in an attempt to find the best values to use. For the angle deviations, it was decided that there should be more allowed deviation in head-to-tail angle than plane-to-plane angle, deviation being the allowed variation from perfectly parallel geometries. This is due to the fact that the geometry of the ligand is such that the reaction is most likely to occur perpendicular to the ligand plane, so the allowed deviation in plane-to-plane angle should be equal to or lower than the head-to-tail angle deviation. For the plane-to-plane angle deviation, it was found that allowed deviations less than 30° showed little or no possible reaction events. Head-to-tail angle deviations lower than 45° also had this same issue. Additionally, the number of possible reaction events greatly increased for head-to-tail deviations greater than 45°. The choice became narrowed to head-to-tail angle deviations of 45° with plane-to-plane angle deviations of either 30° or 45°.

Table 19 shows the results of the 30° allowed plane-to-plane angle deviation case, and Table 20 shows the result for the 45° case. The trends for both cases are the same, so it was decided that the plane-to-plane angle deviation should be constrained to account for the hypothesized geometry. Therefore, the 30° case was used for the rest of the analysis.

Table 19: Number of possible reaction events for allowed plane-to-plane angle deviations of 30°.

Oligomer	Max. Cobalt Separation			
	12 Å	10 Å	9 Å	8 Å
Dimer	0	0	0	0
Trimer	35	16	9	1
Tetramer	826	513	437	385

Table 20: Number of possible reaction events for allowed plane-to-plane angle deviations of 45°.

Oligomer	Max. Cobalt Separation			
	12 Å	10 Å	9 Å	8 Å
Dimer	0	0	0	0
Trimer	227	116	76	28
Tetramer	1785	1244	1072	848

It should be noted that regardless of what deviation was used, the dimer never had any noticeable reaction events due to the fact that for these simulations, it preferred a head-to-head conformation. Additionally, these results were used to pick the maximum allowable cobalt separation distance. Due to the reaction mechanism, the ideal reaction event for the simulated systems would be one in which two counter-ions are not between the cobalt centers. A maximum separation distance of 10 Å was chosen because it decreases the possibility of counting a reaction event where two counter-ions are on the interior of two cobalt centers.

Table 21 shows the reaction scores for the different sizes of the Jacobsen oligomer.

Table 21: Reaction scores for the Jacobsen oligomer.

Oligomer	Reaction Score
Dimer	0
Trimer	4
Tetramer	64

While only these were modeled because these are the oligomers that Ready and Jacobsen actually made, it is believed that the pentamer would see a decrease in reaction score compared to the tetramer. One major problem with these results is that the trend reported by Ready and Jacobsen was not reproduced. They separated the three oligomer sizes that were produced and performed the epoxide reaction at equivalent catalyst loadings. They found that the optimal was the trimer, followed by the dimer, with the tetramer being the least optimal[98]. However, this is most likely due to the approximate nature of the extended force field rather than a problem with the method. Because the Co-salen is part of the backbone for their oligomer, rather than tethered to another support, the flexibility of the salen ligand must be modeled to a high degree of accuracy to correctly capture the possible reaction events. The extended MMFF94 force field had a relatively low RMSD for MD simulations, which indicates that the simulated Co-salen may be more stiff than the actual structure. This is not surprising, especially since the force constants for the cobalt center were obtained from ESFF. ESFF was entirely parameterized based on crystal structures, and this parameterization will cause the force constants to produce structures that are more rigid. Additionally, the reaction was performed using a different counter-ion, which could also have an effect on the number of possible reaction events, as this counter-ion size will have a large effect on the separation distances and geometries between catalyst sizes.

6.3.2 Solvent Model Sensitivity Analysis

Table 22 shows the reaction scores for the different solvent models used. The values for the dimer are not shown, as they are all zero. The number following "Born" is the value used for the solvent dielectric constant. These results highlight the fact that the solvent model used does have an effect. One interesting aspect of this is that the ratios of reaction score decreases as the solvent dielectric constant increases, while it is virtually unchanged for the lowest dielectric constant used and the case with no solvent, which means that low dielectric constants essentially mimic Coulomb's Law. This result serves to show that these

Table 22: Reaction scores for the Jacobsen oligomer as a function of solvent model used.

Oligomer	Solvent Model	Reaction Score
Trimer	Born 14	4
Trimer	Born 47	29
Trimer	Born 80	71
Trimer	None	15
Tetramer	Born 14	64
Tetramer	Born 47	261
Tetramer	Born 80	221
Tetramer	None	267

results are consistent with what is expected, as Coulomb's Law is the limit as the dielectric constant goes to one. While the trend for which oligomer is considered optimal does not change, the reaction score is greater for higher dielectric constants used. This is due the screening of the electrostatic interactions. This means that as two catalysts with positive charges approach each other, the electrostatic repulsion that is present is decreased. The effect of this is an increased amount of time at a closer range. While this system has solvent effect, the relative trends should hold true, so long as the simulations are compared with those that were performed with the same parameters.

6.3.3 Comparison of Co-salen PCO Supports

Table 23 shows the reaction scores for the PCO supports. The dimer once again shows no possible reaction events, but this is plausible for this particular system as this would cause increased strain on the support from van der Waals interactions. It would appear that the trimer and tetramer are the optimal support sizes for both the cis and trans isomers, with the pentamer also a good support for the cis isomer. The results for the lPCO support show that the 48 repeat unit polymer, which was synthesized and showed poor experimental reactivity[91], had the lowest reaction score of those simulated. Comparing these results to those of the Jacobsen oligomer shows that the trans tetramer macrocycles have roughly double the reaction score of the top-performing Jacobsen oligomer. The cis trimer cPCO

Table 23: Reaction scores for the PCO supports grouped by PCO isomerism.

PCO Type	# of Repeat Units	cis Reaction Score	trans Reaction Score
cPCO	2	0	0
cPCO	3	620	38
cPCO	4	155	124
cPCO	5	140	24
cPCO	6	58	20
cPCO	7	72	25
cPCO	8	28	25
cPCO	9	61	22
cPCO	10	26	10
lPCO	48	6	6

Table 24: Individual catalyst efficiency for the PCO supports grouped by PCO isomerism.

PCO Type	# of Repeat Units	cis Efficiency	trans Efficiency
cPCO	2	0	0
cPCO	3	929	57
cPCO	4	310	249
cPCO	5	350	60
cPCO	6	173	58
cPCO	7	251	88
cPCO	8	112	102
cPCO	9	275	100
cPCO	10	132	50
lPCO	48	150	146

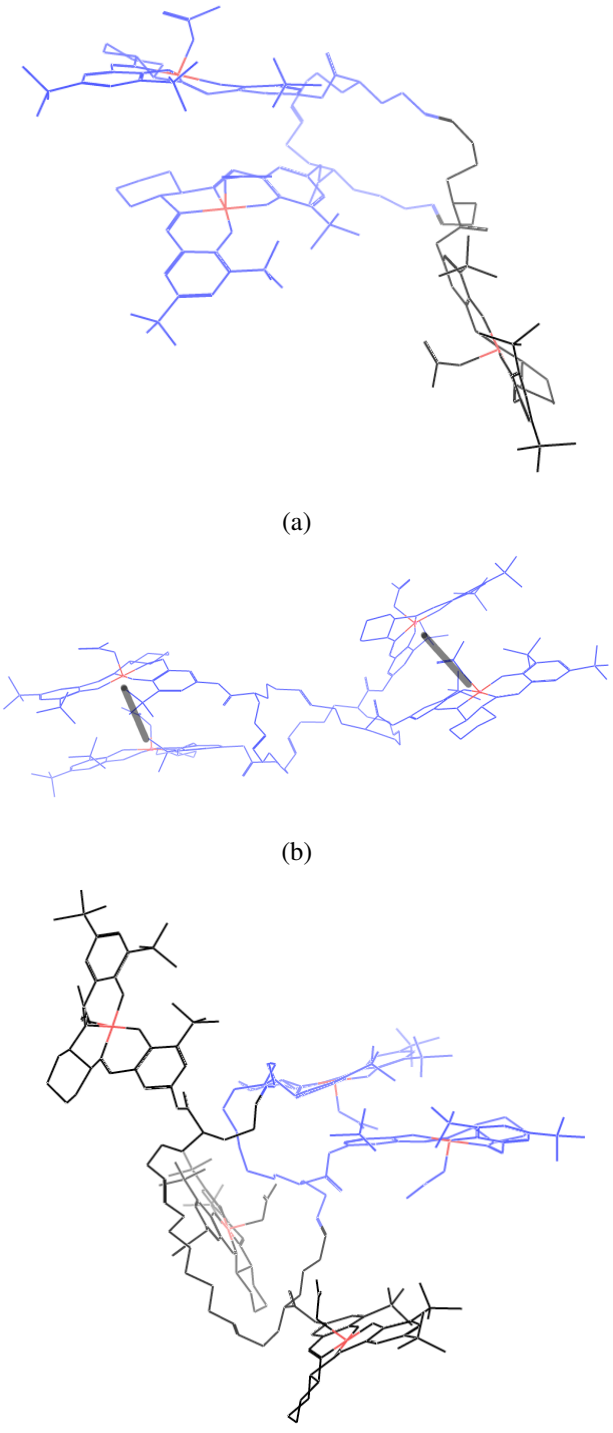
support has roughly ten times the reaction score of the Jacobsen oligomer. This likely means that the macrocyclic PCO support has a design which allows the backbone to work more synergistically with this particular reaction compared to the Jacobsen oligomer.

Table 24 shows the individual catalyst efficiency. These results show the importance of factoring in the effect of performing these experiments at equivalent loading. If only the individual catalyst efficiency were taken into account, it would seem that the actual catalyst sites are equally efficient for several of the support sizes. For example, the 48 repeat unit lPCO has a higher individual catalyst efficiency than all of the other trans supports except

for the tetramer. However, once the loading is taken into account, this particular support has the worst reaction score, and this is confirmed through experimental work[91]. This confirms that the lPCO is worse on a per loading basis than a mixture of low MW cPCO supports.

In order to fully identify the reason for these results, the actual simulations must be examined to determine the structural cause. It would be especially beneficial to determine why the *cis*-cPCO trimer performs better than the tetramer, and why the *trans*-cPCO tetramer performs better than the pentamer and trimer. Representative structures for *trans*-cPCO can be found in Figure 58, and Figure 59 shows structures for *cis*-cPCO. While these do not show all of the generated conformations, these can show why particular cyclic oligomers produce the optimal reaction score.

For the trans structures, only two catalysts interact for the trimer and pentamer, while two pairs are formed for the tetramer. This would definitely cause the tetramer to have a higher reaction score. For the cis structures, only one pair is formed with the tetramer. However, the trimer has the most interesting behavior. For all of the other structures (both trans and cis), the catalyst pairs that are identified are constant throughout the entire simulation. The pairs which are formed are unassociated and become associated again over the course of the simulations. This is caused by the geometry of the cPCO backbone, which seems to move the two catalysts relative to each other as the torsion backbone angles are oscillate between their extrema. However, this could be also be due to poor sampling of the conformation space. The cis trimer actually has one catalyst which interacts with both adjacent catalysts. This motion is more efficient for this type of reaction mechanism, so it results in a substantially higher reaction score. The reason that this motion works better than single pairs is that a single pair will spend a good portion of time outside of the minimum separation distance. In the case of the cis trimer, the catalyst site that moves spends very little time outside of the minimum separation distance. As it leaves its pairing from one site, it only spends a short time before it forms a pair with the adjacent site.

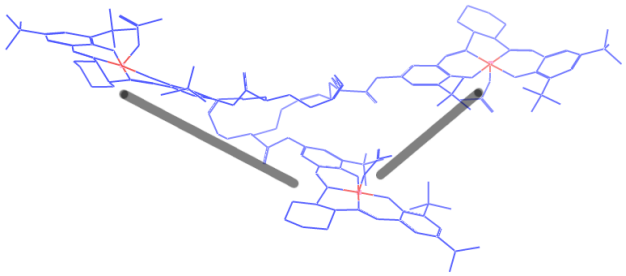


(a)

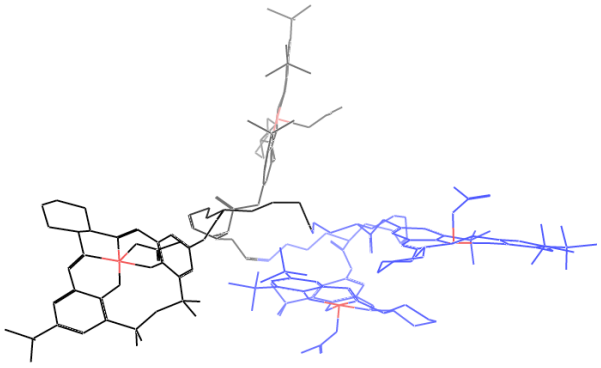
(b)

(c)

Figure 58: Snapshots of representative structures for *trans*-cPCO with (a) 3, (b) 4, and (c) 5 repeat units. The black repeat units are not involved in a reaction event, whereas the blue units are involved in a reaction event. For the case with more than one pair, gray lines are used to indicate the pairing.



(a)



(b)

Figure 59: Snapshots of representative structures for *cis*-cPCO with (a) 3 and (b) 4 repeat units. The black repeat units are not involved in a reaction event, whereas the blue units are involved in a reaction event. For the case with more than one pair, gray lines are used to indicate the pairing.

It should be noted here that these results are not definitive. The best that can be expected of this type of analysis is that a rough trend is found which should be experimentally verified. For example, experiments should be performed on 3-5 unit cis and trans cPCO supports to determine the optimal support, but all of the larger repeat unit structures can be ignored. This type of analysis would be most useful for reducing the total number of experiments needed to perform. In this case, Zheng and coworkers found that cPCO oligomers in the range 2-10 (predominantly trimer) are more efficient than a 48 unit lPCO support[91]. This is consistent with the simulation results, but more experiments are required to determine the exact optimal oligomer.

6.4 Summary

Supports for the Co-salen catalyst were simulated and compared to find the optimal support for this particular catalyst. The MMFF94 force field was extended to include cobalt parameters for this particular force field. These parameters were developed by combining force constants from the ESFF force field, equilibrium bond lengths and angles from X-ray diffraction, and charges from *ab initio* quantum calculations. This force field produced energy minimized Co-salen structures with very good agreement with both the X-ray diffraction and *ab initio* geometries. The RMSD was below 1 Å for both cases.

The use of a reaction score was employed as a way to compare the different support scenarios. This reaction score combines the individual catalyst efficiency with the global reaction probability based on the number of available supports in the system. The Jacobsen oligomer reaction score turned out to have a trend that differed from the experimental data by only one repeat unit. This is most likely due to the fact that the dynamics of the Co-salen complex is much more essential to the overall dynamics of the support, and the force field that was used was parameterized using crystal structures. This would serve to make the structures more rigid, which would, in turn, change the dynamics of the system. An examination of the implicit solvent model showed that the number of possible reaction

events increased with an increase in the dielectric constant of the solvent. This is most likely due to the increased implicit electrostatic Additionally, it was found that the optimal support is in the 3-5 repeat unit range for the macrocyclic PCO supports, and they all performed substantially better than the linear PCO supports. Of the macrocyclic PCO supports, the ones that were optimal for both the cis and trans isomers were those that performed more than one cobalt pairing, though the cis performed better overall. This was true whether the pairs were constant throughout the simulation or whether the pairs changed.

CHAPTER VII

CONCLUSIONS AND FUTURE WORK

7.1 Conclusions

7.1.1 Cyclopentane Simulations

Based on the results of numerous simulations, it was found that the MD thermostat algorithm and not just the force field had a significant effect on the pseudorotational dynamics of cyclopentane. This suggests that, in order to produce an accurate simulation of polymers with such cyclic groups, attention must be paid to the thermostat algorithm, the details of which are often neglected in applied simulation. In general, extended system thermostat algorithms are preferred for cyclopentane because they produce the correct symmetry of motion about the local energy minimum, and the pseudorotational parameters are relatively insensitive to the NVT temperature control coupling paramater Q. This insensitivity of the pseudorotational dynamics to the coupling parameter allows Q to be adjusted to produce the correct fluctuations that will determine thermodynamic and transport properties. If these properties are not a desired output for the simulation, the dynamics of the simulation will remain consistent for any reasonable value of Q, where a reasonable Q value is one in which the thermostat produces fluctuations consistent with the microcanonical ensemble. In contrast, the LFB algorithm, a weak coupling algorithm, exhibited the incorrect symmetry of motion for effective values of Q and a significant dependence of the dynamics on the coupling parameter Q.

Similar accuracy was obtained for all the extended system algorithms for the same choice of force field, indicating that, at least for short simulations, all of the algorithms are equally accurate. While previous studies indicated that the NP algorithm accumulated very little error in the conserved energy quantity for the algorithm, both the NP and the NH-SV

algorithm showed good long term stability, which was shown from the results of very long simulations. At higher temperatures, the NP algorithm is slightly better than NH-SV over long time periods. It was also found that simulations of cyclopentane will be similar for the liquid and gas phases due to the torsion part of the force field being the most important energy contribution.

While the barrier to planarity of the cyclopentane ring was closest to the experimental value for the MMFF94 and AMBER94 force fields, this was not the best parameter to determine the optimal force field. The optimal force field appears to be one which mimics the structure predicted by *ab initio* quantum calculations, since such a force field will properly reproduce the experimental values for the pseudorotational amplitude and velocity. The optimal force field to use for cyclopentane of those investigated was found to be AMBER94, but MMFF94 also displayed dynamics reasonably close to the experimental value.

7.1.2 TSMC Simulations

By attempting to match the value for the limiting value of the characteristic ratio to experiment for poly(*cis*-1,4-butadiene), it was found that MMFF94 was a much better force field for simulating single and double bonds in a polymer backbone compared with AMBER94. This, coupled with the result of similar performance for AMBER94 and MMFF94 for cyclopentane, indicated that MMFF94 would be the best choice of all tested force fields for simulating ROMP PNB.

The main reason to examine ROMP PNB was to determine its conformation to determine if it would be an adequate support for immobilized homogeneous catalysis. The hypothesis for determining whether one polymer might be better than another is that a polymer with a random coil conformation can potentially block catalyst sites with the backbone because of the lack of structural order in the backbone. A polymer which adopted a helix-kink conformation would likely have more accessible catalyst sites because of the order present in the backbone, which would cause the catalyst sites to remain external to the

backbone structure. This would be analogous to the difference in accessibility of a catalyst attached inside of a pore of an inorganic substrate compared to one attached to the surface of the inorganic substrate.

There were three main characteristics that are thought to be required for a polymer to adopt a helix-kink conformation. These were derived from work done on vinylic PNB, PTMSP, and PTBA. The first hypothesis is that the polymer must alternate rotatable and non-rotatable bonds in the polymer backbone. The second is that the non-rotatable bonds must have cis isomerism. The third is that there must be a bulky group which is present at the non-rotatable bonds, preferably alternating sides of the backbone.

The TSMC simulations of ROMP PNB with no side groups displayed the expected behavior, i.e., the low-energy conformation is a random coil. The TSMC simulations that were performed on ROMP PNB that had a side group present on the cyclopentane ring were inconclusive, though they did provide a preliminary glimpse of the likely final conformation of the different configurations of ROMP PNB. This conformation is again that of a random coil. These results were considered inconclusive mainly due to the limitations of TSMC to these particular systems. This was based on the lack of an analytic way to analyze the polymer conformation, as well as the need for larger systems that caused the computational time to be such that the simulations could not be reasonably repeated.

7.1.3 ROMP PNB Accessibility

Because the results of the TSMC simulations were inconclusive, it was decided to use MD to model all of the ROMP PNB configurations. The use of alpha shapes was employed to determine the accessibility of the catalyst sites that were attached to the ROMP PNB side group. A number of assumptions were employed. These assumptions include unperturbed polymer dimensions, a dilute polymer-supported catalysis system, single catalyst and reactant reaction kinetics, and spherically approximated reactants and catalyst sites. These assumptions were employed in order to decrease computation time and reduce complexity.

A multitude of configurational changes unrelated to the polymer backbone were simulated. In addition to the backbone configurations, the polymer size, reactant size, catalyst size, and catalyst linker length were all varied to determine their effect on accessibility. All of these variations displayed the expected effect on accessibility consistent with a random coil. Accessibility decreased as reactant size increased, increased as catalyst size increased, and increased as the linker length was increased. What was unexpected was that all of the configurations of ROMP PNB had the same conformation – that of a random coil polymer. However, it was discovered that very large side groups had an effect on the conformation that gave it some local helical characteristics that did not propagate throughout the entire backbone. This would put this specific polymer configuration in a class of polymers that exhibit some backbone order, but not enough to consider it a rigid rod polymer.

Analyzing the backbone torsion and bond angles for the double bond and cyclopentane pseudo-bond revealed the most likely cause for this behavior. The reason is that the cyclopentane group was too flexible to be considered non-rotatable. It was found that even though the cyclopentane ring is constrained by being part of the polymer backbone, the assumption that it can be considered a non-rotatable bond is invalid. The cyclopentane ring still provides enough flexibility to the backbone that it is unable to constrain the backbone into the desired helix-kink conformation.

7.1.4 ROMP PNBDC Accessibility

The results of the ROMP PNB simulations showed that the cyclopentane ring did not conform to the assumption that it was a non-rotatable pseudo-bond. This was corrected by adding a fused dicarboximide ring to the cyclopentane ring. This was shown to effectively limit the motion of the cyclopentane ring. The simulations that were performed with ROMP PNBDC showed that none of the final conformations had a similar rigid structure as the vinylic PNB had, though the TR configuration did have a loose helical conformation.

The original hypotheses for a helix-kink polymer stated that the polymer must be cis

across all non-rotatable bonds, it must alternate rotatable and non-rotatable bonds, and it must have a bulky side group on the non-rotatable bonds. In this case, the addition of a fused ring provided the desired rotatable/non-rotatable bond structure, and there was a cis conformation available. However, what was overlooked was the fact that the double bond in the backbone had no bulky side group. This is most likely the reason that this structure did not adopt the expected helix-kink conformation. This work on both ROMP PNB and ROMP PNBDC actually serves more to refine the original hypotheses than disprove them. The work on ROMP PNB showed that lacking the proper rotatable/non-rotatable bond structure will not allow a helical conformation. Additionally, the work on ROMP PNBDC showed that the addition of a bulky side group is necessary on all of the non-rotatable bonds. Having a bulky side group on every other non-rotatable bond may not apply the proper steric hindrance required for the polymer to maintain the helix-kink structure.

Even though a rigid helix-kink structure was not found among the final conformations of ROMP PNBDC, a loose helix was discovered for the TR configuration. This particular helix is very broad and takes over 10 repeat units to make one coil of the helix, compared to three repeat units from vinylic PNB. The intermolecular order in the backbone could be the explanation for a number of unique properties that have been found experimentally. It would seem that this particular configuration of this polymer falls into a unique class of polymers which have this intermolecular order without any change in $\langle s^2 \rangle_0$ compared to a random coil.

7.1.5 PCO Co-salen Simulations

The work on Co-salen systems was motivated by experimental work done by Zheng, Jones, and Weck into Co-salen catalysts supported by poly(cyclooctene)[91]. They found that the PCO supports had a synergistic effect with the Co-salen reaction mechanism, and it was desired to try to find an optimum through simulation. Using a combination of ESFF parameters, X-ray diffraction structures, and *ab initio* geometries and charges, parameters

were fit to MMFF94 force field equations that mimicked the Co-salen structure with a RMSD of less than 1 Å compared to both the X-ray diffraction and *ab initio* geometries. The analysis that was chosen for these systems involved a "reaction score," which combined the individual catalyst efficiency with the overall catalyst efficiency.

The model systems used were an oligomer synthesized by Ready and Jacobsen and macrocyclic and linear PCO polymers, each with cis and trans isomers. The macrocyclic PCO supports were found to have a higher reaction score than both the Jacobsen oligomer and the linear PCO polymer. This leads to the conclusion that the macrocyclic PCO support is the optimal support structure of those studied. The cis isomer seems to have much higher reaction scores than the trans isomer, which would suggest that the cis isomer is the better support. Additionally, it was found that the optimal number of repeat units was 3 for the cis macrocycles and 4 for the trans macrocycles.

Examination of the optimal supports showed a trend in what made that support better than the others. Of the macrocyclic PCO supports, the ones that were optimal for both the cis and trans isomers were those that performed more than one cobalt pairing. For the trans tetramer, this consisted of two pairs that were consistent throughout the simulation. For the cis trimer, one of the catalysts alternated pairings with both neighbors. This double pairing is what gave these particular structures increased reaction scores compared to all of the other structures.

7.2 Recommendations and Future Work

The most useful aspect of cyclopentane for computer-based molecular simulations is that it is a compound which has known and experimentally verified molecular motion. Barth, Leimkuhler, and Reich have proposed a number of systems as a test set for molecular dynamics, and to an extent, force fields. They suggested that MD algorithms should be tested by water, butane, a Lennard-Jones fluid, and the alanine dipeptide[107]. It is recommended that this test set be extended to include cyclopentane, as it was shown to provide a way to

perform a robust comparison of both MD algorithms and force fields.

To extend the accessibility metric used into a more useful comparison metric, it is recommended that future work take into account the issue of a dilute system, molecule size, and linker length. For this work, these are all similar to experimental conditions, which means that accessibility is not an issue for these systems. However, if it were decided that a longer polymer was needed or that the system was no longer dilute, these issues become more prevalent, and this method would be much more useful.

Additionally, work has been done on object models for the accessibility of facets to the exterior surface[108]. This method would be a useful extension of the alpha shapes application. Alpha shapes are already comprised of facets. By using the information from individual facets, the accessibility can be found for specific molecular regions. For example, a catalyst like Co-salen, where the active region is specific to a small area on one side of the molecule, would be approximated as a sphere where the alpha shape is a tetrahedron in which one of the facets is the only one considered for accessibility. This would allow the continued use of both a spherical catalyst approximation and alpha shapes. The link between the simulation and reality would be strengthened while keeping the simulation complexity at a lower level.

The work on ROMP PNBDC could be extended to other polymers to determine the reason that the TR configuration has a helical conformation. Because this helical conformation is not the same as the vinylic PNB helix, there is no reason that the original hypotheses would be the same for this loose helical class of polymers. Additionally, it would be interesting to determine if a less polar fused ring would produce a similar structure. Also, this work should be extended to perform bulk simulations that will generate simulate WAXD patterns. Experiments should be performed to generate WAXD patterns of similar structures to be able to have a comparison of simulation and experimental data.

The work on the Co-salen systems was an initial first-pass effort into these interesting supported systems. There are several possible directions that future studies can take. The

first would be to determine the proper experimental perceived dielectric from the solvent on the support system. Also, if this dielectric changes due to the change in solution from the reaction, it would be necessary to capture that so that both the initial and final stages of the reaction can be simulated. This is very important, as the number of possible reaction events should increase with the solvent dielectric. If the solvent dielectric decreases over the course of the reaction, this would have an effect on the efficiency of a particular support.

In addition, future studies of Co-salen supports should probably not contain the full counter-ion, but rather just the hydroxy group. This is because when the actual reaction occurs, the ligand with a hydroxy group is the actual structure of one of the Co-salen ligands involved in the reaction. Also, an approximate reaction separation distance should be found through high-level quantum calculations. Having both of these should improve the results of the reaction score determination and ensure that it more closely matches the actual reaction conformations.

Finally, the analysis of the possible reaction events should include the reaction probability rather than just the difference in number of catalyst supports. The inclusion of reaction probability would provide the analysis with a more robust method of determining the overall reaction score. It would also more realistically take into account the difference in number of repeat units of the supports. This is mainly because this probability would manifest itself in a slope change on a plot of the enantiomeric excess over time. The probability would take into account the fact that at the start of the reaction, there is only a one-in-four chance that a given solvent molecule in the reaction space is the desired reactant. Additionally, it would take into account the fact that each successful reaction event decreases the probability of the proper reaction geometry containing the proper reactant molecule in the reaction space. The addition of this probability would effectively extend these simulations to having the ability to compare the actual reaction profile with time.

APPENDIX A

FORCE FIELD EQUATIONS

The energy for a given force field is determined from Equation 1. A more complete form of this equation is

$$E_{total} = \sum E_{stretch} + \sum E_{bend} + \sum E_{torsion} + \sum E_{oop} + \sum E_{cross} + \sum E_{es} + \sum E_{vdw}, \quad (22)$$

where $E_{stretch}$ is the energy associated from a bond stretching between two bonded atoms; E_{bend} is the energy associated from the angle bend between three bonded atoms; $E_{torsion}$ is the energy associated with the torsion angle between four bonded atoms; E_{oop} is the out-of-plane bending term at tricoordinate centers; E_{cross} is energy associated with any additional cross terms; E_{es} is the energy associated with the nonbonded electrostatic pair interactions; and E_{vdw} is the energy associated with the van der Waals dispersion interactions, though usually bonded atoms do not include nonbonded interactions for the nearest three bonded neighbors. Each term is summed over the energy values from each interaction, i.e., the total bond stretch energy is the sum of each bond stretch term from each individual bond. There are several different equations for each term, and different force fields use different equations. Rather than reproducing all of the possible force field equations, the equations for each energy term that are described here will be limited to those used in MMFF94[109, 56, 110, 111, 112, 113] and ESFF[104], as these are the ones that were used to determine parameters for cobalt. The equations for ESFF will be limited to those that correspond to cobalt atoms as used in this work. For a more detailed description of possible force field equations, the literature for an individual force field will have the equations available for that force field. Additionally, Jensen[114] has compiled a good overview of more possible force field equations.

A.1 Bond Stretch Energy

MMFF94 uses a quartic form for the bond stretch energy. This equation for the bond stretch energy between two atoms i and j is

$$E_{stretch} = k_2(r_{ij} - R_0)^2 + k_3(r_{ij} - R_0)^3 + k_4(r_{ij} - R_0)^4, \tag{23}$$

where k_x is the bond stretch force constant for that term, r_{ij} is the distance between atoms i and j, and R_0 is the equilibrium bond length for atoms i and j. ESFF uses a much different form for the bond stretch energy. ESFF uses a Morse potential, which is

$$E_{stretch} = D_{ij}\left(1 - e^{-\alpha_{ij}(r_{ij}-R_0)}\right)^2, \tag{24}$$

where D_{ij} is the bond dissociation energy and α is the force constant.

A.2 Angle Bend Energy

MMFF94 uses a similar polynomial form for the angle bend energy for normal angles. The equation for the angle bend energy from the angle formed between atoms i, j, and k is

$$E_{bend} = k_{\theta 2}(\theta_{ijk} - \theta_0)^2 + k_{\theta 3}(\theta_{ijk} - \theta_0)^3, \tag{25}$$

where $k_{\theta x}$ is the force constant for that term, θ is the angle formed between i, j, and k, and the subscripts have the same meaning as before. ESFF uses

$$E_{bend} = \frac{k_\theta}{\sin^2\theta_0}(\cos\theta_{ijk} - \cos\theta_0)^2, \tag{26}$$

where the symbols have the same meaning. Both MMFF94 and ESFF use a separate equation for linear angles which is

$$E_{bend} = k_\theta(1 + \cos\theta_{ijk}) \tag{27}$$

for MMFF94, and

$$E_{bend} = k_\theta \cos^2 \theta_{ijk} \tag{28}$$

for ESFF. ESFF also uses this equation for perpendicular angles.

A.3 Torsion Energy

The torsion energy describes the angle that exists between atoms i, j, k, and l. For MMFF94, this equation is

$$E_{torsion} = \frac{V_1}{2}(1 + \cos \Phi_{ijkl}) + \frac{V_2}{2}(1 - \cos 2\Phi_{ijkl}) + \frac{V_3}{2}(1 + \cos 3\Phi_{ijkl}), \tag{29}$$

where V_x is the x-fold Fourier component for the four atoms, and Φ_{ijkl} is the torsion angle. ESFF does not use torsion terms for cobalt.

A.4 Out-of-plane Bending

This term describes how far a given center is out of a defined plane. This term is only used at tricoordinate centers, and it is mainly used to add an energy penalty either to unrealistic pyramidalization (sp2 hybridized atoms) or to planar sp3 hybridized atoms that should be pyramidal. Both MMFF94 and ESFF use this term, and it is described by

$$E_{oop} = k_{ijk;l}\, \chi^2_{ijk;l}, \tag{30}$$

where $k_{ijk;l}$ is the force constant, and $\chi_{ijk;l}$ is the Wilson angle. This angle is defined as the angle between the bond $j - l$ and the plane $i - j - k$.

A.5 Cross Terms

Cross terms for the potential energy include any terms that are a combination of any of the previous terms. In this case, MMFF94 uses one cross term that combines bond stretch and angle bend interactions. This stretch-bend equation is

$$E_{cross} = E_{strb} = [k_{ijk}(r_{ij} - R_0) + k_{kji}(r_{kj} - R_0)](\theta_{ijk} - \theta_0), \tag{31}$$

where the force constants couple the $i - j$ and $k - j$ stretches to the $i - j - k$ bend.

A.6 van der Waals Interactions

MMFF94 uses a "Buffered 14-7" form of the van der Waals energy, which is

$$E_{vdw} = \epsilon_{ij}\left(\frac{1.07R_{ij}}{r_{ij} + 0.07R_{ij}}\right)^7\left(\frac{1.12R_{ij}^7}{r_{ij}^7 + 0.12R_{ij}^7} - 2\right), \tag{32}$$

where R_{ij} is the minimum energy distance, ϵ is the minimum energy, and r_{ij} is the current distance between the atoms. This equation is true for all pairs of atoms in the system that are either not bonded or are separated by three or more chemical bonds. Therefore, van der Waals 1-4 interactions are included and are not scaled. ESFF uses a simpler Lennard-Jones 6-9 potential, which is

$$E_{vdw} = \epsilon_{ij}\left(2\frac{R_{ij}^9}{r_{ij}^9} - 3\frac{R_{ij}^6}{r_{ij}^6}\right). \tag{33}$$

A.7 Electrostatic Energy

The electrostatic energy for ESFF will not be listed here, as it was not used in this work. However, charges were fit to MMFF94 charges, so MMFF94 electrostatic energy will be described here. The partial charges are calculated and used in a "buffered coulombic" form of the electrostatic equation. These equations are

$$q_i = q_i^0 - \sum w_{ik} \tag{34}$$

$$E_{es} = \frac{C\ q_i q_j}{D\ (r_{ij} + 0.05)}, \tag{35}$$

where q_i and q_j are the partial charges for atoms i and j, q^0 is the formal charge, w_{ik} is the bond charge increment ($w_{ik} = -w_{ki}$), C is a unit conversion constant, D is the dielectric

constant, and r_{ij} is the distance between i and j. The pair interactions for electrostatics follow almost the same rules as the van der Waals interactions, with one exception. For the electrostatic energy, electrostatic 1-4 interactions are scaled by 0.75.

APPENDIX B

MOLECULAR DYNAMICS ALGORITHM EQUATIONS

There were five different molecular dynamics algorithms that were compared to each other in various degrees in Chapter 2. These algorithms are the Berendsen temperature controller integrated with the leapfrog Verlet algorithm (LFB)[39], the Nosé-Hoover temperature controller integrated by the Störmer-Verlet leapfrog algorithm (NH-SV)[40], the Nosé-Hoover temperature controller integrated by an implicit integration scheme based on a modification of the velocity Verlet algorithm (NH)[9], the symplectic Nosé-Poincaré temperature controller integrated by the generalized leapfrog algorithm (NP)[41], and Nosé's explicitly integrated implementation of the Nosé-Poincaré algorithm (NPe)[63].

This section provides the full formulation of each of these algorithms for one iteration. Each algorithm is written in real variables. The notation follows that of Bond, Leimkuhler, and Laird[41], where the step in time from t^n to $t^{n+\Delta t}$ is designated by the superscripts. For each algorithm, $\boldsymbol{x}$ signifies position, $\boldsymbol{v}$ is the velocity, $\boldsymbol{a}$ is the acceleration, $U(\boldsymbol{x})$ is the potential energy, K is the kinetic energy, Δt is the timestep, T is the temperature, and m is the mass. The subscript i is used to denote that each operation is done on each individual atom. The target temperature is denoted by T_0, g is used to denote the degrees of freedom, and k_b is Boltzmann's constant. For the NH algorithms, Q is the "mass," ξ is the "momentum," and η is the "position" of the thermostat. For the NP algorithms, these variables are denoted by Q, π, and s, respectively. Additionally, these algorithms also use H_0, which is the initial Nosé Hamiltonian.

Two other definitions are necessary. The first is the kinetic energy,

$$K = \frac{1}{2}\sum m_i \boldsymbol{v}_i^2,$$

which is the standard definition of kinetic energy. The second is temperature, which is

defined from equipartion by

$$\begin{aligned} K &= \frac{3}{2} N k_b T \\ T &= \frac{\sum m_i \boldsymbol{v}_i^2}{3 N k_b}. \end{aligned}$$

The LFB algorithm:

$$\begin{aligned} \boldsymbol{a}_i^n &= -\frac{\nabla_i U(\boldsymbol{x}^n)}{m_i} \\ \lambda &= \left[1 + \frac{\Delta t}{Q} \left(\frac{T_0}{T^{n-1/2}} - 1 \right) \right]^{1/2} \\ \boldsymbol{v}_i^{n+1/2} &= \lambda \left(\boldsymbol{v}_i^{n-1/2} + \boldsymbol{a}_i \Delta t \right) \\ \boldsymbol{x}_i^{n+1} &= \boldsymbol{x}_i^n + \boldsymbol{v}_i^{n+1/2} \Delta t \end{aligned}$$

The NHimp algorithm:

$$\begin{aligned} \boldsymbol{v}_i^{n+1/2} &= \boldsymbol{v}_i^n + \left(\boldsymbol{a}_i^n - \xi^n \boldsymbol{v}_i^n \right) \frac{\Delta t}{2} \\ \xi^{n+1/2} &= \xi^n + \left(K^n - g k_b T_0 \right) \frac{\Delta t}{2Q} \\ \boldsymbol{x}_i^{n+1} &= \boldsymbol{x}_i^n + \boldsymbol{v}_i^{n+1/2} \Delta t \\ \eta^{n+1} &= \eta^n + \xi^{n+1/2} \Delta t \\ \boldsymbol{a}_i^{n+1} &= -\frac{\nabla_i U(\boldsymbol{x}^{n+1})}{m_i} \\ \boldsymbol{v}_i^{n+1} &= \boldsymbol{v}_i^{n+1/2} + \left(\boldsymbol{a}_i^{n+1} - \xi^{n+1} \boldsymbol{v}_i^{n+1} \right) \frac{\Delta t}{2} \\ \xi^{n+1} &= \xi^{n+1/2} + \left(K^{n+1} - g k_b T_0 \right) \frac{\Delta t}{2Q} \end{aligned}$$

The NHSV algorithm:

$$
\begin{aligned}
\boldsymbol{v}_i^{n+1/2} &= \boldsymbol{v}_i^n + \left(\boldsymbol{a}_i^n - \xi^n \boldsymbol{v}_i^{n+1/2}\right)\frac{\Delta t}{2} \\
\boldsymbol{x}_i^{n+1} &= \boldsymbol{x}_i^n + \boldsymbol{v}_i^{n+1/2}\Delta t \\
\xi^{n+1} &= \xi^n + \left(K^{n+1/2} - g k_b T_0\right)\frac{\Delta t}{Q} \\
\boldsymbol{x}_i^{n+1} &= \boldsymbol{x}_i^n + \boldsymbol{v}_i^{n+1/2}\Delta t \\
\eta^{n+1} &= \eta^n + \left(\xi^{n+1} + \xi^n\right)\frac{\Delta t}{2} \\
\boldsymbol{a}_i^{n+1} &= -\frac{\nabla_i U(\boldsymbol{x}^{n+1})}{m_i} \\
\boldsymbol{v}_i^{n+1} &= \boldsymbol{v}_i^{n+1/2} + \left(\boldsymbol{a}_i^{n+1} - \xi^{n+1}\boldsymbol{v}_i^{n+1/2}\right)\frac{\Delta t}{2}
\end{aligned}
$$

The NP algorithm:

$$
\begin{aligned}
\boldsymbol{v}_i^{n+1/2} &= \boldsymbol{v}_i^n + \boldsymbol{a}_i^n\frac{\Delta t}{2} \\
\pi^{n+1/2} &= \pi^n + \left[K^{n+1/2} - g k_b T_0(1 + \ln s^n) - U(\boldsymbol{x}^n) - \frac{(\pi^{n+1/2})^2}{2Q} + H_0\right]\frac{\Delta t}{2} \\
s^{n+1} &= s^n + (s^n + s^{n+1})\frac{\pi^{n+1/2}}{Q}\frac{\Delta t}{2} \\
\boldsymbol{x}_i^{n+1} &= \boldsymbol{x}_i^n + \left(\frac{s^n}{s^{n+1}} + 1\right)\boldsymbol{v}_i^{n+1/2}\frac{\Delta t}{2} \\
\boldsymbol{a}_i^{n+1} &= -\frac{\nabla_i U(\boldsymbol{x}^{n+1})}{m_i} \\
\pi^{n+1} &= \pi^{n+1/2} + \left[K^{n+1/2}\left(\frac{s^n}{s^{n+1}}\right)^2 - g k_b T_0(1 + \ln s^{n+1}) - U(\boldsymbol{x}^{n+1}) - \frac{(\pi^{n+1/2})^2}{2Q} + H_0\right]\frac{\Delta t}{2} \\
\boldsymbol{v}_i^{n+1} &= \boldsymbol{v}_i^{n+1/2}\frac{s^n}{s^{n+1}} + \boldsymbol{a}_i^{n+1}\frac{\Delta t}{2}
\end{aligned}
$$

The NPe algorithm:

$$
\begin{aligned}
s^{n+1/2} &= s^n \left(1 + \frac{\pi^n}{2Q}\frac{\Delta t}{2}\right)^2 \\
\pi^* &= \frac{\pi^n}{\left(1 + \frac{\pi^n}{2Q}\frac{\Delta t}{2}\right)} \\
\boldsymbol{v}_i^{n+1/2} &= \boldsymbol{v}_i^n \frac{s^n}{s^{n+1/2}} + \boldsymbol{a}_i^n \frac{\Delta t}{2} \\
\boldsymbol{x}_i^{n+1} &= \boldsymbol{x}_i^n + \boldsymbol{v}_i^{n+1/2}\Delta t \\
\boldsymbol{a}_i^{n+1} &= -\frac{\nabla_i U(\boldsymbol{x}^{n+1})}{m_i} \\
\boldsymbol{v}_i^{n+1} &= \left(\boldsymbol{v}_i^{n+1/2} + \boldsymbol{a}_i^{n+1}\frac{\Delta t}{2}\right)\frac{s^{n+1}}{s^{n+1/2}} \\
\pi^{**} &= \pi^* + \left[K^{n+1/2} - \frac{1}{2}\left(U(\boldsymbol{x}^n) + U(\boldsymbol{x}^{n+1})\right) - g k_b T_0 (1 + \ln s^{n+1/2}) + H_0\right]\Delta t \\
s^{n+1} &= s^{n+1/2}\left(1 + \frac{\pi^{**}}{2Q}\frac{\Delta t}{2}\right)^2 \\
\pi^{n+1} &= \frac{\pi^{**}}{\left(1 + \frac{\pi^{**}}{2Q}\frac{\Delta t}{2}\right)}
\end{aligned}
$$

APPENDIX C

MOLECULAR DYNAMICS CODE

The following is the molecular dynamics code that was used for the poly(norbornene) and poly(cyclooctene) simulations. This code was originally the MOE MD code found in the "md.svl" file from the 2005.06 version of MOE, but it was changed to add functionality and decrease computation time. All code found here and in subsequent appendices is written in the Scientific Vector Language (SVL), which is the programming language used in MOE. Any calls to functions that are not contained in the code are calls to functions provided by the MOE API.

The major changes included a unit conversion fix, the addition of SETTLE water constraints, improved long-range corrections, the addition of NH-SV, improvements to the LINCS constraint implementation, and improvements to the GUI (not included here). Several of these improvements, which were added based on the submission of these code changes to the Chemical Computing Group, can be currently found in the 2007.09 and newer versions of MOE. It is important to note that any code which was unchanged from the 2005.06 version of MOE is copyrighted by the Chemical Computing Group.

```
#svl
// md.svl molecular dynamics
//
// 20-oct-2006 (as) changed NH to full vel step version (more accurate)
//  removed explicit NP since it is less accurate than
//     implicit version
// 20-jun-2006 (as) cleaned up code contributions
// 31-may-2006 (as) cleaned up GUI
// 26-may-2006 (as) added Nose-Hoover
// 24-may-2006 (as) fixed velocity init, added SETTLE
// 30-apr-2006 (as) improved long range corrections (mainly for MMFF94)
// 27-apr-2006 (as) converted back to A,kcal units
// 18-jul-2005 (pl) NaN message fix, cool temp, new browse protocol
//   29-jun-2005 (ac) minor change to browser parameter interpretation
// 05-apr-2005 (pl) minimum 20K temperature, added browser
// 16-nov-2004 (pl) converted to nm/kJ units
```

```
// 28-jul-2003 (pl) rewrote for Nose-Poincare-Anderson hamiltonian
// 20-jun-2003 (pl) remove linear and angular momentum periodically
// 08-jun-2003 (pl) rewrote
// 15-apr-2002 (sb) confirm overwrite + browse button
// 01-jun-1999 (jd) added open in database viewer checkbox
// 05-feb-1999 (al) bubble help typo
// 06-mar-1997 (pl) fixed atoms ...
// 02-dec-1996 (pl) new function names
// 21-oct-1996 (pl) bubble help
// 17-oct-1996 (pl) removed set_velocity
// 04-oct-1996 (pl) notice
//
// COPYRIGHT (C) 1996-2005 CHEMICAL COMPUTING GROUP INC.  ALL RIGHTS RESERVED.
//
// PERMISSION TO USE, COPY, MODIFY AND DISTRIBUTE THIS SOFTWARE IS HEREBY
// GRANTED PROVIDED THAT: (1) UNMODIFIED OR FUNCTIONALLY EQUIVALENT CODE
// DERIVED FROM THIS SOFTWARE MUST CONTAIN THIS NOTICE; (2) ALL CODE DERIVED
// FROM THIS SOFTWARE MUST ACKNOWLEDGE THE AUTHOR(S) AND INSTITUTION(S); (3)
// THE NAMES OF THE AUTHOR(S) AND INSTITUTION(S) NOT BE USED IN ADVERTISING
// OR PUBLICITY PERTAINING TO THE DISTRIBUTION OF THE SOFTWARE WITHOUT
// SPECIFIC, WRITTEN PRIOR PERMISSION; (4) ALL CODE DERIVED FROM THIS SOFTWARE
// BE EXECUTED WITH THE MOLECULAR OPERATING ENVIRONMENT (MOE) LICENSED FROM
// CHEMICAL COMPUTING GROUP INC.
//
// CHEMICAL COMPUTING GROUP INC. DISCLAIMS ALL WARRANTIES WITH REGARD TO THIS
// SOFTWARE, INCLUDING ALL IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS,
// AND IN NO EVENT SHALL CHEMICAL COMPUTING GROUP INC. BE LIABLE FOR ANY
// SPECIAL, INDIRECT OR CONSEQUENTIAL DAMAGES OR ANY DAMAGES WHATSOEVER
// RESULTING FROM LOSS OF USE, DATA OR PROFITS, WHETHER IN AN ACTION OF
// CONTRACT, NEGLIGENCE OR OTHER TORTIOUS ACTION, ARISING OUT OF OR IN
// CONNECTION WITH THE USE OR PERFORMANCE OF THIS SOFTWARE.
//
// References:
//
// READ ME:
// Allen, M.P., Tildesley, D.J.;
// Computer Simulation of Liquids
// (1987) Oxford University Press, New York
//
// Nose-Poincare (NVT ensemble):
// Bond, S.D., Leimkuhler, B.J., Laird, B.B.;
// The Nose-Poincare Method for Constant Temperature Molecular Dynamics
// J. Comp. Phys 151 (1999) 114-134
//
// Nose-Poincare-Andersen (NPT ensemble):
// Sturgeon, J.B., Laird, B.B.;
// Syplectic Algorithm for Constant-Pressure Molecular Dynamics Using
// a Nose-Poincare Thermostat
// J. Chem. Phys. 112 (2000) 3474-3482
//
// LINCS:
// Hess, B., Bekker, H., Berendsen, H.J.C., Fraaije, J.G.E.M.
// LINCS: A Linear Constraint Solver for Molecular Simulations
// J. Comp. Chem. 18 (1997) 1463-1472
```

```
//
// SETTLE:
// Miyamoto, S., Kollman, P.A.;
// SETTLE: An Analytical Version of the SHAKE and RATTLE Algorithm
// for Rigid Water Models
// J. Comp. Chem. 13 (1992) 952-962
//
// Original paper by Nose:
// Nose, S.;
// A Molecular Dynamics Method for Simulations in the Canonical Ensemble
// Mol. Phys. 52 (1984) 255-268
//
// Original paper by Hoover:
// Hoover, W.G.;
// Canonical Dynamics: Equilibrium Phase-Space Distributions
// Phys. Rev. A 31 (1985) 1695-1697
//
// Nose-Hoover Stoermer-Verlet Algorithm:
// Holian, B.L., De Groot, A.J., Hoover, W.G., Hoover, C.G.;
// Time-reversible Equilibrium and Nonequilibrium Isothermal-Isobaric
// Simulations with Centered-difference Stoermer Algorithms
// Phys. Rev. A 41 (1990) 4552-4553
//
// TO DO
// - low temperature fluctuations cause NaNs

#set title 'Molecular Dynamics'
#set class 'MOE:simulation'
#set version '2006.10'

const DISPLAY_ERRORS = 0; // display constrain errors in MOE window?
const BOX_MIN = 1; // minimum size to enable periodic boundaries

function fwrite_MOE, fread_MOE;

const MD_DEFAULTS = [
    dt: 0.002,// time step (ps)
    sample: 0.5,// sample period (ps)
    T0: 300,// initial velocities temperature
    algorithm: 'NH',// MD algorithm
    QT: 1,// temperature response time (ps)
    QP: 0.5,// pressure response time (ps)
    waterConstraints: 'settle',// ['none', 'bonds', 'settle']
    bondConstraints: 'light',// ['none', 'light', 'all']
    constraintTol: 1e-9,// tolerance for holonomic constraints
    savePosition: 0,// save positions to database
    saveVelocity: 0,// save velocities to database
    saveAcceleration: 0 // save accelerations to database
];

// MD Units:
//
// length:   Angstroms (fundamental)
// time:   ps
```

```
// mass:   amu (g/mol)
//
// energy:   kcal/mol (derived)
// pressure: kPa

const PCONV = 1.43953e-7; // pressure convert: kPa -> kcal/mol/A**3
const VAECONV = 418.4; // velocity and acceleration convert:
// kcal/g -> A**2/ps**2
// kcal/g/A -> A/ps**2
// inv VAECONV converts g*A**2/ps**2/mol
//     -> kcal/mol

// ------------------------------ UTILITIES ----------------------------------

#if 0
local function segext [seg, len_x]
    local last_seg = mod [len_x, seg];
    if seg then seg = resize [seg, ceil (len_x / seg)]; endif
    if last_seg then seg(length seg) = last_seg; endif
    return seg;
endfunction

local function stretch [x, seg]
    if length seg == 1 then seg = resize [seg, length x]; endif
    return x[pscan droplast ctof seg];
endfunction

local function s_add [x, seg]
    if length seg == 1 then seg = segext [seg, length x]; endif
    x = mget [pscan x, rotl ctom seg];
    return unpack [x - rotrpoke [x, 0], 0, seg];
endfunction
#endif

local function vcross [u,v] // u x v
    local normu = maxE abs u, normv = maxE abs v;
    u = u * [invz normu];
    v = v * [invz normv];
    return [normu * normv] * (rotl u * rotr v - rotl v * rotr u);
endfunction

local function div1p1m x // (1+x)/(1-x)
    local idx = x_pack (x > 0);
    local f = - tanh (0.5 * log abs x);
    return put [f, idx, inv f[idx]];
endfunction

// To retain compatibility with MOE without db_PrevEntry we need wrapper
// functions to simulate quickly getting the last entry

local function mdb_Append [mdb, data] = db_Write [mdb, 0, data];
local function mdb_LastEntry mdb = db_PrevEntry [mdb, 0];

local function NearestImage [dpos, box]
```

```
    box = box * (box >= BOX_MIN and box < REAL_MAX);
    if orE (box > 0) then dpos = dpos - box * round (dpos * invz box); endif
    return dpos;
endfunction

// ---------------------------------- RESTART -------------------------------

// In order to restart the calculation after a termination of some kind, we
// write the global data required in MOE into the output database.

local function tvar v = token swrite ['{v}',v];

local function storefile filename // storable filename
    local i;

    const PREFIX = tr [
[ cat [string MOE, "/"],"$MOE/" ],
[ cat [string HOME, "/"],"~/" ],
[ cat [string TMP, "/"],"$TMP/" ]
    ];

    filename = string filename;

    for i = 1, l_length PREFIX loop
if keep[filename, length PREFIX(1)(i)] === PREFIX(1)(i) then
    filename = drop [filename, length PREFIX(1)(i)];
    filename = cat [PREFIX(2)(i), filename];
    break;
endif
    endloop

    return token filename;
endfunction

// dynamics options are written to the the 'MD:Options' environment variable
// as a token tagged vector.

local function SaveRestartOptions [mdb, opt]
    opt = opt | m_uniq tags opt;
    opt = opt | indexof [tags opt, tags MD_DEFAULTS];
    db_SetEnv [mdb, 'MD:Options', tvar opt];
endfunction

local function GetRestartOptions mdb
    return first first sread [string db_GetEnv [mdb, 'MD:Options'], '{v}'];
endfunction

// the forcefield filename is stored into the database in the environment
// variable 'MD:PotentialFilename'.  We also store the setup in
// the variable 'MD:PotentialSetup' as a tokenized svl var

local function SaveRestartPotential mdb
    local potfile = (pot_Info[]).filename;
    db_SetEnv [mdb, 'MD:PotentialFilename', storefile potfile];
```

```
    db_SetEnv [mdb, 'MD:PotentialSetup', tvar PotSetup[]];
endfunction

local function GetRestartPotential mdb
    local filename = token db_GetEnv [mdb, 'MD:PotentialFilename'];
    if filename <> '' then pot_Load filename; endif
    PotSetup first first sread [
string db_GetEnv [mdb, 'MD:PotentialSetup'], '{v}'
    ];
endfunction

// the molecular system is stored in the environment variable
// 'MD:MolecularSystem' as a MOE file

local function SaveRestartSystem mdb
    local f = fopenw '';
    fwrite_MOE [f, Chains[]];
    fseek [f,0];
    local data = freadb [f, 'char', INT_MAX];
    fclose f;
    db_SetEnv [mdb, 'MD:MolecularSystem', token data];
    local chains = Chains[];
endfunction

local function GetRestartSystem mdb
    local data = string db_GetEnv [mdb, 'MD:MolecularSystem'];
    local f = fopenw '';
    fwriteb [f, 'char', data];
    fseek [f,0];
    local chains = fread_MOE f;
    fclose f;
    return chains;
endfunction

// ------------------ CONSTRAINED BOND LENGTH SOLVER ------------------------

// ConstrainInit initializes the data structures required for maintaining bond
// length constraints.

local function ConstrainInit [opt, atoms, invmass]
    local N = atoms + 1; // constraint coding
    local idx, mask, seg;

    local A = [], B = [], lenAB = [], wA = [], wB = [], wlen = [], settle = [];

// obtain the water constraints: 'none', 'bonds', 'settle'.  If not
// 'none' then add the bonds in water.  If settle, then use SETTLE.

    local wmask = sm_Match [ '[#8Q0!$(*~[!D1])]', atoms];
    local Owater = atoms | wmask;

    if opt.waterConstraints === 'settle' then

local sA = indexof [Owater, atoms];
```

```
local sB = sA + 1;
local sC = sA + 2;

// remove any fixed atoms
[sA,sB,sC] = [sA,sB,sC] || [
    invmass[sA] > 1e-6 and invmass[sB] > 1e-6 and invmass[sC] > 1e-6
];

// find ideal lengths to be used in SETTLE
local lenABstl = second pot_Parm_str [atoms[sA], atoms[sB]];
local lenAC = second pot_Parm_str [atoms[sA], atoms[sC]];
local lenBC = sqrt (sqr lenABstl + sqr lenAC - 2 * lenABstl * lenAC *
    cos second pot_Parm_ang [atoms[sB], atoms[sA], atoms[sC]]
    );
settle.rc = lenBC/2;
local mass = inv invmass;

// The average distance between the oxygen and each hydrogen is used
// (in case the two are different)
settle.ra = (mass[sB]+mass[sC]) / (mass[sA] + mass[sB] + mass[sC]) *
    sqrt( sqr((lenABstl + lenAC)/2) - sqr settle.rc);
settle.rb = sqrt( sqr( (lenABstl + lenAC)/2 ) - sqr settle.rc)
    - settle.ra;

settle.A = sA; settle.B = sB; settle.C = sC;
settle.idx = cat [sA,sB,sC];

    elseif opt.waterConstraints === 'bonds' then
wA = stretch [indexof [Owater, atoms], aBondCount Owater];
wB = indexof [cat aBonds Owater, atoms];
wlen = second pot_Parm_str [atoms[wA], atoms[wB]];

    endif

// obtain the non-water constraints: 'none', 'light', 'all'.  If
// not none, then add the bond lengths to light atoms.  If 'all' then
// add all bond lengths.

    if opt.bondConstraints === 'light' or opt.bondConstraints === 'all' then
A = cat apt rep [x_id atoms, aBondCount atoms];
B = indexof [cat aBonds atoms, atoms];

[A,B] = [A,B] || [A and B and A < B];
[A,B] = [A,B] || not [wmask[A] or wmask[B]];

if opt.bondConstraints === 'light' then
    [A,B] = [A,B] || [invmass[A] >= inv 1.1 or invmass[B] >= inv 1.1];
endif

lenAB = second pot_Parm_str [atoms[A],atoms[B]];
    endif

// add the bond constraints together and remove any fixed atoms
```

```
    [A,B,lenAB] = apt cat [[A,B,lenAB], [wA,wB,wlen]];
    [A,B,lenAB] = [A,B,lenAB] || [invmass[A] > 1e-6 and invmass[B] > 1e-6];

// for each atom jA(i) that joins >1 constraint, calculate
// the set of constraint indices, jC(i), that it joins

    local [jA,jC] = [ cat [A,B], cat rep [igen l_length[A,B], 2] ];
    [jA,jC] = apt get [[jA,jC], [x_sort jA]];

    jA = jA | (mask = (jA <> rotrpoke [jA, -1]));
    jC = split [jC, seg = mtoc mask];

    [jA,jC] = [jA,jC] || [seg > 1];

// for each constraint, conI(i), with an atom shared by another
// constraint, calculate the set of constraints indices conJ(i)
// that have an atom in common and set conA to be that atom

    local i;

    local conI = rep [0, l_length [jA,jC]];
    local conJ = rep [[], length conI], conA = rep[[], length conI];

    for i = 1, l_length [jA,jC] loop
local n = length jC(i);
local pairs = [stretch [jC(i), n], cat rep [jC(i), n]];
pairs = pairs || [neE pairs];
conI(i) = pairs(1);
conJ(i) = pairs(2);
conA(i) = rep [jA(i), l_length pairs];
    endloop

    conI = cat conI; conJ = cat conJ; conA = cat conA;
    [conI,conJ,conA] = apt get [[conI,conJ,conA], [x_sort conJ]];
    [conI,conJ,conA] = apt get [[conI,conJ,conA], [x_sort conI]];

    local conseg = mtoc (conI <> rotrpoke [conI, -1]);

// set the conC[i,j] to be:
// sign * invmass[c] * diagS[i] * diagS[con[i,j]],
// where sign is -1 if atomA[i]=atomA[con[i,j]] or
// atomB[i]=atomB[con[i,j]] and 1 otherwise and c is the
// atom coupling constraints i and con[i,j]

    local diagS = sqrt (invmass[A] + invmass[B]);
    local diagSinv = inv diagS;

    local conC  = (
  invmass[conA]
* (diagSinv[conI] * diagSinv[conJ])
* select[ -1, 1, A[conI] == A[conJ] or B[conI] == B[conJ] ]
    );

    local [Aidx,Amask] = sam A;
```

```
    local [Bidx,Bmask] = sam B;

    return [
invmass: invmass,
atomA: A,
atomB: B,
Aidx: Aidx,
Aseg: mtoc Amask,
Aput: A[Aidx] | Amask,
Bidx: Bidx,
Bseg: mtoc Bmask,
Bput: B[Bidx] | Bmask,
lenAB: lenAB,
diagSinv: diagSinv,
conI: conI,
conIu: conI | ctom conseg,
conJ: conJ,
conseg: conseg,
conC: conC,
settle: settle
    ];
endfunction

local function Bmatrix [cons, cell, pos]
    local xA = cons.atomA, xB = cons.atomB;
    local rAB = NearestImage [apt get[pos,[xA]] - apt get[pos,[xB]], cell];
    local dAB = norm rAB;
    return [rAB * [invz dAB], dAB];
endfunction

local function Bmul [cons, B, x] = add ( // B x
    B * (apt get [x, [cons.atomA]] - apt get [x, [cons.atomB]])
);

local function addMBmul [cons, x, B, y] // x + M**-1 B' y
    if l_length B then
y = B * [y];
local minv = cons.invmass, Aput = cons.Aput, Bput = cons.Bput;
x = apt put [x, [Aput], apt get [x, [Aput]]
    + [minv[Aput]] * apt s_add[apt get [y, [cons.Aidx]], [cons.Aseg]]
];
x = apt put [x, [Bput], apt get [x, [Bput]]
    - [minv[Bput]] * apt s_add[apt get [y, [cons.Bidx]], [cons.Bseg]]
];
    endif
    return x;
endfunction

// invBMBmul solves the matrix equation (B M**-1 B') sol = rhs where B
// the bond direction matrix.  The BMB matrix has invmass(a) + invmass(b)
// on the diagonal and the off-diagonal elements are only non-zero
// when two constraints are connected, then the elements cos x / m(c)
// where x is the angle between the bonds and m(c) is the mass of the
// atom connecting the bonds.
```

```
//
// We write (B M**-1 B') as S (I - A) S where S is a diagonal matrix with
// sqrt (invmass(a) + invmass(b)) where a and b are the atoms of a constraint.
// (I - A) can be inverted with a power series expansion as long as all of
// the eigenvalues are less than 1.  This is true for molecular structures
// with bond length constraints as long as there are not too many small
// rings.

local function invBMBmul [cons, tol, B, rhs, A, maxit]
    local iter;

    local z = rep [0, l_length [cons.atomA, cons.atomB]];
    local I = cons.conI, J = cons.conJ, segJ = cons.conseg;

    if isnull A then
A = cons.conC * (add (apt get [B, [I]] * apt get [B, [J]]));
    endif
    if isnull maxit then
maxit = 4;
    endif

    local idx = cons.conIu; // destinations
    rhs = rhs * cons.diagSinv;
    local sol = rhs; // the solution

    for iter = 1, maxit loop
local delta = s_add [A * rhs[J], segJ];
rhs = put [z, idx, delta];
sol[idx] = sol[idx] + delta;
    until max abs delta <= tol endloop

    return sol * cons.diagSinv;
endfunction

// addTmul calulates x + Ty, T = M**-1 B' (B M**-1 B')**-1

local function addTmul [cons, tol, B, x, y, A]
    y = invBMBmul [cons, tol, B, y, A];
    return addMBmul [cons, x, B, y];
endfunction

// ConstrainPositions calculates a least-action correction to a given set
// of positions.  The corrections occur in two stages.  First, we calculate
// the corrections, c, that satisfy the the constrainted minimization problem
//
// c'c / 2  - a' (B (r + N c) - d) N = inverse(M)
//
// Which sets the correct lengths *but along the old bond directions B*.  Then
// make an additional correction to set the lengths along the new directions.
//
// For the first correction, we must have (c' - a' B N = 0), hence
//
// (B r - d) = B N c = (B N B') N a
//
```

```
// Therefore N**-1 (B N B')**-1 (Br - d) = a; hence
//
// N c = N B' (B N B')**-1 (Br - d)
//
// which is the desired first stage correction.  The second correction
// involves setting the length along the old bond to
//
// p(i) = sqrt (2 * sqr d(i) - l(i))
//
// where l(i) is the bond length after the first correction.  We then set
//
// r_new = (I - T B) r + Tp T = N B' (B N B')**-1
//
// We are given the B matrix (which is based upon the old positions)
//
// SETTLE Constraints
// The SETTLE conctraints are calculated as explained in the SETTLE article.
// Essentially, since water forms a triangle, SETTLE uses the fact that the
// lengths of the triangle are fixed, so the angles must also be fixed. Each
// water molecule after an unconstrained move is moved to a new coordinate
// system where the origin is at the center of mass of the water molecule. The
// constraints are applied here, then the molecule is moved back into Cartesian
// coordinates. Since this is an analytic solution of RATTLE, this algorithm is
// very speedy due to the lack of iterations.

local function ConstrainPositions [cons, tol, cell, B, pos, pos0, dt]
    local iter, dAB, delta;
    local unconsPos = pos, iWc = 0;

    if length cons.lenAB then

local I = cons.conI, J = cons.conJ, segJ = cons.conseg;
local xA = cons.atomA, xB = cons.atomB, d = cons.lenAB;

local A = cons.conC * (add (apt get [B, [I]] * apt get [B, [J]]));

delta = d - Bmul[cons, B, pos];

pos = addTmul [cons, tol, B, pos, delta, A];

for iter = 1, 10 loop
    dAB = norm NearestImage [
apt get[pos,[xA]] - apt get[pos,[xB]], cell
    ];
    delta = sqrt maxE [0, (2 * sqr d - sqr dAB)] - d;
    if max abs delta < tol then break; endif
    pos = addTmul [cons, tol, B, pos, delta, A];
endloop
    endif

    // SETTLE
    if length cons.settle.A then

local mass = inv cons.invmass;
```

```
// Calculate center of mass with the new water coordinates
local wA = cons.settle.A, wB = cons.settle.B, wC = cons.settle.C;
local posA = apt get[pos,[wA]];
local posB = apt get[pos,[wB]];
local posC = apt get[pos,[wC]];
local d1 = (posA*[mass[wA]] + posB*[mass[wB]] + posC*[mass[wC]]) *
    [inv (mass[wA] + mass[wB] + mass[wC])];

// Find ra, rb, rc, theta, phi, psi
local ra = cons.settle.ra;
local rb = cons.settle.rb;
local rc = cons.settle.rc;

// Find new values (X',Y',Z') for coordinates. The new coordinate
// system is defined by the initial (previous move's) plane parallel
// to the X'Y' plane and A1 (new oxygen position) must be in the
// Y'Z' plane.

local pos0A = apt get[pos0,[wA]];
local pos0B = apt get[pos0,[wB]];
local pos0C = apt get[pos0,[wC]];
local AB0 = pos0B - pos0A;
local AC0 = pos0C - pos0A;
local prime;
prime(3) = vcross [AB0, AC0];
prime(1) = vcross [posA,prime(3)];
prime(2) = vcross [prime(3),prime(1)];

// Normalize the new vectors and make the transformation matrix
local normprime = inv app norm prime;
local transmat = [
    prime(1)*[normprime(1)],prime(2)*[normprime(2)],
    prime(3)*[normprime(3)]
];

// get the new coordinates for triangle 0 and triangle 1
// (these are in prime space)
local A1, B1, C1;
AB0 = app add ([AB0]*transmat);
AC0 = app add ([AC0]*transmat);
A1 = app add ([posA-d1]*transmat);
B1 = app add ([posB-d1]*transmat);
C1 = app add ([posC-d1]*transmat);

// Solve for theta, phi, and psi

local theta, phi, psi;
phi.sin = A1(3)/ra;
phi.cos = sqrt (1 - sqr phi.sin);
psi.sin = (B1(3) - C1(3))/(2*rc*phi.cos);
psi.cos = sqrt (1 - sqr psi.sin);

// to get theta, need to solve for alpha, beta, and gamma
```

```
local alpha, beta, gamma, a2,b2,c2;
a2 = [0,ra*phi.cos];
b2 = [-rc*psi.cos,-rb*phi.cos-rc*psi.sin*phi.sin];
c2 = [rc*psi.cos,-rb*phi.cos+rc*psi.sin*phi.sin];
alpha = b2(1)*(AB0(1)-AC0(1)) + b2(2)*AB0(2) + c2(2)*AC0(2);
beta = b2(1)*(AC0(2)-AB0(2)) + b2(2)*AB0(1) + c2(2)*AC0(1);
gamma = B1(2)*AB0(1) - B1(1)*AB0(2) + C1(2)*AC0(1) - C1(1)*AC0(2);
theta.sin = (alpha*gamma - beta*sqrt(sqr alpha + sqr beta - sqr gamma))
    / (sqr alpha + sqr beta);
theta.cos = (gamma - alpha*theta.sin)/beta; // Avoids sqr/sqrt.
//theta.cos = sqrt (1 - sqr theta.sin); //alternate cos theta

// Get the new positions and transform them back to Cartesian
local a3,b3,c3, A3,B3,C3, trtransmat;
a3 = [- a2(2)*theta.sin,a2(2)*theta.cos,A1(3)];
b3 = [b2(1)*theta.cos - b2(2)*theta.sin,
    b2(1)*theta.sin + b2(2)*theta.cos,B1(3)];
c3 = [c2(1)*theta.cos - c2(2)*theta.sin,
    c2(1)*theta.sin + c2(2)*theta.cos,C1(3)];
trtransmat = tr transmat;
A3 = (app add ([a3]*trtransmat)) + d1;
B3 = (app add ([b3]*trtransmat)) + d1;
C3 = (app add ([c3]*trtransmat)) + d1;
local pidx = cons.settle.idx;
local pnew = app cat [
    [A3(1),B3(1),C3(1)],[A3(2),B3(2),C3(2)],[A3(3),B3(3),C3(3)]
];
pos = apt put [pos,[pidx],pnew];
    endif

    // The constraint virial is calculated by r dot (dr_i*m_i/sqr dt),
    // where dr_i is the difference in distance between the constrained
    // and unconstrained positions.
    if (length cons.lenAB or length cons.settle.A) then
iWc = (pos - unconsPos) * [inv cons.invmass] / sqr dt / VAECONV;
iWc = -add app add (pos * iWc);
    endif

    return [pos,iWc];

endfunction

// ConstrainVelocities adjusts the velocities to a given set of positions.
// (I - TB) * vel is the smallest mass-weighted change to the velocities to
// enforce the constraints on the velocities
// We return the time step scaled lagrange multiplers.  (BM**-1B')**-1 B v0.
//
// SETTLE
// The SETTLE velocity constraints simply remove motion along the bond. This
// is also fast because it is non-iterative.

local function ConstrainVelocities [cons, tol, B, vel, pos, dt]
    local iter, y = 0, d, unconsVel = vel, iWc = 0;
    if l_length B then
```

```
local I = cons.conI, J = cons.conJ, segJ = cons.conseg;
local xA = cons.atomA, xB = cons.atomB;
d = cons.lenAB;

local A = cons.conC * (add (apt get [B, [I]] * apt get [B, [J]]));

local delta = -Bmul[cons, B, vel];

vel = addTmul [cons, tol, B, vel, delta, A];

    endif

    if length cons.settle.A then
local wA = cons.settle.A, wB = cons.settle.B, wC = cons.settle.C;
local mass = inv cons.invmass;

// get relative velocities between atoms
local vA0 = apt get[vel,[wA]];
local vB0 = apt get[vel,[wB]];
local vC0 = apt get[vel,[wC]];
local vAB = vB0-vA0, vBC = vC0 - vB0, vCA = vA0 - vC0;

// get unit vectors along the bonds and get bond lengths
local pA = apt get[pos,[wA]];
local pB = apt get[pos,[wB]];
local pC = apt get[pos,[wC]];
local rAB = pB - pA, rBC = pC - pB, rCA = pA - pC;
local eAB = rAB/[norm rAB], eBC = rBC/[norm rBC], eCA = rCA/[norm rCA];
rAB = norm rAB; rBC = norm rBC; rCA = norm rCA;

// find all of the angles using law of cosines
local cos;
cos.A = (sqr rAB + sqr rCA - sqr rBC)/(2*rAB*rCA);
cos.B = (sqr rAB + sqr rBC - sqr rCA)/(2*rAB*rBC);
cos.C = (sqr rBC + sqr rCA - sqr rAB)/(2*rBC*rCA);

// solve for constraint forces in the bond directions
local vab = add (eAB*vAB), vbc = add (eBC*vBC),vca = add (eCA*vCA);
local ma = mass[wA],mb = mass[wB],mc = mass[wC];
local tab, tbc, tca;
d = (2*sqr (ma+mb) + 2*ma*mb*cos.A*cos.B*cos.C - 2*sqr (mb*cos.A)
    - ma*(ma+mb)*(sqr cos.B + sqr cos.C))/(2*mb);
tab = ma/d * (vab * (2*(ma+mb) - ma *sqr cos.C)
    + vbc * (mb*cos.C*cos.A - (ma+mb)*cos.B)
    + vca * (ma*cos.B*cos.C - 2*mb*cos.A));
tbc = (vab * ma*(mb*cos.C*cos.A - (ma+mb) *cos.B)
    + vbc * (sqr(ma+mb) - sqr (mb*cos.A))
    + vca * ma* (mb*cos.B*cos.A - (ma+mb)*cos.C)) / d;
tca = ma/d * (vab * (ma*cos.C*cos.B - 2*mb *cos.A)
    + vbc * (mb*cos.A*cos.B - (ma+mb)*cos.C)
    + vca * (2*(ma+mb) - ma*sqr cos.B));

local vA = vA0 + ([tab]*eAB - [tca]*eCA) / (2*[ma]);
```

```
local vB = vB0 + ([tbc]*eBC - [tab]*eAB) / (2*[mb]);
local vC = vC0 + ([tca]*eCA - [tbc]*eBC) / (2*[mc]);
local vidx = cons.settle.idx;
local vnew = app cat [
    [vA(1),vB(1),vC(1)],[vA(2),vB(2),vC(2)],[vA(3),vB(3),vC(3)]
];
vel = apt put [vel,[vidx],vnew];
    endif

    // The constraint virial is calculated by r dot (dr_i*m_i/sqr dt),
    // where dr_i is the difference in distance between the constrained
    // and unconstrained positions. This idea is carried to the velocities,
    // and dt/2 is used because this correction occurs only for a half-stepped
    // velocity. The new eqn. is r dot ( dv_i*mi/(dt/2) )
    if (length cons.lenAB or length cons.settle.A) then
iWc = (vel - unconsVel) * [inv cons.invmass] / dt / VAECONV * 2;
iWc = -add app add (pos * iWc);
    endif

    return [vel,iWc];
endfunction

// --------------------------------- VELOCITIES -----------------------------

// CenterOfMass calculates sum_i m[i] M**-1 r[i] where m[i] is the mass of
// atom i, r[i] is a 3-vector and M = sum_i m[i].  We are given the
// inverse masses, some of which may be zero (fixed atoms).  In this case
// M**-1 = 0 if any single atom has invmass[i] = 0.  If there are fixed
// atoms then, (m[i] M**-1) == (invmass[i] == 0).

local function CenterOfMass [pos, invmass]
    local mask = (invmass == 0), nfixed = iadd mask;
    if nfixed then return app add (pos || mask) * inv nfixed; endif
    local mass = inv invmass;
    return app add (pos * [mass * inv add mass]);
endfunction

// RemoveLinearVelocity adjusts the given velocities so that the total linear
// momentum is zero.  This is done by finding the additive corrections c[i]
// such that satisfies the following constrained optimiztion.
//
// sum_i m[i] c[i]'c[i] / 2 - a' sum_i m[i] (v[i] + c[i])
//
// If L is the current linear momentum, then c[k] = -L / sum_i m[i] is the
// solution.  This change will not alter the angular momentum.  This change
// will not alter constraints that are independent of rotation and
// translation (e.g., bond lengths).  This code correctly handles the
// case of one or more fixed atoms (invmass = 0 so inv total mass = 0)

local function RemoveLinearVelocity [vel, invmass]
    local mask = (invmass == 0);
    if orE (mask) then return apt select [0, vel, [mask]]; endif
    local mass = inv invmass;
    return vel - app add ([mass] * vel) * invz add mass;
```

```
endfunction

// RemoveAngularVelocity removes the total angular momentum from the
// velocities by adding a correction to each velocity to produce a
// new velocity v[k].  This is done with
//
// A = sum_i m[i] P[i] x V[i] // angular momentum
// v[k] = V[k] + c[k] // new velocities
// c[k] = B x p[k] // P[k] is (p[k] x) matrix
// B = S**-1 A
// S = sum_i m[i] P[i] P[i] // S is symmetric
//
// where P[k] is the p[k] cross operator matrix.  These corrections minimize
// the constrained functions
//
// sum_i 0.5 * m[i] |v[i] - V[i]|**2 - B' sum_i m[i] p[i] x v[i]
//
// Note that this correction will not change the linear momentum since.  The
// bond length constraints and first time derivatives are unaffected.

local function RemoveAngularVelocity [pos, vel, invmass]
    local fixed = (invmass == 0);
    if orE fixed then return apt select [0, vel, [fixed]]; endif

    pos = pos - CenterOfMass [pos, invmass];

    local Mp = [inv invmass] * pos; // mass * pos
    local A = app add vcross [Mp, vel]; // angular momentum

// compute the angular intertial tensor S

    local pm_norm2 = add add (Mp * pos);

    local S = [
[ add (Mp(1) * pos(1)), add (Mp(2) * pos(1)), add (Mp(3) * pos(1)) ],
[ add (Mp(1) * pos(2)), add (Mp(2) * pos(2)), add (Mp(3) * pos(2)) ],
[ add (Mp(1) * pos(3)), add (Mp(2) * pos(3)), add (Mp(3) * pos(3)) ]
    ] - [
[ pm_norm2, 0,        0        ],
[ 0,        pm_norm2, 0        ],
[ 0,        0,        pm_norm2 ]
    ];

// solve for the lagrange multiplier B

    local [U, D, W] = __svd S; // S = U D W'
    (D | abs D <= max abs D * EPS_M) = 0;
    local B = app add (W * [invz D * add (U * A)]); // S**-1 A

    return vel + vcross[B,pos];
endfunction

// ------------------------------ INTEGRATOR ---------------------------------
```

```
// LRC calculates the long range correction for the energy. This will add
// corrections consistent with the potential energy model with two exceptions
// (listed below). The general form is E_lrc = A / Rc**(m-3) - B / Rc**(n-3).
// An assumption made for all cases is that (m>n and Rij>Rc), so the dispersion
// term will contain the largest contribution. The final equation is:
//
// E(r) = vdwWeight * (-2 PI N**2 C6 (m/(m-n)) / (V * (n-3) * Rc**(n-3) ))
//
// where C6 is the average dispersion constant
//
// <C6> = sum_{i,j=1,M} N[i] N[j] / N**2 C6[ij]
//
// Where M is the number of types, and there are N[i] particles of type i.
//
// The LRC to the virial looks like:
//
// W = E_lrc * <n> / (3V)
//
// Where <n> is the average value of the dispersion exponent
//
// <n> = sum_{i,j=1,M} N[i] N[j] / N**2 N[ij]
//
// The two deviations from the potential energy model are:
// 1) if you are using MMFF94, the buffering constants are assumed to be
//    so small that they are essentially zero
// 2) State parameters (Tij in the model, set by aSetState atoms) are not
//    supported

local function LRC atoms
    local atype = aMMType atoms;
    local [idx, mask] = sam atype;

    atoms = atoms[idx] | mask; // representative atom
    local ntype = mtoc mask; // Ni for each type

    local C6 = 0;
    local i;
    local rc = (PotSetup[]).cutoffOff;
    local W_n = 0;

    for i = 1, length atoms loop
local [r,eps,m,n] = (pot_Parm_vdw [atoms(i), atoms])[[2,3,4,5]];
C6 = C6 + (
    ntype(i) * add (ntype * eps * pow [r, n]
      * m * inv (m-n) * inv (n-3)
        * inv pow [rc,(n-3)])
);
W_n = W_n + ntype(i) * add (ntype * n);
    endloop

    [C6,W_n] = [C6,W_n] * invz sqr add ntype;

    return [
    C6 * -(2*PI) * sqr length atoms * (PotSetup[]).vdwWeight,
```

```
    W_n
];
endfunction

// MolecularVolume calculates the volume of the molecule using a simple
// approximation based on pairwise contacts.

local function MolecularVolume [atoms, pos]
    local amask = (
    not (aAtomicNumber atoms == 0 and aBondCount atoms == 0)
and not aInert atoms
    );
    atoms = atoms | amask;
    pos = pos || [amask];

    if not length atoms then return 0; endif

    local r1 = 1.1 * aRadius atoms;
    local prox = prox_open [ 2 * max r1, pos, r1];
    local [seg,idx,d] = prox_find [prox, pos, r1];
    prox_close prox;

    local mask = split[idx,seg] <> x_id atoms;
    local r2 = split [r1[idx], seg] || mask;

    d = split [d, seg] || mask;
    d = maxE [abs (r1 - r2), minE [d, (r1 + r2)]];

    local x1 = (sqr r1 - sqr r2 + sqr d) * invz (d + d);
    local x2 = (sqr r2 - sqr r1 + sqr d) * invz (d + d);

    local vol = (4*PI/3) * cube r1 - maxE [ 0, app add (
  (2/3) * cube r1 - x1 * sqr r1 + (1/3) * cube x1
+ (2/3) * cube r2 - x2 * sqr r2 + (1/3) * cube x2
    )];

    return max [1, add vol];
endfunction

// GetContext calculates the context variables necessary for molecular
// dynmics routines.  We return a tagged vector.

local function GetContext opt
    local chains = Chains[];
    local atoms = cat cAtoms chains;
    local mass = maxE [0, aMass atoms];
    local invmass = select [0, inv mass, aFixed atoms];
    return [
mol: mol_Extract chains,
atoms: atoms,
inv_mass: invmass,
constr: ConstrainInit [opt, atoms, invmass]
    ];
endfunction
```

```
// DynamicsInit creates an empty dynamics database.  The main Dynamics
// routine can then be called to advance the loaded system.

global function DynamicsInit [mdbfile, opt];
    const MDB_FIELDS = [
't','H','U','K','T','P','V','Ua','Ub','Uab','pos','vel','acc'
    ];
    const MDB_FIELDS_D = [ 'xpt', 'box' ];

    write ['DynamicsInit [ {}, {v} ]\n', mdbfile, opt ];

    local mdb = db_Open [mdbfile,'create'];
    db_EnsureField[mdb, 'mol', 'molecule'];
    apt db_EnsureField[mdb, MDB_FIELDS, 'float'];
    apt db_EnsureField[mdb, MDB_FIELDS_D, 'double'];

#if 0
    db_SetEnv [ mdb,
'browse:Dynamics',
'dbvbrowse_DynamicsAnimator(mol,box,pos,t,T,P,V,U)'
    ];
#endif

    SaveRestartSystem mdb;
    SaveRestartPotential mdb;
    SaveRestartOptions [mdb,opt];

    db_Close mdb;
endfunction

// SampleConfiguration calculates the quantities of interest in the sytem
// and writes them to the database.   We are given the instanteous variables
// of the system in the form of a tagged vector.

local function SampleConfiguration [mdb, ctx, ivar, opt]
    if not mdb then return; endif

    local atoms = ctx.atoms;
    aSetPos [atoms, ivar.pos];

// calculate the interaction energies with the MD_INTERACTION
// object set

    local Ua = ivar.U, Ub = 0, Uab = 0;

    local iset = uniq cat oAtoms oGetCollection 'MD_INTERACTION';

    if length iset then
local old_prio = task_prio 0;
local mask = indexof [atoms, iset];
local org_inert = aInert atoms;

aSetInert [atoms, org_inert or mask];
```

```
Ua = first Potential [X:ivar.pos, dX:0];
aSetInert [atoms, org_inert or not mask];
Ub = first Potential [X:ivar.pos, dX:0];
aSetInert [atoms, org_inert];

Uab = ivar.U - Ua - Ub;
task_prio old_prio;
    endif

// calculate the standard thermodynamic quantities

    local box = BoxSize[]; // box dimensions

    local mol = ctx.mol;
    mol(4) = put[mol(4), [MOL_ATOM_X,MOL_ATOM_Y,MOL_ATOM_Z], ivar.pos];

    local data = [
mol: mol,
t: ivar.t,// time
pos: cat ivar.pos,// current positions
vel: cat ivar.vel,// current velocities
acc: cat ivar.acc,// current accelerations
xpt: ivar.xpt,// Npt coordinates
box: box * BoxEnable[],// periodic box (A)
V: ivar.V,// volume
K: ivar.K,// kinetic energy
U: ivar.U,// potential energy
T: ivar.T,// temperature
P: ivar.P / PCONV,// pressure
H: ivar.H,// enthalpy
Ua: Ua,// A-set energy
Ub: Ub,// B-set energy
Uab: Uab // interaction energy
    ];

    local nent = db_nEntries mdb;
    local prevent = mdb_LastEntry mdb;
    local ent = mdb_Append [mdb, data];

    if db_NextEntry[mdb, 0] == prevent then prevent = 0; endif

    if prevent and mod [nent, 1000] <> 0 then
local wdata = [];
if isfalse opt.savePosition     then wdata.pos = []; endif
if isfalse opt.saveVelocity     then wdata.vel = []; endif
if isfalse opt.saveAcceleration then wdata.acc = []; endif
db_Write [mdb, prevent, wdata];
    endif

    db_Flush mdb;
endfunction

// DynamicsReload reloads all dynamics restart data from the mdb file
```

```
global function DynamicsReload mdbfile
    local mdb = db_Open mdbfile;
    local chains = GetRestartSystem mdb;
    GetRestartPotential mdb;
    db_Close mdb;
    if call[] then chains = []; endif
    return chains;
endfunction

// Dynamics advances an existing database system through time by integrating
// the quations of motion.  If there are atoms in the system we assume that
// we are continuing a dynamics run from the given database (or starting
// a new one).  To restart, clear the sytem.

global function Dynamics [
    mdbfile,// trajectory file
    tctl,// upper time limit (inclusive) of simulation (ps)
    Tctl,// desired temperature (Kelvin)
    Pctl,// desired pressure (kPa)
    opt // as in MD_DEFAULTS
]
    local mol_msg = 0; // graphics window message

    local algorithm; // controller algorithm

    local t, T, P; // time, temperature, pressure

    local pos, vel, acc; // atomic pos, vel, force/mass

    local Xt, Xp, Xt_v, Xp_v; // extended system coordinates
    local QT, QP, eta; // extended system coordinates

    local iUex = 0; // extended system potential energy
    local iKex  = 0; // extended system kinetic energy
    local iH    = 0; // total NPA hamiltonian
    local H0    = []; // iUex + iKex at t=0

    local iU = 0; // forcefield potential
    local iK    = 0; // atomic kinetic energy
    local iW = 0; // virial (r.f) from potential

    local iB    = [[],[],[]]; // constraint B matrix (laminated)
    local iBr   = []; // constraint bond lengths (Br)
    local iY = []; // constrain lagrange multipliers
    local iWc   = 0; // constraint virial (r.g)

    local const_P = 1, const_T = 1;

    write ['Dynamics [ {}, {v}, {v}, {v}, {v} ]\n',
mdbfile, tctl, Tctl, Pctl, opt
    ];

    if isnull Tctl or andE (Tctl == 0) then Tctl = 300;  const_T = 0; endif
    if isnull Pctl or andE (Pctl == 0) then Pctl = 0  ;  const_P = 0; endif
```

```
    if isnull tctl then tctl = 1000; endif

    [Tctl,Pctl] = apt resize [[Tctl,Pctl],2];
    Pctl = Pctl * PCONV * const_P;

// Open the trajectory database:
//
// 1) Read out the stored options and add in the new ones
// 2) Read and set the forcefield parameters
// 3) Setup the periodic box from the last data file entry

    local lastent = [];

    local mdb = db_Open mdbfile;

    if db_nEntries mdb > 0 then
lastent = db_Read [mdb, mdb_LastEntry mdb];
    endif

    opt = cat [opt, GetRestartOptions mdb, MD_DEFAULTS];
    opt = opt | indexof [tags opt, tags MD_DEFAULTS];
    opt = opt | m_uniq tags opt;

    if length Chains[] == 0 then // reload molecular data?
Close[force:1];
GetRestartSystem mdb;
GetRestartPotential mdb; // potential + setup
    endif

    local ctx = GetContext opt; // dynamics context
    local cons = ctx.constr; // constrain context


    // This function moves the atoms based on volume change (rscale).
    // Used for NPT dynamics.
    function ScalePositions [rscale, pos]
if rscale == 1 then return pos; endif
local c = CenterOfMass [pos, ctx.inv_mass];
return c + [select [1, rscale, ctx.inv_mass == 0]] * (pos - c);
    endfunction


    // Calculates the KE given a velocity
    function KineticEnergy vel = 0.5 * add (invz ctx.inv_mass * add sqr vel)
/ VAECONV;


    if isnull lastent then
if length tctl == 1 then tctl = prepend[tctl,0]; endif
t = tctl(1); // start time
pos = aPos ctx.atoms;
vel = [];
H0 = [];
Xt = 1;
```

```
Xt_v = 0;
Xp_v = 0;
Xp = select [
    mul BoxSize[], MolecularVolume [ctx.atoms,pos], orE BoxEnable[]
];
eta = 0;
    else
t = lastent.t; // continue time (ps)
if length tctl == 1 then tctl = prepend[tctl,t]; endif

vel = split [lastent.vel, length lastent.vel / 3];
pos = split [lastent.pos, length lastent.pos / 3];

aSetPos [ctx.atoms, pos];

if orE (lastent.box > 0) then BoxSize lastent.box; endif
BoxEnable (lastent.box > 0);
[Xt,Xt_v,Xp,Xp_v,eta,H0] = lastent.xpt;
    endif

    lastent = [];
    local box = orE BoxEnable [];
    local celldim = select [
BoxSize[], REAL_MAX, BoxEnable[] and BoxSize[]>=1
    ];

    // Make sure cutoff is properly setup in case there is a box. This uses
    // a 2 Angstrom spline, which *should* be adequate
    if box then
PotSetup[cutoffOn: (min (celldim/2) - 2),
    cutoffOff: (min (celldim/2))];
    endif

    local conserve_lmom = not orE (ctx.inv_mass <= 0);
    local conserve_amom = not orE (ctx.inv_mass <= 0) and not box;

// determine how long the simulation is to run

    local tmax  = tctl(2); // upper time limit
    local dt    = opt.dt; // time step (ps)
    local saveI = max [1,ceil(opt.sample/dt)]; // save period (iterations)
    local iter  = round (t / dt); // current iteration number
    local maxit = ceil (tmax / dt); // upper iteration limit

    local dof = max [1,// calculated degrees of freedom
  3 * add (ctx.inv_mass > 0) // moving atoms
- 3 * conserve_lmom // translation
- 3 * conserve_amom // rotation
- add length cons.lenAB // less LINCS constraints
- 3 * add length cons.settle.A // less SETTLE constraints
    ];

    T = max [20, Tctl(1) + sub Tctl * (t - tctl(1)) * invz sub tctl];
    P = Pctl(1) + sub Pctl * (t - tctl(1)) * invz sub tctl;
```

```
    local gkT = dof * KBOLTZ * T;

    QT = (4 * sqr PI) * invz (sqr opt.QT * gkT);
    QP = (4 * sqr PI) * invz (sqr opt.QP * (dof+3) * KBOLTZ * T);

// save restart options

    SaveRestartSystem mdb;
    SaveRestartPotential mdb;
    SaveRestartOptions [mdb, opt];

 // set up the current state

    local tol = max [1e-14, min [opt.constraintTol, cube dt]];

// potential calcualtes the potential energy of the sytem
// We add long-range corrections for the potential and
// pressure calculations.

    local dU_scale = -ctx.inv_mass * VAECONV;
    local E_lrc = 0; // kcal/mol
    local W_n = 0; // required for virial LRC

    // The long-range corrections should only be used for a box. If a box is
    // not used, the user should have set a reasonably large cutoff distance.
    if (PotSetup[]).cutoffEnable and box then
[E_lrc,W_n] = LRC ctx.atoms;
    endif


    // Returns the PE, acceleration, and virial for given positions. The PE
    // and virial have long-range corrections included.
    function potential [pos, V]
local [U,dU,W] = Potential[X: pos, W:1];
W = add W;

U = U + (E_lrc / V);
W = W - (E_lrc / V) * W_n; // W = dU/dV * 3V

return [U, dU * [dU_scale],W];
    endfunction


    [iB,iBr] = Bmatrix [cons, celldim, pos];
    [pos,iWc] = ConstrainPositions [cons, tol, celldim, iB, pos, pos, dt];

    [iU,acc,iW] = potential [pos, Xp]; // acc = M**-1 f
    acc = apt select [0, acc, [ctx.inv_mass <= 0]];
    [iB,iBr] = Bmatrix [cons, celldim, pos];

    Xp_v = Xp_v * const_P;
    Xt_v = Xt_v * const_T;
```

```
    // Assign initial velocities
    if isnull vel then
vel = randN rep [
    sqrt(VAECONV*(opt.T0*KBOLTZ)*(ctx.inv_mass*not aInert ctx.atoms)),
    3
];

// need to adjust velocities for constraints
// these options differ only in the initial constraint virial
//local temp_pos = pos - vel * dt;
//[temp_pos,iWc] =
//    ConstrainPositions [cons, tol, celldim, iB, temp_pos, pos, dt];
//vel = (pos - temp_pos)/dt;

// OR
[vel,iWc] = ConstrainVelocities [cons, tol, iB, vel, pos, dt];

vel = apt select [0, vel, [ctx.inv_mass <= 0]];

if conserve_lmom then
    vel = RemoveLinearVelocity [vel, ctx.inv_mass];
endif
if conserve_lmom then
    vel = RemoveAngularVelocity [pos, vel, ctx.inv_mass];
endif
vel = vel * sqrt (
    (dof*KBOLTZ*opt.T0) * invz (2 * KineticEnergy vel)
);
    else
[vel,iWc] = ConstrainVelocities [cons, tol, iB, vel, pos, dt];
    endif

    iK = KineticEnergy vel;

    // do algorithm specific stuff. This means set up the proper
    // equations for iUex and iKex (conserved quantity in each algorithm)
    if opt.algorithm === 'NPA' then

iUex = iU + gkT * log Xt + (P * Xp);
iKex = iK + 0.5 * (sqr Xp_v * QP + sqr Xt_v * QT);

    elseif opt.algorithm === 'NH' then

iUex = iU + gkT * eta;
iKex = iK + 0.5 * sqr Xt_v / QT;
    endif


    if isnull H0 then
H0 = iKex + iUex;
    elseif abs (iKex + iUex - H0) > 10 then
H0 = iKex + iUex;
    endif
```

```
    iUex = iUex - H0;


    // IntegrateNPA advances the system with the Sturgeon & Laird
    // integration for the Nose-Poincare-Anderson equations of motion
    function IntegrateNPA h
local hh = 0.5 * h;
gkT = dof * KBOLTZ * T;
local pos0 = pos;

    // advance the velocities for one half step
vel = vel + hh * acc;
[vel,iWc] = ConstrainVelocities [cons, tol, iB, vel, pos, h];

local hK = KineticEnergy vel;
local hP = (2*hK - (iW + iWc)) / (3*Xp);

Xp_v = Xp_v + const_P * (hh * Xt) * (hP - P);

local C = -Xt_v + hh * (iUex + gkT - hK + sqr Xp_v * QP/2);
Xt_v = const_T * (sqrt (1 - h*C*QT) - 1) / (hh*QT);

local R = cbrt Xp;
local s = div1p1m (Xt_v * (hh*QT));

Xp = Xp + hh * Xt * (1 + s) * Xp_v * QP;
Xt = Xt * s;
  R = cbrt Xp / R;

pos = ScalePositions [R, pos + hh * (1 + inv (sqr R * s)) * vel];
pos = first ConstrainPositions [cons, tol, celldim, iB, pos, pos0, dt];
if box and R <> 1 then
    BoxSize (celldim = celldim * R);
endif

vel = (pos - ScalePositions[R,pos0]) / hh / (1 + inv (sqr R * s));

[iU,acc,iW] = potential [pos, Xp];
[iB,iBr] = Bmatrix [cons, celldim, pos];

iUex = (iU + gkT * log Xt + P * Xp) - H0;

hK = hK / sqr (R * s);
vel = vel * inv (s * R) + hh * acc;
[vel,iWc] = ConstrainVelocities [cons, tol, iB, vel, pos, h];

hP = (2*hK - (iW + iWc)) / (3*Xp);

Xt_v = Xt_v - const_T * hh*(iUex+gkT-hK+sqr Xp_v *QP/2+sqr Xt_v *QT/2);
Xp_v = Xp_v + const_P * (hh * Xt) * (hP - P);

iK   = KineticEnergy vel;
iKex = iK + 0.5 * (sqr Xp_v * QP + sqr Xt_v * QT);
```

```
    endfunction


    // IntegrateNHSV integrates the Nose-Hoover equations of motion
    // using the Stormer-Verlet algorithm
    function IntegrateNHSV h
local hh = 0.5 * h;
local pos0 = pos;
gkT = dof * KBOLTZ * T;

vel = (vel + hh * acc) / (1 + hh * Xt_v);

eta = eta + const_T * hh * Xt_v;

pos = pos + h * vel;
pos = first ConstrainPositions [cons, tol, celldim, iB, pos, pos0, dt];

vel = (pos - pos0)/h;

iK = KineticEnergy vel;
Xt_v = Xt_v + const_T * h * QT * (2 * iK - gkT);
eta = eta + const_T * hh * Xt_v;

[iU,acc,iW] = potential [pos,Xp];
[iB,iBr] = Bmatrix [cons, celldim, pos];

vel = vel + 0.5 * h * (acc - Xt_v * vel);
[vel,iWc] = ConstrainVelocities [cons, tol, iB, vel, pos, h];

iK   = KineticEnergy vel;

iKex = iK + sqr Xt_v / QT / 2;
iUex = iU + gkT * eta - H0;

    endfunction


// Display manages the display in the MOE window (or printout
// to the cli if in batch).  NOTE: WE ASSUME THAT ALL VARIBLES
// OF INTEREST HAVE BEEN CALCULATED.

    local last_display = clock[]; // time of last display

    function Display []
const FRAMERATE = MOE_BATCH * (60*60) + not MOE_BATCH * (1/12);
local iter_time = clock[];

if (iter_time - last_display) < FRAMERATE then
    return;
endif

last_display = iter_time;
aSetPos [ctx.atoms, pos];
if box then BoxSize celldim; endif
```

```
local iV = Xp; // volume
local iT = 2*iK * invz (KBOLTZ * dof); // temperature
local iP = (2*iK - (iW + iWc)) / (3*iV); // pressure
local iH = (iKex + iUex) * Xt; // NPA hamiltonian

local msg = swrite [
    '[MD] t:{n:12.3f}, H:{n:12.6f}, '
    'T:{n:6.9g}, P:{n:8.9g}, V:{n:7.9g}, U:{n:12.3f}\n'
    '       {t:12}    {t:12}  ',
    t, (iH), round (iT),
    round (iP/PCONV), round (iV), (iU)
];

if DISPLAY_ERRORS then
    local com = CenterOfMass [pos, ctx.inv_mass];

    msg = cat [msg, swrite [
'\n     m: {n:14.6e} {n:14.6e}'
'\n     e: {n:14.6e} {n:14.6e}'
,
max [0, max app add ([invz ctx.inv_mass] * vel)],
max [0, max app add vcross[pos-com,[invz ctx.inv_mass]*vel]],
max [0, max abs (iBr - cons.lenAB)],// G
max [0, max abs Bmul [cons, iB, vel]] // dG/dt
    ]];
endif

if not MOE_BATCH then
    mol_msg = Message[mol_msg, token msg];
else
    write ['{}\n', token msg];
endif
    endfunction

// Sample calcualtes the various quantities of interest for
// writing to the trajectory database.  NOTE: WE ASSUME
// ALL INSTANTANEOUS QUANTITIES HAVE BEEN CALCULATED

    function Sample []
local iV = Xp; // volume
local iT = 2*iK * invz (KBOLTZ * dof); // temperature
local iP = (2*iK - (iW + iWc)) / (3*iV); // pressure
local iH = (iKex + iUex) * Xt; // NPA hamiltonian
local ivar = [
    t: t,
    pos: pos,
    vel: vel,
    acc: acc,
    H: iH,
    U: iU,
    K: iK,
    V: iV,
    P: iP,
```

```
    T: iT,
    xpt: [Xt,Xt_v,Xp,Xp_v,eta,H0]
];
SampleConfiguration [mdb, ctx, ivar, opt];
    endfunction

    if db_nEntries mdb == 0 then // first time?
Sample[];
    endif

// ADVANCE THE SYSTEM IN TIME

    loop
T = max [20, Tctl(1) + sub Tctl * (t - tctl(1)) * invz sub tctl];
P = Pctl(1) + sub Pctl * (t - tctl(1)) * invz sub tctl;

QT = (4 * sqr PI) * invz (sqr opt.QT * gkT);
QP = (4 * sqr PI) * invz (sqr opt.QP * 9 * KBOLTZ * T);

if orE not finite [T, P, iUex, iKex, iW, iWc] then
    print [T, P, iUex, iKex, iW, iWc];
    exit 'Numerical integration failure.  The Time Step\n'
 'is possibly too large or there is too much strain\n'
 'energy in the system.'
    ;
endif

Display [];

t = (iter  = iter + 1) * dt;

if conserve_lmom then
    vel = RemoveLinearVelocity [vel, ctx.inv_mass];
endif
if conserve_amom then
    vel = RemoveAngularVelocity [pos, vel, ctx.inv_mass];
endif

if opt.algorithm === 'NPA' then
    IntegrateNPA dt;
elseif opt.algorithm === 'NH' then
    IntegrateNHSV dt;
        endif

if mod [iter, saveI] == 0 then Sample[]; endif

if (iter >= maxit) then break; endif
    endloop

    Message[mol_msg,[]];
    db_Close mdb;

    return min [t, tmax];
endfunction
```

APPENDIX D

ACCESSIBILITY CODE

```
global function Accessibility mdb

    mdb = db_Open mdb;

    db_EnsureField[mdb, 'Probe_Size', 'float'];
    db_EnsureField[mdb, 'Accessible_Tags', 'float'];
    db_EnsureField[mdb, 'Accessibility', 'float'];

    local probe = [1.4, 2, 3, 4, 5, 7, 10, 15, 20];

    local EntryKeys = db_Entries mdb;

        // loop over each database entry

    local i, j, accessible_atoms, accessibility, atoms, amask, mpos, rad, mr2;

        // The positions and radii will need to be scaled by a factor of M to
        // maintain the accuracy of log10M digits (or, more precisely,
        // log2M bits) after the decimal point:
    const M = 512;

        // Setup various variables to perform certain calcs once
    mol_Create cat db_ReadFields[mdb, EntryKeys(1), 'mol'];

        // ignore hydrogens (except those that are "catalyst sites")
    atoms = Atoms[] | (aElement Atoms[] <> 'H' or aName Atoms[]=='XYZ');
    amask = not m_diff[Atoms[],atoms];

    rad = rep[el_VDW_Radius aElement atoms, length probe] + probe;
    mr2 = sqr (M*rad);

    oDestroy Chains[];

    for i=1, length EntryKeys loop

            // export current conformation to MOE from database

        mol_Create cat db_ReadFields[mdb, EntryKeys(i), 'mol'];

        atoms = Atoms[] | amask;

            // Start by getting the positions and radii of the atoms:

        mpos = M * aPos atoms;                                  // atom centers
```

```
        for j=1, length probe loop

            local [faces, links] = ialpha_complex3 [mpos, mr2(j)];

            // The resulting face centers and weights must then be scaled down,
            // again by a factor of M:

            local size = faces(1);                        // #of vertices per face
            local weight = inv sqr M * faces(2);        // face weight
            local centers = inv M * drop [faces, 2];        // face center

            // Calculate the facet number, smx, of all simplices that have
            // nonnegative weight and collect all facet numbers, vtx, of
            // vertices that participate in such simplices:

            local [f, g, attached] = links;                          // face links
            local smx = x_pack (size == max size and weight >= 0.0);
            local vtx = f[x_join [g,smx]];        // OR: vtx = f | m_join [g,smx];

            // Note that while smx is a sorted array of unique numbers, vtx,
            // in general, is unsorted and contains duplicates. The number of
            // duplicates of each vertex is equal to the number of simplices
            // of nonnegative weight that contain that vertex.

            // Now construct a mask, mask, to extract the accessible atoms
            // with atoms|mask:

            local mask = m_join [x_id atoms, vtx];

            accessible_atoms(j) = atoms | (mask and (aName atoms=='XYZ'));

            accessibility(j) = length accessible_atoms(j) / length Residues[];
            accessible_atoms(j) = m_join[Residues[],oParent accessible_atoms(j)];

        endloop

        db_Write [mdb, EntryKeys(i),
            [Probe_Size:probe, Accessible_Tags:cat accessible_atoms,
            Accessibility:accessibility ]];

            // Destroy current molecule to clear MOE for next one

        oDestroy Chains[];

    endloop

    db_Close mdb;

endfunction
```

APPENDIX E

DESCRIPTION OF AND DATA FROM *AB INITIO* CALCULATIONS FOR THE CO-SALEN COMPLEX

The following information was supplied by Tait Takatani to describe the calculations used to obtain the Co-salen structure through *ab initio* quantum calculations. This work was performed by Tait Takatani under the supervision of Prof. David Sherrill in the School of Chemistry and Biochemistry at the Georgia Institute of Technology.

Density functional theory (DFT), executed with the Jaguar suite of programs (Jaguar 5.5, Schrödinger, LLC, Portland OR), was used to compute the optimized singlet state structure of the Co(III) metal-salen model with the BP86 functional[115] and the LACVP* basis set (LAN2DZ for Cobalt and 6-31G* for all other atoms). A frequency computation at the converged geometry was performed to ensure the structure corresponded to a potential energy minimum. Atomic charges were computed by fitting to the DFT electrostatic potential[116, 117, 118].

Cartesian coordinates (x, y, z) for the optimized geometry of the singlet state for the reduced Co-salen complex with the acetoxy counter-ion:

```
Geometry (Angstroms)
 N     -1.9680103790   -0.1242184687    0.0641556408
 N     -0.1630708440    1.7239565025   -0.1441722134
 O      0.0266303159   -1.9346544090    0.7116227391
 O      1.6539633444   -0.0632622838    0.8771604198
 C     -2.2568918387   -2.5449426969    0.1295055377
 C     -0.8910016489   -2.8402356126    0.5025981065
 C      2.1157840659    2.2942729254    0.5059429235
 C      2.4711465418    0.9590039689    0.9216305420
 C     -2.7187900854   -1.1954455744   -0.0065021229
```

C	0.8124336477	2.5906539387	-0.0042068365
H	-3.8023311520	-1.0664323773	-0.1573382197
H	0.6221541310	3.6384788010	-0.2844071064
C	-2.5224268816	1.2454955344	-0.0272647644
H	-2.5713233153	1.6393377875	1.0117731728
C	-1.4540982433	2.0762830637	-0.7744717962
H	-1.3999529995	1.6770989686	-1.8093913773
Co	-0.0673742758	-0.1317415226	0.1839442789
C	-3.9096624430	1.4016715426	-0.6741148833
H	-3.8898280186	0.9659799552	-1.6931716942
H	-4.6734313699	0.8484858606	-0.0966823787
C	-4.2979255826	2.8950078190	-0.7394622506
C	-1.8347023626	3.5651899438	-0.8260947425
H	-1.8352571628	3.9841861807	0.2007475685
H	-1.0903243936	4.1358676751	-1.4116502478
C	-3.2297994118	3.7320627246	-1.4675414733
H	-3.5132968500	4.8006968999	-1.4668901279
H	-3.1799296478	3.4200597184	-2.5301881138
H	-5.2773553698	3.0019975886	-1.2411068588
H	-4.4286602474	3.2822709358	0.2913041313
C	4.6864829805	1.8024968629	1.5466925995
C	3.7803719304	0.7529983062	1.4525174359
H	5.6899308460	1.6113998439	1.9461094695
C	3.0694089691	3.3430222368	0.6128522484
C	4.3406869946	3.1105732343	1.1229442523
C	-1.4803358521	-5.2256168407	0.5210659158
C	-0.5435282971	-4.2152918994	0.6966363775
H	-1.1769195631	-6.2690705757	0.6713387636
C	-3.1891344086	-3.6077984462	-0.0367824088
C	-2.8179571803	-4.9332669054	0.1480116909
H	2.7833068731	4.3490194169	0.2782688737
H	-4.2214700709	-3.3577320900	-0.3160081981
C	1.0944409479	-0.4133520256	-2.3358991614
O	-0.0372353784	-0.5529798463	-1.6448305876
O	2.1609131769	0.0650050900	-1.9555155712
H	-3.5450629824	-5.7404733310	0.0131125493
H	0.4905432274	-4.4334681632	0.9827508436
H	5.0687360895	3.9253686869	1.1939359351
H	4.0418993660	-0.2632961534	1.7637004638
C	0.9007044395	-0.9752509773	-3.7571036302
H	1.7682697964	-0.6942887920	-4.3730155106
H	0.8256342531	-2.0756973252	-3.7094735777
H	-0.0295634744	-0.5985687821	-4.2142095045

Electrostatic point charges for the optimized geometry of the singlet state for the reduced Co-salen complex with the acetoxy counter-ion:

Atomic charges from electrostatic potential:

Atom	N1	N2	O3	O4	C5
Charge	-0.62937	-0.65036	-0.53077	-0.51227	-0.18083
Atom	C6	C7	C8	C9	C10
Charge	0.50648	-0.13198	0.47351	0.22486	0.17309
Atom	H11	H12	C13	H14	C15
Charge	0.07110	0.08720	0.35996	-0.01789	0.49821
Atom	H16	Co17	C18	H19	H20
Charge	-0.00819	1.00709	-0.09596	0.03560	0.04697
Atom	C21	C22	H23	H24	C25
Charge	-0.12096	-0.26740	0.06680	0.08468	-0.03260
Atom	H26	H27	H28	H29	C30
Charge	0.05020	0.02177	0.05492	0.04107	-0.08259
Atom	C31	H32	C33	C34	C35
Charge	-0.26362	0.12233	-0.13705	-0.15224	-0.07453
Atom	C36	H37	C38	C39	H40
Charge	-0.29241	0.12076	-0.13286	-0.16470	0.11943
Atom	H41	C42	O43	O44	H45
Charge	0.11725	0.70934	-0.49516	-0.47585	0.12379
Atom	H46	H47	H48	C49	H50
Charge	0.16219	0.12297	0.16175	-0.59829	0.14905
Atom	H51	H52			
Charge	0.17481	0.16069			

sum of atomic charges: 0.000000

REFERENCES

[1] Knözinger, H. and Kochloefl, K., *Ullmann's Encyclopedia of Industrial Chemistry Online*, ch. Heterogeneous Catalysis and Solid Catalysts. Wiley-Interscience, 2003.

[2] Behr, A., *Ullmann's Encyclopedia of Industrial Chemistry Online*, ch. Organometallic Compounds and Homogeneous Catalysis. Wiley-Interscience, 2000.

[3] Gates, B. C., *Kirk-Othmer Encyclopedia of Chemical Technology Online*, vol. 5, ch. Catalysis, pp. 200–254. Wiley-Interscience, 2002.

[4] Clark, J. H. and Macquarrie, D. J., *Kirk-Othmer Encyclopedia of Chemical Technology Online*, vol. 5, ch. Supported Catalysts, pp. 322–344. Wiley-Interscience, 2002.

[5] Thomas, J. M., "Design, synthesis, and in situ characterization of new solid catalysts," *Angewandte Chemie-International Edition*, vol. 38, pp. 3588–3628, 1999.

[6] Bergbreiter, D. E., "Using soluble polymers to recover catalysts and ligands," *Chemical Reviews*, vol. 102, pp. 3345–3384, 2002.

[7] Haag, R. and Roller, S., "Polymeric supports for the immobilisation of catalysts," *Topics in Current Chemistry*, vol. 242, pp. 1–42, 2004.

[8] Floudas, C. A., Aggarwal, A., and Ciric, A. R., "Global optimum search for nonconvex nlp and minlp problems," *Computers and Chemical Engineering*, vol. 13, pp. 1117–1132, 1989.

[9] Frenkel, D. and Smit, B., *Understanding Molecular Simulation*. Academic Press, 2nd ed., 2002.

[10] Flory, P. J., *Statistical Mechanics of Chain Molecules*. Wiley-Interscience, 1969.

[11] Mattice, W. L. and Suter, U. W., *Conformational Theory of Large Molecules: The Rotational Isomeric State Model in Macromolecular Systems*. John Wiley & Sons, 1994.

[12] Karafilidis, C., Hermann, H., Rufinska, A., Gabor, B., Mynott, R. J., Breitenbruch, G., Weidenthaler, C., Rust, J., Joppek, W., Brookhart, M. S., Thiel, W., and Fink, G., "Metallocene-catalyzed c7-linkage in the hydrooligomerization of norbornene by sigma-bond metathesis: Insight into the microstructure of polynorbornene," *Angewandte Chemie-International Edition*, vol. 43, no. 18, pp. 2444–2446, 2004.

[13] Ivin, K. J. and Mol, J. C., *Olefin Metathesis and Metathesis Polymerization*. Academic Press, San Diego, 1997. p. 407.

[14] Gaylord, N. G., Deshpande, A. B., Mandal, B. M., and Martan, M., "2,3- and 2,7-bicyclo[2.2.1]hept-2-ene. preparation and structures of polynorbornene," *Journal of Macromolecular Science and Chemistry*, vol. 11, pp. 1053–1070, 1977.

[15] Goodall, B., Benedikt, G., McIntosh, L., Banres, D., and Rhodes, L., "Novel heat-resistant cyclic olefin polymers made using single component nickel and palladium catalysts," *Polymeric Materials Science and Engineering*, vol. 76, pp. 56–57, 1997.

[16] Kaminsky, W. and Noll, A., "Copolymerization of norbornene and ethene with homogeneous zirconocenes methylaluminoxane catalysts," *Polymer Bulletin*, vol. 31, no. 2, pp. 175–182, 1993.

[17] Grove, N. R., Zhao, Q., Kohl, P. A., Bidstrup-Allen, S. A., Shick, R. A., Goodall, B. L., McIntosh, L. H., and Jayaraman, S. ., "New olefinic interlevel dielectric materials for microelectronics," *Advancing Microelectronics*, vol. 23, pp. 16–18, 1996.

[18] Czornyj, G., Asano, M., Beliveau, R. L., Garrou, P., Hiramoto, H., Ikeda, A., Kreuz, J. A., and Rohde, O., *Microelectronics Packaging Handbook: Part II*, p. 548. Chapman and Hall, 1997.

[19] Goodall, B., Benedikt, G., McIntosh, L., Banres, D., and Rhodes, L., "Addition polymers derived form norbornene-functional monomers and process therefor," International Patent WO/1995/014048, The B.F. Goodrich Company, 1995.

[20] Hoskins, T., Chung, W. J., Agrawal, A., Ludovice, P. J., Henderson, C. L., Seger, L. D., Rhodes, L., and Shick, R. A., "Bis(trifluoromethyl)carbinol-substituted polynorbornenes: dissolution behavior," *Macromolecules*, vol. 37, pp. 4512–4518, 2004.

[21] Ahmed, S., Ludovice, P. J., and Kohl, P., "Microstructure of 2,3 erythro di-isotactic polynorbornene," *Computational and Theoretical Polymer Science*, vol. 10, pp. 221–233, 2000.

[22] Callander, D. B., *Molecular modeling of polymer free volume distribution*. PhD thesis, Georgia Institute of Technology, 2005.

[23] Chung, W. J., *Molecular modeling of structure-property relationship for palladium catalyzed poly(norbornene) and its derivatives*. PhD thesis, Georgia Institute of Technology, 2003.

[24] Wilks, B. R., Chung, W. J., Ludovice, P. J., Rezac, M. E., Meakin, P., and Hill, A. J., "Structural and free-volume analysis for alkyl-substituted palladium-catalyzed poly(norbornene): A combined experimental and monte carlo investigation," *Journal of Polymer Science Part B-Polymer Physics*, vol. 44, pp. 215–233, Jan. 2006.

[25] Kilpatrick, J. E., Pitzer, K. S., and Spitzer, R., "The thermodynamics and molecular structure of cyclopentane," *Journal of the American Chemical Society*, vol. 69, no. 10, pp. 2483–2488, 1947.

[26] Sundaralingam, M., "Stereochemistry of nucleic acids and their constituents. iv. allowed and preferred conformations of nucleosides, nucleoside mono-, di-, tri-, tetraphosphates, nucleic acids and polynucleotides," *Biopolymers*, vol. 7, pp. 821–860, 1969.

[27] Altona, C. and Sundaralingam, M., "Conformational-analysis of sugar ring in nucleosides and nucleotides - new description using concept of pseudorotation," *Journal of the American Chemical Society*, vol. 94, no. 23, pp. 8205–8212, 1972.

[28] Altona, C., "Dna, the versatile vector of life: two-dimensional nmr studies," *Journal of Molecular Structure*, vol. 141, pp. 109–125, 1986.

[29] Harvey, S. C. and Prabhakaran, M., "Ribose puckeringstructure, dynamics, energetics, and the pseudorotation cycle," *Journal of the American Chemical Society*, vol. 108, no. 20, pp. 6128–6136, 1986.

[30] Cui, W. L., Li, F. B., and Allinger, N. L., "Simulation of conformational dynamics with the mm3 force-fieldthe pseudorotation of cyclopentane," *Journal of the American Chemical Society*, vol. 115, no. 7, pp. 2943–2951, 1993.

[31] Cornell, W. D., Ha, M. P., Sun, Y., and Kollman, P. A., "Application of a simple diagonal force field to the simulation of cyclopentane conformational dynamics," *Journal of Computational Chemistry*, vol. 17, no. 13, pp. 1541–1548, 1996.

[32] Dobado, J. A., Molina, J. M., and Espinosa, M. R., "A comparative molecular mechanics, semiempirical and ab-initio study of saturated 5-membered rings," *Theochem-Journal of Molecular Structure*, vol. 109, pp. 205–212, 1994.

[33] Wu, A., Cremer, D., Auer, A. A., and Gauss, J., "Extension of the karplus relationship for nmr spin-spin coupling constants to nonplanar ring systems: Pseudorotation of cyclopentane," *Journal of Physical Chemistry A*, vol. 106, no. 4, pp. 657–667, 2002.

[34] Verlet, L., "Computer experiments on classical fluids .i. thermodynamical properties of lennard-jones molecules," *Physical Review*, vol. 159, no. 1, p. 98, 1967.

[35] Gear, C. W., *Numerical Initial Value Problems In Ordinary Differential Equations*. Prentice-Hall, 1971.

[36] Allen, M. P. and Tildesley, D. J., *Computer Simulation of Liquids*. Oxford University Press, 1987.

[37] Hockney, R. W., "The potential calculation and some applications," *Methods in Computational Physics*, vol. 9, pp. 136–211, 1970.

[38] Swope, W. C., Andersen, H. C., Berens, P. H., and Wilson, K. R., "A computer-simulation method for the calculation of equilibrium-constants for the formation of physical clusters of moleculesapplication to small water clusters," *Journal of Chemical Physics*, vol. 76, no. 1, pp. 637–649, 1982.

[39] Berendsen, H. J. C., Postma, J. P. M., Vangunsteren, W. F., Dinola, A., and Haak, J. R., "Molecular-dynamics with coupling to an external bath," *Journal of Chemical Physics*, vol. 81, no. 8, pp. 3684–3690, 1984.

[40] Holian, B. L., Degroot, A. J., Hoover, W. G., and Hoover, C. G., "Time-reversible equilibrium and nonequilibrium isothermal-isobaric simulations with centered-difference stoermer algorithms," *Physical Review A*, vol. 41, no. 8, pp. 4552–4553, 1990.

[41] Bond, S. D., Leimkuhler, B. J., and Laird, B. B., "The nose-poincare method for constant temperature molecular dynamics," *Journal of Computational Physics*, vol. 151, no. 1, pp. 114–134, 1999.

[42] Hunenberger, P., "Thermostat algorithms for molecular dynamics simulations," *Advanced Computer Simulation Approaches for Soft Matter Sciences I*, vol. 173, pp. 105–149, 2005.

[43] Pitzer, K. S. and Donath, W. E., "Conformations and strain energy of cyclopentane and its derivatives," *Journal of the American Chemical Society*, vol. 81, pp. 3213–3218, 1959.

[44] Durig, J. R. and Wertz, D. W., "Vibrational spectra and structure of small-ring compounds .x. spectroscopic evidence for pseudorotation in cyclopentane," *Journal of Chemical Physics*, vol. 49, no. 5, pp. 2118–2121, 1968.

[45] Carreira, L. A., Jiang, G. J., Person, W. B., and Willis, J. N., "Spectroscopic determination of barrier to planarity in cyclopentane," *Journal of Chemical Physics*, vol. 56, no. 4, pp. 1440–1443, 1972.

[46] Bauman, L. E. and Laane, J., "Pseudorotation of cyclopentane and its deuteriated derivatives," *Journal of Physical Chemistry*, vol. 92, no. 5, pp. 1040–1051, 1988.

[47] Adams, W. J., Geise, H. J., and Bartell, L. S., "Structure, equilibrium conformation, and pseudorotation in cyclopentane an electron diffraction study," *Journal of the American Chemical Society*, vol. 92, no. 17, pp. 5013–5019, 1970.

[48] Poupko, R., Luz, Z., and Zimmermann, H., "Pseudorotation in cyclopentanean experimental-determination of the puckering amplitude by nmr in oriented solvents," *Journal of the American Chemical Society*, vol. 104, no. 20, pp. 5307–5314, 1982.

[49] Han, S. J. and Kang, Y. K., "A pseudorotation model and ring-puckering of cyclopentane," *Theochem-Journal of Molecular Structure*, vol. 362, no. 2, pp. 243–255, 1996.

[50] Rosas, R. L., Cooper, C., and Laane, J., "Evaluation of molecular mechanics methods for the calculation of the barriers to planarity and pseudorotation of small ring molecules," *Journal of Physical Chemistry*, vol. 94, no. 5, pp. 1830–1836, 1990.

[51] Ikeda, T., Lord, R. C., Malloy, Jr., T. B., and Ueda, T., "Far-infrared spectra of ring compounds. viii. the effect of a finite central barrier on pseudorotation in five-membered rings," *Journal of Chemical Physics*, vol. 56, no. 4, pp. 1434–1439, 1972.

[52] Henry, B. R., Hung, I. F., Macphail, R. A., and Strauss, H. L., "Local mode description of the ch-stretching overtone spectra of the cycloalkanes and cycloalkenes," *Journal of the American Chemical Society*, vol. 102, no. 2, pp. 515–519, 1980.

[53] Rao, S. T., Westhof, E., and Sundaralingam, M., "Exact method for the calculation of pseudorotation parameters p, tau-m and their errors: a comparison of the altona-sundaralingam and cremer-pople treatment of puckering of 5-membered rings," *Acta Crystallographica Section A*, vol. 37, pp. 421–425, 1981.

[54] Dembo, R. S. and Steihaug, T., "Truncated-newton algorithms for large-scale unconstrained optimization," *Mathematical Programming*, vol. 26, pp. 190–212, 1983.

[55] Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., and Kollman, P. A., "A 2nd generation force-field for the simulation of proteins, nucleic-acids, and organic-molecules," *Journal of the American Chemical Society*, vol. 117, no. 19, pp. 5179–5197, 1995.

[56] Halgren, T. A., "Merck molecular force field. i. basis, form, scope, parameterization, and performance of mmff94," *Journal of Computational Chemistry*, vol. 17, pp. 490–519, 1996.

[57] Fabricius, J., Engelsen, S. B., and Rasmussen, K., "The consistent force field .5. pef95sac: Optimized potential energy function for alcohols and carbohydrates," *Journal of Carbohydrate Chemistry*, vol. 16, no. 6, pp. 751–772, 1997.

[58] Jorgensen, W. L., Maxwell, D. S., and TiradoRives, J., "Development and testing of the opls all-atom force field on conformational energetics and properties of organic liquids," *Journal of the American Chemical Society*, vol. 118, no. 45, pp. 11225–11236, 1996.

[59] MacKerell, Jr., A. D., Bashford, D., Bellott, M., Dunbrack Jr., 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, III, W. E., Roux, B., Schlenkrich, M., Smith, J. C., Stote, R., Straub, J., Watanabe, M., Wiorkiewicz-Kuczera, J., Yin, D., and Karplus, M., "All-atom empirical potential for molecular modeling and dynamics studies of proteins," *Journal of Physical Chemistry B*, vol. 102, pp. 3586–3616, 1998.

[60] Engh, R. A. and Huber, R., "Accurate bond and angle parameters for x-ray protein-structure refinement," *Acta Crystallographica Section A*, vol. 47, pp. 392–400, 1991.

[61] Morishita, T., "Fluctuation formulas in molecular-dynamics simulations with the weak coupling heat bath," *Journal of Chemical Physics*, vol. 113, no. 8, pp. 2976–2982, 2000.

[62] Dean, J. A., *Lange's Hanbook of Chemistry*. McGraw-Hill Professional, 1998.

[63] Nose, S., "An improved symplectic integrator for nose-poincare thermostat," *Journal of the Physical Society of Japan*, vol. 70, no. 1, pp. 75–77, 2001.

[64] Cho, K. and Joannopoulos, J. D., "Ergodicity and dynamic properties of constant-temperature molecular-dynamics," *Physical Review A*, vol. 45, no. 10, pp. 7089–7097, 1992.

[65] Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E., "Equations of the state calculation by fast computing machines," *Journal of Chemical Physics*, vol. 21, pp. 1087–1091, 1953.

[66] Madras, N. and D., S. A., "The pivot algorithm: a highly efficient monte carlo method for the self-avoiding walk," *Journal of Statistical Physics*, vol. 50, pp. 109–186, 1988.

[67] Jacobson, S. H., "Molecular modeling studies of polymeric gas separation and barrier materials: structure and transport mechanisms," *Polymers for Advanced Technologies*, vol. 5, pp. 724–732, 1994.

[68] Fried, J. R. and Goyal, D. K., "Molecular simulation of gas transport in poly[1-(trimethylsilyl)-1-propyne]," *Journal of Polymer Science Part B: Polymer Physics*, vol. 36, pp. 519–536, 1998.

[69] Pollino, J. M., Stubbs, L. P., and Weck, M., "Living romp of exo-norbornene esters possessing pdii scs pincer complexes or diaminopyridines," *Macromolecules*, vol. 36, pp. 2230–2234, 2003.

[70] Carlise, J. R. and Weck, M., "Side-chain functionalized polymers containing bipyridine coordination sites: polymerization and metal-coordination studies," *Journal of Polymer Science Part A: Polymer Chemistry*, vol. 42, pp. 2973–2984, 2004.

[71] Hansen, S. M., Rominger, F., Metz, M., and Hofmann, P., "The first grubbs-type metathesis catalyst with cis stereochemistry: Synthesis of [(eta(2)-dtbpm)cl2ru=ch-ch=cme2] from a novel, coordinatively unsaturated dinuclear ruthenium dihydride," *Chemistry-A European Journal*, vol. 5, pp. 557–566, Feb. 1999.

[72] Hamilton, J. G., *Handbook of Metathesis*, vol. 3. John Wiley & Sons, 2003.

[73] Mark, J. E., "Random-coil configurations of cis-1,4-polybutadiene and cis-1,4-polyisoprene. theoretical interpretation," *Journal of the American Chemical Society*, vol. 88, pp. 4354–4359, 1966.

[74] Nash, S. G., "A survey of truncated-newton methods," *Journal of Computational and Applied Mathematics*, vol. 124, pp. 45–59, 2000.

[75] Tripos, "Tripos bookshelf." Website, 2007. http://graphics.med.yale.edu:5080/TriposBookshelf/sybyl/molcad/surfaces2.html.

[76] Lee, B. K. and Richards, F. M., "The interpretation of protein structures: Estimation of static accessibility," *Journal of Molecular Biology*, vol. 55, pp. 379–400, 1971.

[77] Connolly, M. L., "Analytical molecular surface calculation," *Journal of Applied Crystallography*, vol. 16, pp. 548–558, 1983.

[78] Connolly, M. L., "Solvent-accessible surfaces of proteins and nucleic acids," *Science*, vol. 221, pp. 709–713, 1983.

[79] Edelsbrunner, H., "Weighted alpha shapes," Tech. Rep. Report UIUCDCS-R-92-1760, Department of Computer Science, University of Illinois at Urbana-Champaign, 1992.

[80] Edelsbrunner, H., Facello, M. A., Fu, P., and Liang, J., "Measuring proteins and voids in proteins," *Proceedings of the 28th Hawaii International Conference on Systems Science*, vol. 5, pp. 256–264, 1995.

[81] Lu, N. and Kofke, D. A., "Accuracy of free-energy perturbation calculations in molecular simulation. i. modeling," *Journal of Chemical Physics*, vol. 114, pp. 7303–7311, 2001.

[82] Hess, B., Bekker, H., Berendsen, H. J. C., and Fraaije, J. G. E. M., "Lincs: A linear constraint solver for molecular simulations," *Journal of Computational Chemistry*, vol. 18, pp. 1463–1472, Sept. 1997.

[83] Sattigeri, J. A., Shiau, C. W., Hsu, C. C., Yeh, F. F., Liou, S., Jin, B. Y., and Luh, T. Y., "Remarkable enhancement of second-order nonlinear optical properties of polynorbornenes having pendant chromophores. use of hyper-rayleigh scattering to estimate the tacticity of rigid rod polymers," *Journal of the American Chemical Society*, vol. 121, pp. 1607–1608, Feb. 1999.

[84] Lin, W. Y., Murugesh, M. G., Sudhakar, S., Yang, H. C., Tai, H. C., Chang, C. S., Liu, Y. H., Wang, Y., Chen, I. W. P., Chen, C. H., and Luh, T. Y., "On the rigidity of polynorbornenes with dipolar pendant groups," *Chemistry-A European Journal*, vol. 12, pp. 324–330, 2006.

[85] Diaz, K., Vargas, J., Del Castillo, L. F., Tlenkopatchev, M. A., and Aguilar-Vega, M., "Polynorbornene dicarboximides with cyclic pendant groups: synthesis and gas transport properties," *Macromolecular Chemistry and Physics*, vol. 206, pp. 2316–2322, 2005.

[86] Contreras, A. P., Cerda, A. M., and Tlenkopatchev, M. A., "Synthesis of high-tg polymers by ring-opening metathesis polymerization of n-cycloalkylnorbornene dicarboximide," *Macromolecular Chemistry and Physics*, vol. 203, pp. 1811–1818, 2002.

[87] Hino, T., Inoue, N., and Endo, T., "Ring-opening metathesis copolymerization behaviors of cyclooctene and norbornene bearing a five- or six-membered ring cyclic

carbonate," *Journal of Polymer Science Part A-Polymer Chemistry*, vol. 43, p. 6599, Dec. 2005.

[88] Hino, T., Inoue, N., and Endo, T., "Detailed study of the ring-opening metathesis polymerization of norbornene bearing a five- or six-membered ring cyclic carbonate along with volume expansion," *Journal of Polymer Science Part A-Polymer Chemistry*, vol. 44, pp. 395–405, Jan. 2006.

[89] Bielawski, C. W. and Grubbs, R. H., "Highly efficient ring-opening metathesis polymerization (romp) using new ruthenium catalysts containing n-heterocyclic carbene ligands," *Angewandte Chemie-International Edition*, vol. 39, no. 16, pp. 2903–2906, 2000.

[90] Tlenkopatchev, M. A. Personal Communication, 2007.

[91] Zheng, X., Jones, C. W., and Weck, M., "Ring-expanding olefin metathesis: A route to highly active unsymmetrical macrocyclic oligomeric co-salen catalysts for the hydrolytic kinetic resolution of epoxides," *Journal of the American Chemical Society*, vol. 129, pp. 1105–1112, 2007.

[92] Ready, J. M. and Jacobsen, E. N., "Highly active oligomeric (salen)co catalysts for asymmetric epoxide ring-opening reactions," *Journal of the American Chemical Society*, vol. 123, no. 11, pp. 2687–2688, 2001.

[93] Zhang, W., Loebach, J. L., Wilson, S. R., and Jacobsen, E. N., "Enantioselective epoxidation of unfunctionalized olefins catalyzed by (salen)manganese complexes," *Journal of the American Chemical Society*, vol. 112, pp. 2801–2803, 1990.

[94] Martinez, L. E.and Leighton, J. L., Carsten, D. H., and Jacobsen, E. N., "Highly enantioselective ring opening of epoxides catalyzed by chiral (salen)cr(iii) complexes," *Journal of the American Chemical Society*, vol. 117, pp. 5897–5898, 1995.

[95] Tokunaga, M., Larrow, J. F., Kakiuchi, F., and Jacobsen, E. N., "Asymmetric catalysis with water: Efficient kinetic resolution of terminal epoxides by means of catalytic hydrolysis," *Science*, vol. 277, no. 5328, pp. 936–938, 1997.

[96] Annis, D. A. and Jacobsen, E. N., "Polymer-supported chiral co(salen) complexes: Synthetic applications and mechanistic investigations in the hydrolytic kinetic resolution of terminal epoxides," *Journal of the American Chemical Society*, vol. 121, no. 17, pp. 4147–4154, 1999.

[97] Nielsen, L. P. C., Stevenson, C. P., Blackmond, D. G., and Jacobsen, E. N., "Mechanistic investigation leads to a synthetic improvement in the hydrolytic kinetic resolution of terminal epoxides," *Journal of the American Chemical Society*, vol. 126, no. 5, pp. 1360–1362, 2004.

[98] Ready, J. M. and Jacobsen, E. N., "A practical oligomeric [(salen)co] catalyst for asymmetric epoxide ring-opening reactions," *Angewandte Chemie-International Edition*, vol. 41, no. 8, pp. 1374–1377, 2002.

[99] Jain, S., Zheng, X., Jones, C. W., Weck, M., and Davis, R. J., "Importance of counterion reactivity on the deactivation of co-salen catalysts in the hydrolytic kinetic resolution of epichlorohydrin," *Inorganic Chemistry*, vol. 42, pp. 8887–8896, 2007.

[100] Leung, W.-H., Chan, E. Y. Y., Chow, E. K. F., Williams, I. D., and Peng, S.-M., "Metal complexes of a chiral quadridentate schiff base," *Journal of the Chemical Society, Dalton Transactions*, pp. 1229–1236, 1996.

[101] Chapman, J. J., Day, C. S., and Welker, M. E., "Enantioselective diels-alder reactions of optically active (buta-1,3-dien-2-yl)(salen)cobalt(iii) complexes," *European Journal of Organic Chemistry*, pp. 2273–2282, 2001.

[102] Cohen, C. T., Thomas, C. M., Peretti, K. L., Lobkovsky, E. B., and Coates, G. W., "Copolymerization of cyclohexene oxide and carbon dioxide using (salen)co(iii) complexes: synthesis and characterization of syndiotactic poly(cyclohexene carbonate)," *Dalton Transactions*, pp. 237–249, 2006.

[103] Beachy, M. D., Chasman, D., Murphy, R. B., Halgren, T. A., and Friesner, R. A., "Accurate ab initio quantum chemical determination of the relative energetics of peptide conformations and assessment of empirical force fields," *Journal of the American Chemical Society*, vol. 119, pp. 5908–5920, 1997.

[104] Shi, S., Yan, L., Yang, Y., Fisher-Shaulsky, J., and Thacher, T., "An extensible and systematic force field, esff, for molecular modeling of organic, inorganic, and organometallic systems," *Journal of Computational Chemistry*, vol. 24, pp. 1059–1076, 2003.

[105] Malek, K., Jansen, A. P. J., Li, C., and van Santen, R. A., "Enantioselectivity of immobilized mn-salen complexes: A computational study," *Journal of Catalysis*, vol. 246, pp. 127–135, 2007.

[106] Wojciechowski, M. and Lesyng, B., "Generalized born model: analysis, refinement, and applications to proteins," *Journal of Physical Chemistry B*, vol. 108, pp. 18368–18376, 2004.

[107] Barth, E., Leimkuhler, B., and Reich, S., "A test set for molecular dynamics," *Lecture Notes in Computational Science and Engineering*, vol. 24, pp. 73–103, 2002.

[108] Dhaliwal, S., Guptal, S. K., Huang, J., and Priyadarshi, A., "Algorithms for computing global accessibility cones," *Journal of Computing and Information Science in Engineering*, vol. 3, pp. 200–209, 2003.

[109] Halgren, T. A., "Representation of van der waals (vdw) interactions in molecular mechanics force-fieldspotential form, combination rules, and vdw parameters," *Journal of the American Chemical Society*, vol. 114, pp. 7827–7843, Sept. 1992.

[110] Halgren, T. A., "Merck molecular force field. ii. mmff94 van der waals and electrostatic parameters for intermolecular interactions," *Journal of Computational Chemistry*, vol. 17, pp. 520 – 552, 1996.

[111] Halgren, T. A., "Merck molecular force field. iii. molecular geometries and vibrational frequencies for mmff94," *Journal of Computational Chemistry*, vol. 17, pp. 553–586, 1996.

[112] Halgren, T. A. and Nachbar, R. B., "Merck molecular force field. iv. conformational energies and geometries for mmff94," *Journal of Computational Chemistry*, vol. 17, pp. 587 – 615, 1996.

[113] Halgren, T. A., "Merck molecular force field. v. extension of mmff94 using experimental data, additional computational data, and empirical rules," *Journal of Computational Chemistry*, vol. 17, pp. 587 – 615, 1996.

[114] Jensen, F., *Introduction to Computational Chemistry*. John Wiley & Sons, 1999.

[115] Becke, A. D., "Density-functional exchange-energy approximation with correct asymptotic behavior," *Physical Review A*, vol. 38, pp. 3098–3100, 1988.

[116] Chirlian, L. E. and Francl, M. M., "Atomic charges derived from electrostatic potentials: A detailed study," *Journal of Computational Chemistry*, vol. 8, pp. 894–905, 1987.

[117] Woods, R. J., Khalil, M., Pell, W., Moffat, S. H., and Smith, V. H. J., "Derivation of net atomic charges from molecular electrostatic potentials," *Journal of Computational Chemistry*, vol. 11, pp. 297–310, 1990.

[118] Breneman, C. M. and Wiberg, K. B., "Determining atom-centered monopoles from molecular electrostatic potentials. the need for high sampling density in formamide conformational analysis," *Journal of Computational Chemistry*, vol. 11, pp. 361–373, 1990.