ATOMISTIC SIMULATIONS OF ATACTIC POLYPROPYLENE: INITIAL GUESS OPTIMIZATION AND LONG-RANGE CORRECTION EVALUATION

A Dissertation
Presented to
The Academic Faculty

by

Nohemí D. Treviño-Garrido

In Partial Fulfillment
of the Requirements for the Degree
Master of Science in the
Department of Chemical and Biomolecular Engineering

Georgia Institute of Technology
December 2021

ATOMISTIC SIMULATIONS OF ATACTIC POLYPROPYLENE: INITIAL GUESS OPTIMIZATION AND LONG-RANGE CORRECTION EVALUATION

Approved by:

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

Dr. Martha Grover
School of Chemical and Biomolecular Engineering
Georgia Institute of Technology

Dr. Carson Meredith
School of Chemical and Biomolecular Engineering
Georgia Institute of Technology

Date Approved: August 27, 2021

To Jesus, my Lord and Savior, and to Mitchel, my loving husband.

ACKNOWLEDGEMENTS

First, I would like to express my gratitude to my research advisor, Pete Ludovice, for his mentorship, patience for teaching, and investment into my scientific career. He not only taught me invaluable skills in programming, data analysis, and molecular modelling fundamentals, but he also taught me confidence, tenacity and to find humor in everyday life. I am honored to have worked with such an admirable researcher and kind individual.

I would like to thank the remaining members of my research committee, Carson Meredith and Martha Grover. There are not enough words to express the gratitude I have for their tremendous support, kindness, and belief in me during my graduate school experience. It was because of them that I was able to join this remarkable chemical engineering Master's program at GT as well as be an ACS fellow. As a chemist prior to graduate school, transitioning into engineering was significantly facilitated by these outstanding individuals. I would also like to thank Alexa Dobbs, another ACS fellow, for her companionship and support during this experience. I would like to thank our collaborators at USF for their aid in this research effort. Particularly, I would like to thank Sahar Zenoozi for her friendship, work in polystyrene, and brainstorming sessions.

I would like to thank my mother, Margarita, for her unconditional love and my husband, Mitchel, for his optimism and strength; I could not have done it without them. I would like to thank Rosemont Baptist church and the wonderful women of Casey's Bible study for their countless prayers and genuine, Christ-like love. Lastly and most importantly, I would like to thank God for being my true friend, for His redemption, and for His innumerable blessings even if I don't deserve them.

TABLE OF CONTENTS

Acknowledgements **iv**

List of Tables **vii**

List of Figures **viii**

List of Symbols and Abbreviations **xi**

SUMMARY **xvii**

CHAPTER 1. Introduction **1**
1.1 Motivation **1**
1.2 Objectives **8**
1.3 Overview of Thesis **11**

CHAPTER 2. Methods **12**
2.1 Components of the Initial Guess Generator **13**
2.1.1 Polymerization 14
2.1.2 Chain Configuration Statistics 16
2.1.3 Conformation Statistics: Rotational Isomeric State Model 18
2.2 Simulations **23**
2.2.1 Background 23
2.2.2 Forcefield 24
2.2.3 Energy Minimization 27
2.2.4 Molecular Dynamics 30

CHAPTER 3. Parameter and Initial Structure Optimization **34**
3.1 Parameters of Interest **34**
3.1.1 Shape and Size Parameters 35
3.1.2 Density and Molecular Weight 39
3.1.3 Diffusion and Mean Square Displacement 39
3.1.4 Energetic Parameters 40
3.2 Initial Guess Optimization **45**
3.2.1 3-step Minimization 52
3.2.2 Density and Asymmetry of Chains 52
3.2.3 Monomer Minimization 61
3.2.4 Constant Experimental Density 64
3.2.5 Structure Filtration 65
3.2.6 Forcefield Modification Methods 67
3.2.7 RIS Model Modification 69

CHAPTER 4. Long-Range corrections **71**
4.1 Introduction **71**
4.2 Long-Range Energy Corrections **71**
4.2.1 Dispersion Energy and Pressure Corrections 74

4.2.2 Electrostatic Energy Corrections 76
4.2.3 Proposed Extrapolation Method 78

CHAPTER 5. Final Remarks 86
5.1 Conclusions 86
5.2 Recommendations 87

Appendix A. Polybuild SVL Code 89

Appendix B. R_g^2, LRC, and CED Calculation Code 101

Appendix C. Tacticity Generator Matlab code 106

Appendix D. Modified Forcefield 107

Appendix E. Journal Licenses/Permissions 121

REFERENCES 122

LIST OF TABLES

Table 2-1	Plateau Moduli and Entanglement Molecular Weights of Polypropylene. Copyright License is included in Appendix E.	15
Table 2-2	Bond Lengths and Angles of Monomers.	17
Table 2-3	Polypropylene Density and Stereoregularities. Copyright is located in Appendix E.	18
Table 2-4	Torsion Angles and Statistical Weights of Polypropylene.	23
Table 3-1	Isolated aPP Chain Structural Values 60 Initial Guesses were Averaged at Varying Degrees of Polymerization.	38
Table 3-2	Percent Difference in Parameter Values after Molecular Mechanics and Dynamics Using Six Different Methods.	51
Table 3-3	Lower Density Method Statistical Values. Twenty Structures were Comparison of Varying Densities and Relaxation Stages.	56
Table 3-4	p-Values and Sample Size	57
Table 3-5	Densities Observed after NPT Simulations.	59
Table 3-6	Post-MD (NVT) Atactic Polypropylene Parameters Obtained from 15 Structures of 2 Chains and 76 Monomers at Various Densities.	60
Table 3-7	Initial Guess Forcefield Modifications.	68

LIST OF FIGURES

Figure 1-1	Moore's Law	1
Figure 1-2	Phases of a Polystyrene-Polyisoprene (PI) BCP as a Function of χ Times the Chain Size (N).	3
Figure 1-3	Effect of χn on Uniformity or Defectivity.	3
Figure 1-4	Directed Self-Assembly of Block Copolymers Defect Types	4
Figure 1-5	United-Atom Polypropylene Chain.	5
Figure 2-1	Depiction of System after Boxstart, a Function that Randomly Inserts, Translates, and Rotates Initial Monomers within the Periodic Cell.	16
Figure 2-2	Most Popular and Simple Polymer Chain Models. A) Freely-Jointed Chain Model, B) Harmonic Bead-Spring Model, C) Rotational Isomeric State Model.	19
Figure 2-3	Torsional Potential Curve and Torsional States.	22
Figure 2-4	Statistical Matrices Associated with a 3-State RIS Model for Polypropylene.	23
Figure 2-5	Lennard-Jones Potential Energy Curves for HCl and CO.	26
Figure 2-6	Bromine Pseudo-Atoms in aPP Monomer.	27
Figure 2-7	NVT Trajectories. A) 2 Chains of 76 Monomers. B) 8 Chains of 170 Monomers	32
Figure 3-1	CED as a Metric for Entanglement	42
Figure 3-2	Relationship between R_g^2 and CED	44
Figure 3-3	Post-Dynamics R_g^2 Comparison for Various Initial Guess Methods, Average over all Structures and Chains.	46
Figure 3-4	Post-Dynamics Collapse and Characteristic Ratio Comparison for Various Initial Guess Methods	46
Figure 3-5	Post-Dynamics Chain Half R_g^2 Comparison for Various Initial Guess Methods	47

Figure 3-6	Post-Dynamics Chain Half R_g^2 p-Value Comparison for Various Initial Guess Methods	47
Figure 3-7	Post-Dynamics CED Comparison for Various Initial Guess Methods	48
Figure 3-8	Mean Square Displacement After 500ps of Dynamics for Various Initial Guess Methods	48
Figure 3-9	Post-Minimization R_g^2 Comparison for Various Initial Guess Methods	49
Figure 3-10	Post-Minimization Collapse and Characteristic Ratio Comparison for Various Initial Guess Methods	49
Figure 3-11	Post-Minimization Chain Half R_g^2 Comparison for Various Initial Guess Methods.	50
Figure 3-12	Post-Minimization Chain Half R_g^2 p-Value Comparison for Various Initial Guess Methods.	50
Figure 3-13	Post-Minimization Cohesive Energy Density Comparison for Various Initial Guess Methods.	51
Figure 3-14	Shrinking Periodic Cube	54
Figure 3-15	Preliminary Gradient Determination for Polymerization-Minimization Hybrid Method. A) Mean Square Radius of Gyration Comparison. B) Cohesive Energy Density Comparison. C) Potential Energy Comparison.	63
Figure 3-16	Original Method R_g^2 Distribution and Statistical Parameters. Sample size was 64 chains for A, B, and C.	66
Figure 4-1	Construction of a Bulk Periodic System. The cutoff is equal to half the box side length.	73
Figure 4-2	van der Waals Interaction Potential Curve with Spline.	74
Figure 4-3	Potential Energy Curve for aPP Long-Range Energy Correction Calculation. A) 2 Chains of 76 Monomers, B) 8 Chains of 170 Monomers.	80
Figure 4-4	aPP Cohesive Energy Density Convergence of A) 2 Chain, 76 Monomer System, B) 8 Chain, 170 Monomer System.	82

Figure 4-5 Atactic Polystyrene A) Van der Waals and B) Electrostatic Convergence Curves for a System of 4 Chains of 181 Monomers. 84

Figure 4-6 aPP Pressure Convergence Curve for a System of 8 Chains of 170 Monomers. 85

LIST OF SYMBOLS AND ABBREVIATIONS

Chapter 1

DSA	Directed Self-Assembly
BCP	Block copolymer
χ	Flory-Huggins parameter
USF	University of South Florida
MOE	Molecular Operating Environment
SVL	Scientific Vector Language
aPP	Atactic polypropylene
aPS	Atactic polystyrene
PMMA	Poly(methyl methacrylate)
PI	Poly(isoprene)
RIS	Rotational Isomeric State
MD	Molecular Dynamics
MM	Molecular Mechanics

Chapter 2

SVL	Scientific Vector Language
RIS	Rotational Isomeric State
USF	University of South Florida
MOE	Molecular Operating Environment
aPP	Atactic polypropylene
aPS	Atactic polystyrene
MD	Molecular Dynamics

MM	Molecular Mechanics
G_N^o	plateau modulus
M_e	Entanglement Molecular Weight
M_w	Molecular Weight
T_g	Glass Transition Temperature
ρ	Density
χ	Side group torsional angle
φ_i	Intradiad dihedral angle
φ_{i+1}	Interdiad dihedral angle
g	Gauche conformation
t	Trans conformation
g-	Gauche minus conformation
θ	RIS model bond angle
N_s	Number of segments in a freely-jointed chain model
b	Length of each segment
$\vec{R}$	An observed vector in freely-jointed chain model
H	Hamiltonian
Ψ	Wave function
E	Energy or eigenvalues
R	Nucleus position
r	Electron position
r_a	Equilibrium bond distance
ϑ_b	Equilibrium bond angle
Φ_c	Equilibrium dihedral angle

s	Expansion coefficient of the Fourier term
K	Force constants
U	Potential
t	Time
e_A	Specified error
N	Number of atoms
x_k	Atom positions
T	Number of significant figures in the potential value
p_k	Search direction
a_k	Step size
k	Minimization index
τ_e	Relaxation time
ζ	Friction coefficient
D_e	Lennard-Jones energy well
σ^*	Distance of maximum attraction
q	Charge
ϵ_0	Permittivity in a vacuum
ϵ	Permittivity in the medium

Chapter 3

r	End-to-end distance
l	Length of bond
n	Number of bonds
r_f	End-to-end distance of freely-jointed chain
C_∞	Characteristic ratio

ρ_n	Gaussian distribution of end-to-end distance probabilities
R_g	Radius of gyration
C_R	Collapse ratio
MSD	Mean square displacement after MD
N	Number of atoms
x	Position of atoms
t	Time
n	Number of dimensions
E_{coh}	Cohesive energy
$U_{tot,isolated}$	Sum of isolated chain potential energies
U_{tot}	Total potential energy
δ	Hildebrand solubility
V	Volume of periodic cell
gtest	Euclidean norm minimization gradient
ε	Dispersion parameter
σ	Particle size
CED	Cohesive energy density
D_s	Segmental self-diffusivity after MD

Chapter 4

L	Length of periodic box side
N	Number of atoms
g	Pair distribution function
u	Potential energy
r	distance between a pair of atoms

k_B	Boltzmann constant
T	Temperature
P	Pressure
V	Volume
q	Charge
ϵ_0	Permittivity in a vacuum
ϵ	Permittivity of the medium or dielectric constant
R_c	Truncation distance
U_F	Fourier part of Ewald Sum
k	Number of charges
l	Lattice vectors in Fourier space, (l_x, l_y, l_z)
k	$(2\pi/L)$**l**
r	Point charge position
α	associated with the Guassian width
U_{self}	Correction due to self-interaction
$U_{short\text{-}range}$	Short-range interaction contribution to the Ewald Sum
U_{Coul}	Ewald Sum
φ(**r'**)	Multipole Expansion potential at a point **r'**
$\mathbf{r_i}$ (r_i, θ_i, φ_i)	location of charge
Y_n^{-m}	Legendre functions
E_{coh}	Cohesive energy
$U_{tot,isolated}$	Total isolated chain potential energy
U_{tot}	Total energy of a structure
δ	Hildebrand solubility

Chapter 5

R_g^2	Mean square radius of gyration
CED	Cohesive energy density
RIS	Rotational Isomeric State
C_∞	Characteristic ratio
C_R	Collapse ratio
MSD	Mean square displacement after molecular dynamics
σ	Particle size, forcefield parameter

SUMMARY

Molecular simulation is a cost-effective tool for the prediction of structure-property relationships of a variety of materials and the computational design of new materials. For example, a recent research effort has examined the directed self-assembly (DSA) of block copolymers (BCPs) using meso-scale or coarse-grain modelling. The future of DSA will likely involve copolymers that are energetically incompatible so they produce self-assembled phase separation at smaller size scales. Coarse-grain modelling dramatically improves the efficiency of these molecular simulations by grouping a number of bonded backbone atoms into "pseudo-" atoms to produce an approximate representation. This approximation is particularly useful for bulk polymer systems and meso-scale phenomena such as DSA. However, this reduced approach discards the chemical details of a fully atomistic system and may compromise the refined acquisition of some important chemical parameters. Therefore, atomically-detailed simulations are required to accurately predict which copolymer blocks will strongly phase separate. One approach that little research has done in the past is to generate accurate atomistic models and then map to the coarse-grain model for further bulk simulations. Using Molecular Operating Environment (MOE) software, we began atomistic modelling of atactic polypropylene to generate a methodology capable of producing accurate initial guesses. Formulation of such a methodology will result in a significant advance in this field, and the investigation of other polymers including polystyrene and block copolymers.

It was shown by this research, that significant errors still exist in the structural and energetic components of the initial guess using the current methods and that these errors

persist after molecular mechanics and dynamics. Namely, this research showed that after about 75% of initial guess chain polymerization, chains would begin to collapse into compact conformations due to the self-avoiding nature of the chains. This resulted in structural inaccuracies in the initial guess that would not be ameliorated by energy minimization or molecular dynamics. Thus, the effect of density on the structural and energetic components of initial guess generation was investigated in this research. Findings showed significant structural asymmetry in chains generated with the present established initial guess method. Additionally, cohesive energy density, diffusivity, and entanglement of atomistic models of atactic polypropylene are strongly sensitive to density changes and were improved by initial densities that were greater than the experimental density. However, there were only slight changes to structural properties such as mean square radius of gyration, the characteristic ratio, and chain symmetry as density was modified. Additional initial guess methods were developed such as integrating minimization into the polymerization process, filtration of initial guesses based on their structural properties, multi-step minimization, a periodic boundary expansion method, and optimization of forcefield parameters for initial guess generation. Integrating minimization into the polymerization process had the surprising effect of improving chain symmetry. Optimized forcefields, multi-step minimization, and structure filtration had a significantly positive effect on the shape and size of the structures as shown by elevated mean square radius of gyration and characteristic ratio values.

Additionally, a simple, rapid, and empirical method was developed for calculating energy and pressure long-range corrections. This method involved iteratively replicating the periodic lattice of the polymer system and matching the cutoff to the new size of the system.

Thus, the potential energy was dependent on the number of cells in the lattice and the potential energy eventually converged to provide the long-range corrections. The cohesive energy also converged since the isolated energy of the chains is a constant value and could also be rapidly calculated.

CHAPTER 1. INTRODUCTION

1.1 Motivation

The directed self-assembly (DSA) of block copolymers (BCP) has been a subject of intensive investigation over the past several decades due to its potential to extend Moore's Law (Figure 1-1).[1] One of the leading applications of BCPs is in semiconductor fabrication processes since this is far more economical than other alternatives such as EUV lithography and complex multiple-patterning schemes.[1-6]

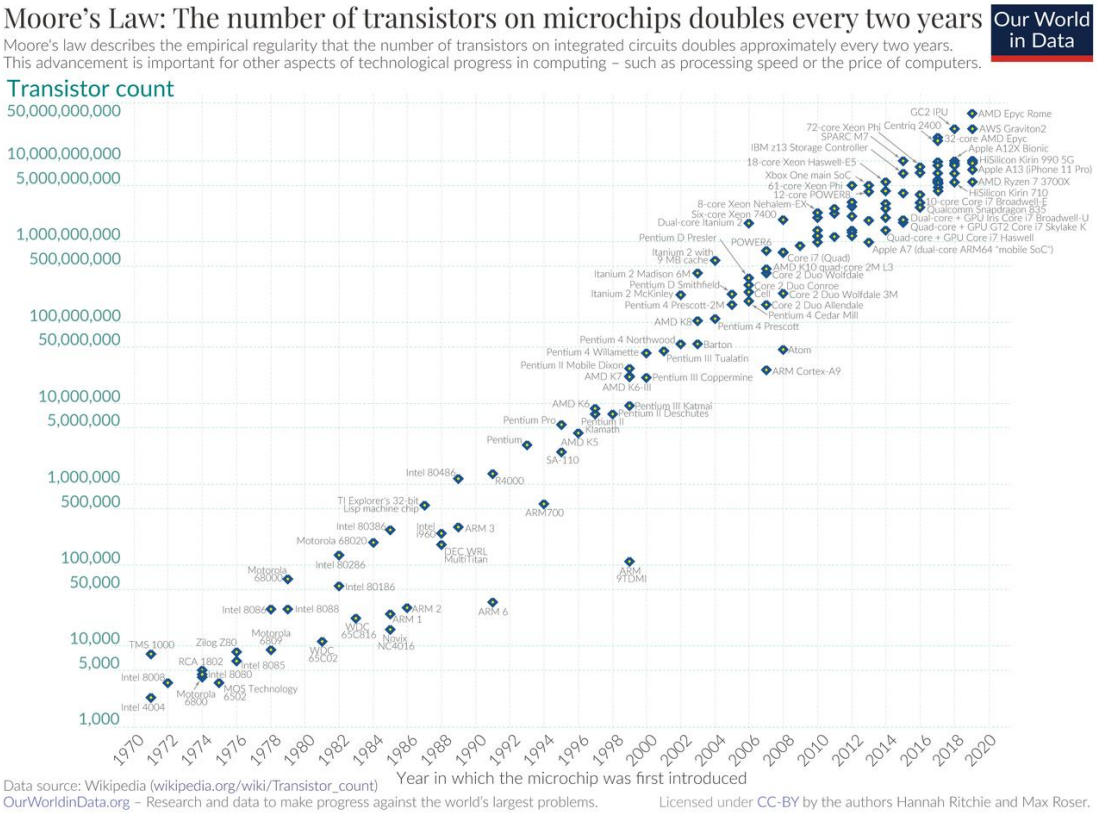

Figure 1-1 Moore's Law

The reason BCPs have such high potential in nanotechnology and photolithography is because of their intrinsic ability to phase-separate and assemble into uniform lamellae or cylindrical structures with tunable dimensions, as well as their numerous possible functional group combinations.[1-5] With the utilization of BCPs, device fabricators can manipulate the energy landscape of these structures through DSA to directly control their alignment and patterning.[1] The structural properties and electronegativities of the blocks can radically affect the self-assembly of the block copolymer and thus produce new and improved processes, morphologies, and applications.[7] DSA of BCPs may be utilized to produce patterns in next generation microelectronics by formulating a very small etch mask from the nanoscale lines or holes (lamellar phase labelled LAM in Figure 1-2) that form from BCP phase separation on a substrate pattern using traditional photolithography.[1, 4-5, 8] Idealized simulation work has shown how the Flory-Huggins χ parameter (which approximates the enthalpy of mixing of the BCP segments) effects this phase separation (see symmetric dashed line in Figure 1-2).[5-6, 8-9]

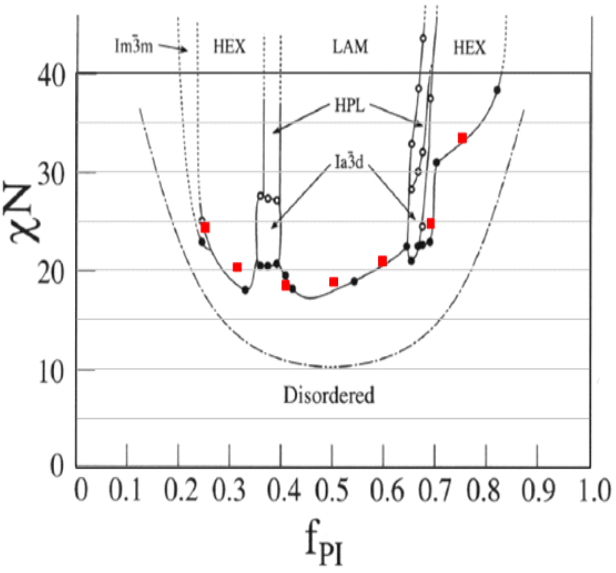

Figure 1-2 Phases of a Polystyrene-Polyisoprene (PI) BCP as a Function of χ Times the Chain Size (N).

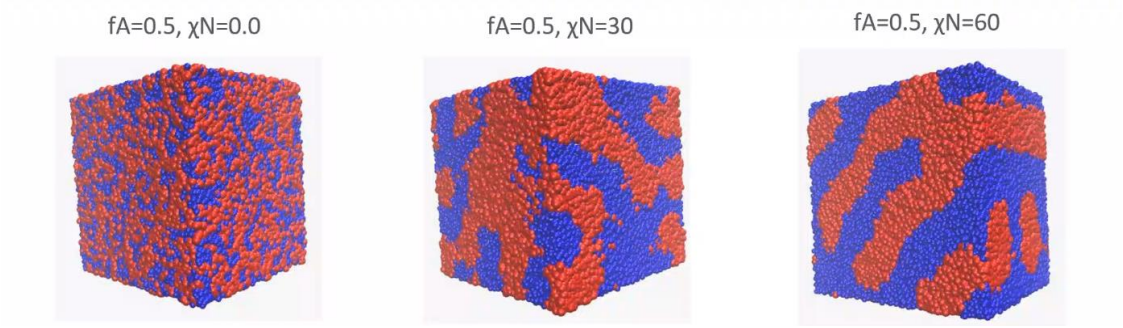

Figure 1-3 Effect of χn on Uniformity or Defectivity.

Despite the potential BCPs have for economical and improved semiconductor fabrication, the density of defects—namely bridge defects—have proven to still be above the ideal for implementation at an industrial-scale.[1] Affinity defects in chemoepitaxial underlayers have been suspected to propagate the bridge defects in the desired BCP-film

overlayer.[1] These defects act as obstacles for circuit signals and eliminating them are an important challenge for improved circuits.

Figure 1-4 Directed Self-Assembly of Block Copolymers Defect Types

The key to understanding and reducing these defects and to design improved BCPs and DSA processes is to first identify the structural and energetic components and their associated effects. Further investigation must be done to identify which structural properties of the BCPs or factors in DSA processing are involved in bridge defectivity.

A recent research effort has examined many polymer parameters that affect DSA using molecular simulations.[1-4, 6-7, 10] Molecular simulations are a cost-effective tool for the prediction of structure-property relationships of a variety of materials and the computational design of new materials. Course-grain or meso-scale modeling dramatically improves the efficiency of polymer molecular simulations by grouping a number of bonded backbone atoms into "pseudo-" atoms to produce an approximate representation. A united atom typically consists of a Kuhn length number of monomers. Figure 1-5 below illustrates how skeletal atoms in polypropylene are grouped to make a united-atom approximation.[11-12] Coarse-grain modeling would be a choice approach for modeling the directed self-

assembly of macromolecules such as block copolymers, since BCPs can contain thousands of atoms and DSA involves forming lamellae that can be several nanometers thick. Coarse-grain modeling is used as a computational expedient for molecular simulations since a detailed atomistic or quantum model would be more computationally expensive.[11]

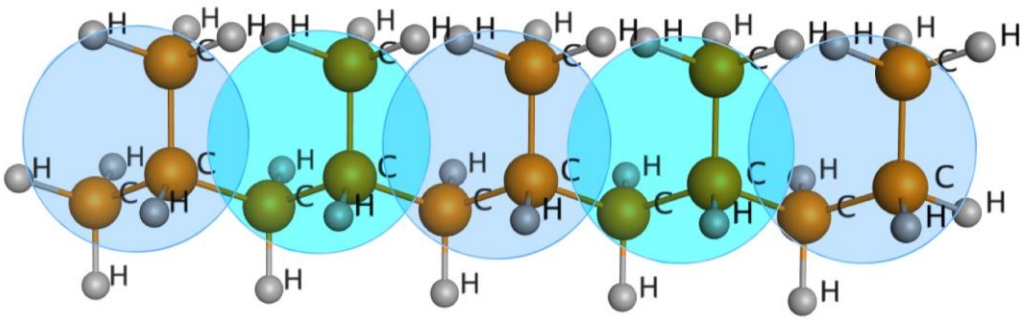

Figure 1-5 United-Atom Polypropylene Chain.

By far, the majority of nanoscale patterning research and BCP DSA meso-scale modeling has been conducted using polystyrene-b-poly(methyl methacrylate) (PS-b-PMMA).[8] The reason is that the average interaction and surface energies of the individual block components are similar or *symmetric* at the annealing temperature, resulting in perpendicular structures and a lack of homopolymer preference for the surface during DSA.[3-4, 8] The degree of similarity between energies of each copolymer interacting with itself is reflected in the symmetry of the resulting phase diagram, and the more symmetric a BCP, the simpler and more ideal the model. This symmetry produces a phase diagram that is also symmetric unlike the asymmetric polystyrene-b-polyisoprene (PS-b-PI) phase diagram seen in Figure 1-2.[2] The future of DSA will likely involve different homopolymers that are asymmetric to produce novel hybrid morphologies and materials.[13] This would

mean that the homopolymer densities, size and shape, and/or cohesive energies are different from each other and that a more in-depth approach such as atomistic modeling would be ideal for understanding the chemical details of the copolymer blocks.[7] A symmetric BCP such as PS-b-PMMA only requires a single Flory-Huggins parameter since the homopolymers are similar, but more complicated, asymmetric BCPs such as PS-b-PI require several parameters and more details about the individual blocks.[8, 13]

Naturally, a coarse-grained reduced approach discards the chemical details of a fully atomistic system and may compromise the refined acquisition of some important chemical parameters and parameter relationships such as polymer size, shape, entanglement, and solubility. This thesis work aimed to do something little research in the past has done which is to improve the accuracy of models by first generating an atomistic model, and then mapping to produce a coarse-grain model. Difficult-to-model polymers and BCPs greatly benefit from such a strategy since, in the past, a qualitative description for the polymer properties has been accepted over quantitative accuracy, leading to significant intrinsic errors. Moreover, even the established atomistic models from previous work do not accurately and completely reproduce relevant properties, thus generating accurate atomistic models is of great import. The accuracy of atomistic models is largely dependent on the initial guess for two main reasons. Firstly, a structure may reach a form of pseudo-equilibrium during molecular mechanics and dynamics which is assumed to be true relaxation of the system. However, the initial conformation may be so unstable, a reasonable amount of mechanics or dynamics may not be sufficient to fully relax the system. This is especially true for more complex polymers including polymers with bulky side groups or BCPs. The closer the initial conformation to the relaxed state, the less

simulation is required. Secondly, chain entanglement is dependent on the polymerization process or initial guess generation since chains generally do not become more entangled during mechanics or dynamics.

Additionally, a comprehensive investigation of the initial parameters that can be modified to systematically optimize polymer properties has never been conducted. Herein, a comparison of all the relevant polymer properties obtained using a wide variety of techniques such as density changes, forcefield optimization, Rotational Isomeric State model optimization, Monte-Carlo-like filtration, and minimization techniques was performed to comprehensively define the protocols for optimal model accuracy. Several of polymer properties including mean square radius of gyration, cohesive energy density, segmental mean square displacement, density, and end-to-end distance were selected to define the accuracy of a model. One attempt to produce an all-atom model was that of Theodorou and Suter who developed a method to predict and measure properties of low molecular weight well-relaxed glassy polymer (atactic polypropylene) systems.[14-15] As such, atactic polypropylene was utilized to establish optimal modelling protocols and evaluate the relevant parameter relationships. Admittedly, polypropylene is not used in DSA but it is the ideal proxy or representative polymer for studying general approaches for initial guess acquisition which is the focus of this work, and the information gathered was essential for collaborators at the University of South Florida to generate accurate polystyrene and PS-b-PI models.

Long-range energy and pressure corrections were also obtained and included in total energy and pressure calculations using an empirical method developed in this thesis work. Long-range corrections are difficult to calculate, especially for charged systems, and

are therefore often omitted from total energy and pressure calculations in molecular modeling work. To elaborate, van der Waals potential energy is generally obtained by direct integration of the atomic energy expression, the instantaneous pressure requires calculation of the virial which includes the virial due to bond constraints in molecular dynamics (also difficult to calculate), and the electrostatic energy is divergent with increasing radial distance and requires conditionally convergent and complex methods such as Ewald Sum or Multipole Expansion.[11, 16] Though difficult to obtain, long-range energy and pressure corrections, may not be insignificant for polymer systems and should be included. As such, a simple, relatively quick, and empirical method for calculating nonbonded energy and pressure corrections was developed in this thesis work. This method involves increasing the dimensions of the periodic lattice and the cutoff until convergence of the energy or pressure is observed. This brute force approach has the advantage that even polar systems can have their own long-range energy correction well characterized without worrying about model assumptions like the conditional convergence of methods like Ewald summation or multipole expansion.

1.2 Objectives

A main objective of this project was to develop an all-atom tailored protocol for modeling various polymers (e.g. polypropylene and polystyrene) at higher molecular weights (above the molecular entanglement weight) using molecular minimization and dynamics to obtain a number of parameters that were consistent with each other and experiment. Minimization and dynamics serve to relax unfavorable conformations; however, a large, poorly constructed initial guess may be stuck in an unfavorable energetic conformation, and minimization and dynamics would not be capable of relaxing the system

within a reasonable simulation timeframe. Thus, the accuracy of a polymer model can mainly be attributed to the upstream initial structure, or guess, and significant effort was allocated into developing an accurate initial guess generator.

The next objective was to evaluate the initial parameters (such as density, forcefield, and RIS models) and the resulting polymer properties to optimize the initial guess generator and minimization protocols. Determination of the key conditions, structural components, and techniques that would result in consistently accurate initial guess generation and efficient energy minimization was a paramount objective in this research. Several techniques were proposed and compared in Chapter 3. Each technique yielded a characteristic set of polymer properties including some particularly improved properties, e.g. energy minimization after each monomer was added during initial guess generation produced highly symmetric chains, reduction of the σ parameter in the forcefield during initial guess generation improved the mean square radius of gyration of the chains, etc. A combination of several methodologies could result in overall improvements in polymer models. Any property tradeoffs were reported, e.g. increasing density improved cohesive energy density but worsened mean square radius of gyration accuracy.

The use of quantum mechanical calculations to fit forcefields can neglect the long-range corrections (LRC) due to periodic boundaries, which are applied to amorphous polymer simulations. Previous work on PVC showed that a commercial forcefield differed from local quantum calculations by the exact amount of this LRC.[17] Thus, a main objective was to develop and apply a method for calculating the nonbonded energy and pressure LRC, using the fact that the LRC depends on the size of the periodic cell. Additionally, since the total energy of the isolated chains is constant, the cohesive energy would also

converge; therefore, application of this method would also serve to calculate cohesive energy density. The results from utilizing this method are presented in Chapter 4.

The specific objectives were as follows:

1. Develop an SVL (Scientific Vector Language) program using the Molecular Operating Environment (MOE) software that would synchronously build multiple atactic polypropylene chains with a specified degree of polymerization inside a periodic cube. The program would incorporate tacticity tunability and a tunable RIS (rotational isomeric state) model to determine torsional states of polymers with various chemical structures.
2. Evaluate parameters of the atactic polypropylene system such as cohesive energy density, mean square displacement, mean square radius of gyration, and characteristic ratio at varying densities to determine the effect of density on initial guess generation and to reduce the asymmetry of growing chains due to their self-avoiding nature.
3. Systematically apply and compare optimization techniques for initial guess generation and energy minimization such as Monte-Carlo-like structure filtration, integrated monomer minimization, initial guess forcefield modification, and RIS model modification to produce an improved and relaxed polymer system.
4. Generate a step-wise periodic cube expansion algorithm to calculate energy and pressure long-range corrections. The original periodic cube should be iteratively

replicated to construct a lattice of polymer system replicates, then ‘built’ to form a single large periodic cube with side lengths equal to the product of the original cube side lengths and the iteration index while modifying the cutoff to be half the size of the newly built cube. With a growing number of lattice replicates, the potential energy will converge. Additionally, we can apply this method to accurately calculate the cohesive energy density.

1.3 Overview of Thesis

Chapter 1 provides the motivation and objectives behind this research. Chapter 2 explains the detailed methods and algorithms utilized to achieve the objectives described in Chapter 1. This includes details, background, and protocols for initial guess generation, energy minimization, and molecular dynamics. Chapter 3 describes the parameters of interest for this research and reports the results obtained from various initial guess generator and minimization optimization methods. It also contrasts the various methods and discusses the advantages and disadvantages for each. Chapter 4 describes the methods utilized to obtain the long-range energy and pressure corrections and expounds on the significance of these calculations for atomistic modeling. Finally, conclusions and recommendations are presented in Chapter 5 based on the results obtained.

CHAPTER 2. METHODS

With the pioneering work of Theodorou and Suter in the field of molecular simulation, a standardized strategy for generating detailed atomistic models of well-relaxed polymers under three-dimensional boundary conditions has been established. Initial polymer structures or guesses are generated, relaxed under a tailored forcefield and molecular mechanics, and subjected to molecular dynamics for further relaxation and diffusion.[11-12] Calculation of mechanical, rheological, and thermodynamic parameters from final models have proven to agree with experiment even though only few initial parameters can be altered other than forcefield parameters to affect predicted parameters.[14-15, 18-23] Unfortunately, this effective and standardized strategy for modelling well-relaxed polymers has not been incorporated into commercial software such as Molecular Operating Environment (MOE), Maestro Schrodinger, GROMACS, LAMMPS, NAMD, AMBER SUITE, etc. In fact, current polymer building software produce structures with significant conformational inaccuracies and require the incorporation of the RIS model. It is important to highlight that the success and accuracy of a final structure is largely dependent on the initial guess for several reasons. Energy minimization does not cause large overall changes to the conformational and configurational characteristics of the chains, therefore the minimized structure closely resembles the initial structure.[24] An inaccurate initial guess is likely to result in steep energetic obstacles that are computationally exhaustive during MM and MD and may result in parameter predictions that do not coincide with experimental data.[24] Additionally, the random placement of initial monomers within the periodic

boundary is important for more even spacing of chains within the box volume, i.e. to reduce the number of structures with volumetric holes or dense areas.[25]

2.1 Components of the Initial Guess Generator

One of the most critical objectives of this investigation was to generate an SVL program that would accurately build amorphous glassy polymer initial guesses and could be easily adapted for various monomeric units, periodic box sizes, and stereochemistries. As such, a significant portion of time was spent perfecting this program, and it was accomplished by programming SVL source code in MOE 2019.01.01 (Molecular Operating Environment). MOE is a drug discovery software developed by Chemical Computing Group that offers ergonomic visualization and simulation/model development of various macromolecules such as DNA and proteins and has proven to be highly effective in polymer atomistic modelling.[7, 26] MOE provided a wide range of useful built-in functions and documentation, forcefield options, and relatively efficient atomistic simulation times. As shown in the commented code in Appendix A, a master global function, PolyBuild300, was constructed that called six local functions: boxstart, monomeradd1, risparams, choosestate, residuebackbone, torsionanglechoice, and LREnergyCalc. The next subsections will further elaborate on the roles of these functions and the components that were necessary for accurate generation of polypropylene and polystyrene initial guesses. It is important to note that this investigation mainly involved polypropylene. However, the results obtained from this thesis were critical for collaborators at University of South Florida to further investigate polystyrene. Polystyrene models were built using the methods described in this chapter and were utilized to calculate electrostatic long-range energy

corrections with the method described in Chapter 4 since polypropylene is modelled with no partially charged atoms.

2.1.1 Polymerization

Initial guesses were built in a 3D periodic cell with an orthorhombic P1 spacegroup. Box dimensions were equilateral and the volume was modified according to the desired initial density. For atactic polypropylene, a density of 0.765g/cc and a molecular weight of 7169.81g/mol (i.e. 8 chains and 170 monomers) were chosen based on the density and moduli experiments conducted by Eckstein et al.[27-28] The entanglement weight of polypropylene is 7050g/mol as shown in Table 2-1.[27] In light of previous simulations conducted by Theodorou and Suter on polypropylene, a smaller system of 2 chains, 76 monomers, and box edge length of 24.075Å was also constructed for comparative purposes.[14-15, 21-23, 29] The experimental density of atactic polystyrene at 300K was 1.07g/cc, therefore that was the density that was chosen for a-PS initial guesses.[30] The entanglement molecular weight of polystyrene is 17,900g/mol; therefore, 8 chains with 172 monomers were created to have chains that were above entanglement weight and to fulfill the density requirement.[31]

Table 2-1 Plateau Moduli and Entanglement Molecular Weights of Polypropylene. Copyright License is included in Appendix E.

polymer	G_N° [a] (Pa)	G_N° [b] (Pa)	M_e [c] (g/mol)	M_w/M_e
i-PPM9	427 000		6 900	126
s-PPM5	1 350 000		2170	222
a-PPM36	418 000	415 000	7050	520

[a] Calculated from eq 2. [b] Calculated from eq 1. [c] Calculated from eq 6 ($T_0 = 463$ K).

A number of initial monomers representing the first monomer of each chain were sequentially inserted into the cell from a monomer database. The database contained only one parent monomer which was repeatedly placed in the periodic cell, randomly (normal distribution) translated somewhere in the cell, and randomly rotated for each initiating chain. For example, if 8 chains were desired, the parent monomer was placed in the periodic cell eight times, then each was translated and rotated. Figure 2-1 illustrates the first monomers placed in the periodic cell after running the boxstart function which randomly inserts, translates, and rotates these initial monomers.

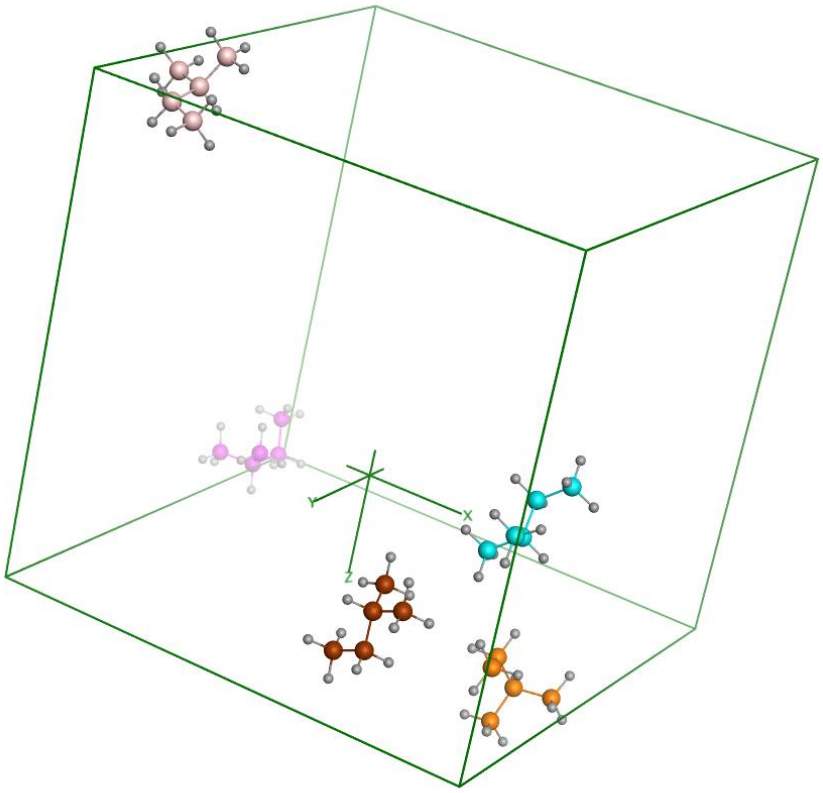

Figure 2-1 Depiction of System after Boxstart, a Function that Randomly Inserts, Translates, and Rotates Initial Monomers within the Periodic Cell.

To mimic bulk conditions, growing polymer chains that extended outside of a periodic boundary were reflected into the box on the exact opposite side of the box.

Calling the monomeradd1 function, the parent monomer was iteratively dealt out and bonded to each polymer chain instead of polymerizing entire chains at a time. Effectively, polymerization of chains occurred synchronously to promote chain entanglement and volume homogeneity.

2.1.2 Chain Configuration Statistics

After polyethylene, polypropylene is the most common and simple organic polymer since it is composed of only one type of monomer and has only a simple methyl side chain. Bond angle and length constants are detailed in Table 2-2 below.[32] These angle and stretch constants were utilized to construct the parent monomer and the initial guess.

Table 2-2 Bond Lengths and Angles of Monomers.

Bond Length (Å)	Bond Angle (º)
C-C, 1.529	CCC, 112.7
C-H, 1.090	CCH, 110.0

Unlike polyethylene, each polypropylene monomer has a chiral center due to its side chain. Atactic polypropylene was used because it is the simplest polymer that forms a glass. For isotactic polypropylene and syndiotactic polypropylene, meso to racemic proportions are 1:0 and 0:1, respectively. Atactic polypropylene, however, consists of a random mixture of meso and racemic diads. Iso-, syndio-, and atactic polypropylene triad stereoregularities were reported by Eckstein, A., *et al*. in their paper on the experimental determination of the plateau modulus and entanglement weights for high molecular weight polypropylene as shown in Table 2-3.[27]

Table 2-3 Polypropylene Density and Stereoregularities. Copyright is located in Appendix E.

polymer	M_n (g/mol)	M_w (g/mol)	M_w/M_n	T_g (K)	T_m (K)	ρ^a (g/cm^3)
i-PPM9	365 000	871 000	2.4	nd[b]	432	0.766
s-PPM5	231 000	483 000	2.1	nd	412	0.762
a-PPM36	1 890 000	3 670 000	1.9	276.9		0.765

[a] At reference temperature $T_0 = 463$ K. [b] Not determined.

polymer	mm^a (%)	mr^a (%)	rr^a (%)
i-PPM9	98.7	0.7	0.6
s-PPM5	0.7	6.3	93.0
a-PPM36	30.2	52.8	17.0

[a] Calculated from the pentad distribution using ^{13}C-NMR data.

Based on these experimental meso and racemic triad proportions, a random number generator was developed in Matlab (Appendix C. Tacticity Generator Matlab code) to produce a tacticity vector of zeros and ones. The random number vector would produce an exact desired proportion of zeros to ones, or meso to racemic stereoregularities—a zero would indicate a meso diad and a one would indicate a racemic diad. Once a monomer was added to a growing chain, the R/S configuration of the chiral center would remain the same as the previous monomer if a zero (meso) was indicated or switch if a one (racemic) was indicated in the random number vector. The first diad of each chain was always meso.

2.1.3 Conformation Statistics: Rotational Isomeric State Model

Polymer theories are established by utilizing simplified general models that incorporate relevant topological and structural characteristics to predict polymer properties and phenomena. The accuracy of a particular prediction represents the ability of the

simplified model to capture the physics of the polymer. Figure 2-2 below shows the most popular simplified models used for polymer simulations:

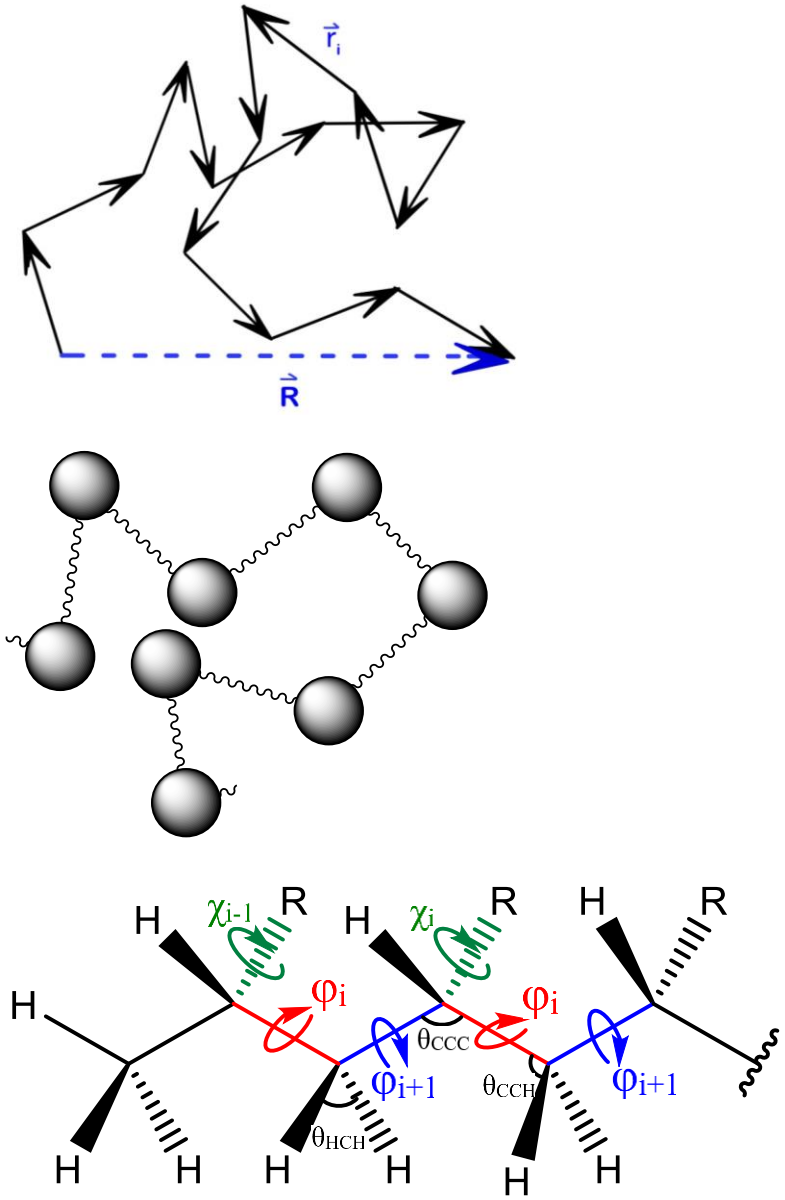

Figure 2-2 Most Popular and Simple Polymer Chain Models. A) Freely-Jointed Chain Model, B) Harmonic Bead-Spring Model, C) Rotational Isomeric State Model.

The simplest model is the freely-jointed chain model, or 'ideal' chain model (Figure 2-2A).[12] This model treats the polymer structure as a random walk where a group of

sequential skeletal atoms constitute a segment (Kuhn length).[33] These segments are rectilinear and there are no energetic interactions between them except for in the bonds connecting them. The freely-jointed segments can point in any direction independent of other segments and can even intersect. The probability of observing a particular vector, $\vec{R}$, between the two ends of the freely-jointed chain can be described with a Gaussian distribution where the mean square of $\vec{R}$ can be found with the simple formula

$$\langle \vec{R}^2 \rangle = N_s b^2 \quad (1)$$

where N_s is number of segments and b is length of the segments. The freely-jointed chain could be used for a system where interactions are only local.

Another coarse-grain ideal chain model is the bead-spring model (Figure 2-2B), or Rouse model, in which several monomers are grouped into a united atom, or 'bead', connected by harmonic springs.[12] The aim of this model is to capture the increase in entropic free energy and decrease in possible conformations as the two ends of the chain are pulled apart. The mean square end-to-end distance can be obtained from the product of $3k_BT$ and the spring constant.

The rotational isomeric state model (RIS) was first developed by P. J. Flory and Volkenstein in the late1960's even before computational polymer simulations were possible and is essentially an adaptation of the statistical Ising model for conformations of a polymer chain.[34-35] This model is a far more detailed and realistic model as it incorporates fixed values for bond angles, lengths (l), and torsional states.[12, 34-35] In the last forty years, more than six hundred RIS polymer models have been established and have proven to

provide the most accurate initial guess conformations; thus, an RIS model was integrated into our program.

The RIS model chooses torsional states based on the torsional state of the adjacent backbone bond. [12, 32-34, 36] The conformational energetics of a polypropylene diad are dependent on four torsional angles, φi, φi+1, χi-1, χi+1 as illustrated in Figure 2-2C. Skeletal diads are comprised of an interdiad bond, φ_i, which connects monomers with each other and an intradiad bond, $\varphi_{i+1,}$ which is already present in the propylene monad substrate. The RIS model approximates discrete conformational states corresponding with local dihedral potential energy minima every 120^o including trans (t), gauche$^+$ (g^+), and gauche$^-$ (g^-). Combination of states for polymer diads results in nine conformational isomers, tg^+, tg^-, g^+t, g^-t, g^+g^+, g^-g^-, g^+g^- and g^-g^+ that have statistical weights associated with them. At times, isomers may be equivalent because of symmetry. The number of states is generally three; however, any number of states may be utilized depending on the number of conformational potential minima of a diad. For example, a two state model may be used if the g^- state is not significantly favorable, a five state model (t, t*, g^+, g*, g^-) may be used, or a thirty-six state model with equivalent statistical weights may be utilized and new isomers chosen according to a Monte Carlo Metropolis criterion if acceptably low minimization cannot be achieved.[37] Figure 2-3 illustrates the torsional states associated with a five-state model developed by Flory and Suter using Newman projection models and the aPP relative statistical weights.[19, 32] Matrices that represent the torsional state statistical weights are represented in Figure 2-4.

The χ_{i-1} and χ_{i+1} dihedral bonds illustrated in Figure 2-2C correspond with rotation of diad side chains. Since the side chain of polypropylene is a mere methyl group which

contains not-so-bulky hydrogens, incorporation of these dihedral bonds with the RIS model would not be beneficial. For PS, incorporation of χ_{i-1} and χ_{i+1} torsional states may be considered in later research since the phenyl side groups are bulkier.

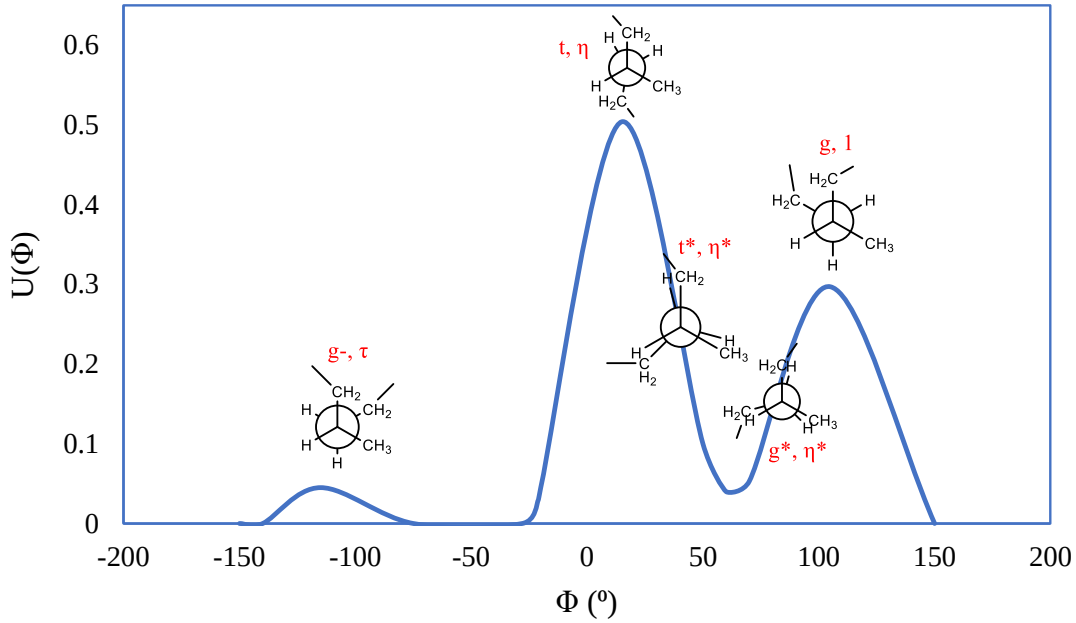

Figure 2-3 Torsional Potential Curve and Torsional States.

Figure 2-4 depicts the aPP RIS statistical probabilities utilized for this work which were obtained from Biskup, Cantow, et al.[38] U' is a matrix that predicts the conformations of the intradiad and U_m" and U_r" are the statistical matrices that predict the conformations of the meso and racemic interdiad bonds, respectively. [32, 34-35, 39] Each column and row corresponds with trans, gauche, or gauch$^-$ conformations. The variables, η, τ, ω, and τ*, are the normalized probability weights associated with a conformation combination, i.e. the probability for a particular conformation to follow another conformation. The values for these statistical weights are listed in Table 2-4.

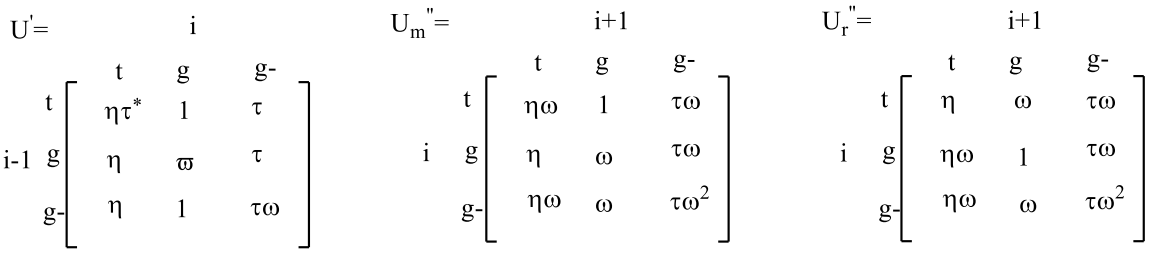

Figure 2-4 Statistical Matrices Associated with a 3-State RIS Model for Polypropylene.

Table 2-4 Torsion Angles and Statistical Weights of Polypropylene.

Torsion Angles		Statistical Wts., ξ	
t	180-Δφ	η	1.0
g	60+Δφ	τ	0.5
g^-	-60	ω	0 to 0.05
		τ*	1.0

The following local functions were developed to implement a three-state RIS model into the initial guess generator: Risparams, Choosestate, torsionanglechoice, LREnergyCalc. The specific code utilized to run these functions is shown in Appendix A.

2.2 Simulations

All simulations were carried out on three Dell Precision 5820 Tower X-Series CPUs with 64 GB of RAM and an i-9, 3.7GHz, Intel 10 core processor with a Microsoft Windows Operating System.

2.2.1 Background

The Schrodinger equation (Equation 2) is the most robust way of describing an atomic system as it predicts very detailed dynamic behavior.

$$H\Psi(R,r) = E\Psi(R,r) \quad (2)$$

H is the Hamiltonian which represents an operator corresponding to a physical property such as position or momentum, E is energy or the eigenvalues, and Ψ is the wave function predicting the positions, for example, of the electrons (r) and nuclei (R).[16] The Born-Oppenheimer Approximation is applied in molecular simulations and assumes that the motion of the nuclei and electrons can be decoupled such that the wave function depends only on the motion of the nuclei, and the Schrodinger equation simplifies to

$$H\Theta(R) = E\Theta(R) \quad (3)$$

However, this equation is still difficult to solve computationally, therefore a forcefield is fitted and applied to the system.

2.2.2 Forcefield

Forcefields are composed of a set of equations and constants related to the molecular geometry and the interaction energies between atoms and are utilized to approximate the potential energy of a molecular mechanics or dynamics simulation.[11-12, 16, 25, 40] The potential energy calculation involves summing the nonbonded and the bonded interaction energies. The bonded interaction energy component is composed of stretch, bond angle, and torsional contributions as shown in Equation 4; these individual components can be calculated using Equations 5 to 7.

$$E_{bonded} = E_{stretch} + E_{angle} + E_{torsion} + E_{inversion} + E_{crossterms} + E_{restraints} \quad (4)$$

$$E_{stretch} = \sum_{a} \left[K_{a,2}(r_a - {r_a}^0)^2 + K_{a,3}(r_a - {r_a}^0)^3 + K_{a,4}(r_a - {r_a}^0)^4\right] \quad (5)$$

$$E_{angle} = \sum_{b} \left[K_{b,2}\left(\vartheta_b - {\vartheta_b}^0\right)^2 + K_{b,3}\left(\vartheta_b - {\vartheta_b}^0\right)^3 + K_{b,4}\left(\vartheta_b - {\vartheta_b}^0\right)^4\right] \quad (6)$$

$$E_{angle} = \sum_{c} \sum_{s=1}^{3} K_{c,s}\left[1 + \cos\left(s\Phi_c - {\Phi_{c,s}}^0\right)\right] \quad (7)$$

where r is the equilibrium bond distance, ϑ is the equilibrium bond angle, Φ is the equibrium torsional angle, and the force constants are represented by K. Nonbonded interaction energies include dispersion/Van der Waals and electrostatic/Coulombic contributions and can be calculated with the following formulas.

$$E_{nonbonded} = E_{VDW} + E_{Coulomb} \quad (8)$$

$$E_{VDW} = \sum_{i>j} D_e \left[2\left(\frac{\sigma^*}{r_{ij}}\right)^{12} - 3\left(\frac{\sigma^*}{r_{ij}}\right)^{6}\right] \quad (9)$$

$$E_{Coulomb} = \left(\frac{1}{4\pi\epsilon_0}\right) \sum_{i>j} \frac{q_i q_j}{\varepsilon r_{ij}} \quad (10)$$

For the VDW potential, D_e represents the energy of maximium attraction or the well depth (also known as ε), σ* equals $1.5^{1/3}\sigma$ and represents the distance of maximum attraction, and

r is the distance between atoms i and j. Lennard-Jones potential curves shown in Figure 2-5 for HCl and CO characterize these parameters.

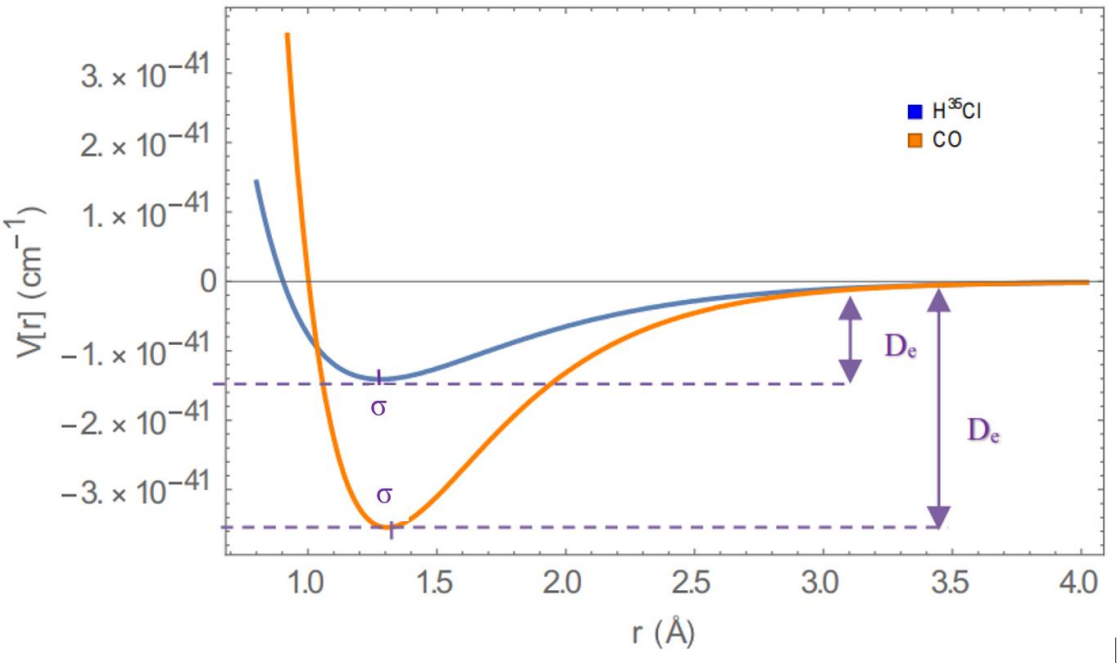

Figure 2-5 Lennard-Jones Potential Energy Curves for HCl and CO.

For the Coulombic calculation, ϵ_0 is the permittivity of the electric field in a vacuum and q represents the point charges for atoms i and j.

Other than the initial guess forcefield modification methods described in Section 3.2.6, the standard OPLS-AA forcefield with the addition of a bromine pseudo-atom was utilized; bromine atoms were utilized because of their large size. Bromine atoms replaced the two hydrogens at the ends of the substate monomers used for polymerization as shown in Figure 2-6. As a new monomer was bonded to a chain unidirectionally, a pseudo-atom was discarded. At the end of the polymerization, the pseudo-atoms on the ends were replaced with a methyl group and hydrogen atom. The bromine atomic mass was set to the mass of a carbon atom, the charge was removed, and the geometry and energetic parameters

were modified to equal those of a carbon atom plus 10% in the forcefield to mimic a pseudo-atom effect.

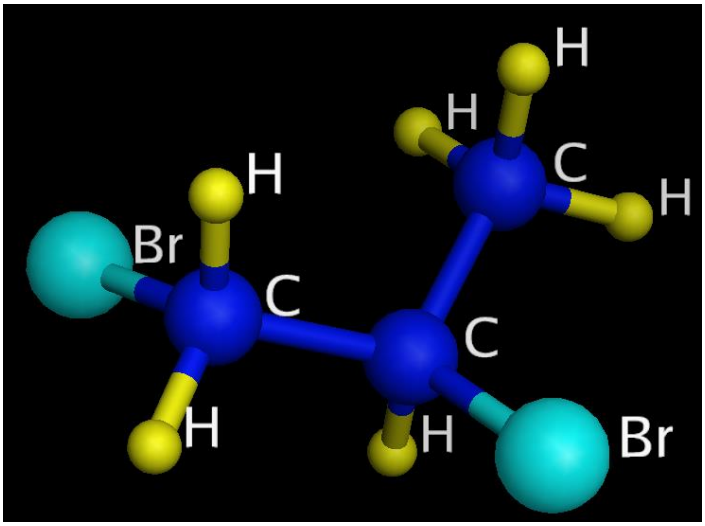

Figure 2-6 Bromine Pseudo-Atoms in aPP Monomer.

2.2.3 Energy Minimization

Since the nucleus is much more massive compared with electrons, quantum effects can be neglected and Equation 4 simplifies to Newton's Equation of Motion below

$$-\frac{dU}{dR} = m\frac{d^2R}{dt^2} \tag{11}$$

where R and m represent the position and mass of the nucleus, respectively, and U is the potential energy.[16] Note that the derivative of the potential energy is simply the force of the atom. Molecular mechanics is a subtype of molecular dynamics where the temperature is set to absolute zero, hence the kinetic energy of a system is also zero and time dependence in Newton's Equation of Motion (Equation 11) can be ignored. It consists of potential energy minimization and aims to find a set of coordinates for a molecule that

correspond with a local minimum in the simplified Newton's Equation of Motion. Various large-scale optimization methods were applied to search for a molecular conformation with zero forces on the atoms. Because the time-dependence is ignored in molecular mechanics, molecules do not diffuse and final conformations are near to the initial geometry.

There were four steps associated with the nonlinear optimization algorithms applied during energy minimization in MOE:

1. Check if any of the convergence criteria have been fulfilled in the iterative step, k; x_k is a vector of the molecules' atomic coordinates. Termination of this step occurs when any one of the following criteria is satisfied.
    a. Root Mean Square Gradient Test:

$$|grad\ U(x_k)| < e_A\sqrt{N}$$

    where U is the potential energy, e_A is a predetermined constant, and N is the number of atoms.

    b. Maximum Iterations Test: The maximum number of minimization iterations input by the user is reached.

    c. Progress Test: A pre-determined constant, T, representing the number of significant figures in the potential value, $U(x_k)$, and required to achieve a desired accuracy is reached. The following conditions must all be satisfied:

$$U(x_{k-1}) - U(x_k) < T(1 + |U(x_k)|)$$

$$|x_{k-1} - x_k| < T^{0.5}(1 + |x_k|)$$

$$|\nabla U(x_k)| \leq T^{0.3\overline{3}}(1 + |U(x_k)|)$$

2. Choose the vector search direction, p_k, indicating a nonlinear optimization algorithm based on the current potential gradient of the system. The Steepest Descent algorithm is typically first applied in the simulation since it is most effective for very high gradients and relaxing the highest energy conformations characteristic of an initial guess. Steepest Descent is a first-order optimization algorithm based on the observation that the descent of the potential occurs fastest in the direction of the negative gradient. Far from the local minimum, convergence of the potential occurs fast, but near the local minimum, convergence becomes slower and exhibits oscillatory behavior. Once the gradient becomes low enough, the Conjugate Gradient method improves upon the Steepest Descent method by repeatedly altering the gradient in the direction of the local minimum; this method is desirable for strained conformations but is ineffective for convergence. Lastly, a Newton-Raphson optimization algorithm, which is the best known optimization algorithm, is applied. Its form is described by the following iterative formula where H is the Hessian matrix:

$$x_{k+1} = x_k - H^{-1}(x_k) \cdot \nabla U(x_k)$$

3. Determine the subsequent step size, a_k, such that

$$U(x_k + a_k p_k) < U(x_k)$$

the step size is a scalar determined by using bicubic interpolation along the search direction.

4. Let x_k equal x_{k+1} and k=k+1 to repeat the scheme until Step 1 is satisfied.

Generally, a minimization normal gradient of 0.00001kcal/mol was applied with a maximum number of iterations of 5000 for all trial runs unless otherwise specified. For integrated initial guess and minimization trials, the minimization gradient varied logarithmically from 0.1 to 100 and maximum iterations were always 100. For 3-step minimization, the gradient was 0.01, 0.001, and 0.0001 for every step, respectively.

2.2.4 Molecular Dynamics

To conduct molecular dynamics is to solve Newton's Equation of Motion for a thermodynamic ensemble or trajectory of N atoms at a constant temperature or pressure. Thermodynamic properties can be computed from molecular dynamics simulations since the trajectory over time of the system provides the momentum and configurational properties of each atom through a 6N-dimensional hyperspace. The trajectory is exploratory in the sense that a conformational search over the system's available space is conducted. However, for the entire conformation space to be searched could theoretically take thousands or millions of steps. It is qualitatively assumed that a plateau in potential energy indicates a sufficiently relaxed system.

A general characteristic of polymer fluids is their propensity to respond viscoelastically to flow to reduce stress from the flow strain.[12] The energy exerted by the flow is converted to elastic energy and stored in the polymer chains. When the flow ceases, the elastic energy is then used to spring back to the unperturbed state—a macroscopic recoil. The time required for bulk conformational rearrangement or *relaxation* in real systems is relatively long—about 5500 nanoseconds for atactic polypropylene at 348K.[12, 28, 41] The Rouse model predicts the "longest relaxation time", or the time required for a

chain to diffuse a length equal to the square root of the chain end-to-end distance. For polymers of higher molecular weights, entanglement constraints between polymer coils further extends the relaxation time. The reptation theory of polymer dynamics, introduced by Pierre-Gille de Gennes, predicts that for entangled chains, the relaxation time is proportional to the molecular mass of a chain, $\tau_e \sim M^3$ and the chain self-diffusivity is related by $D \sim M^{-2}$.[42] For aPP at 348K, τ_e was 5500ns and the monomeric friction coefficient, ζ, was 2.51×10^{-10}Ns/m according to Jan van Meerveld and the Likhtman and McLeish (LM) method.[28]

Polymer chains in some commercial systems are very long, containing thousands or even millions of atoms and molecular weights of hundreds of thousands g/mol or more. The 5500ns required to relax such a large and real atactic polypropylene system would not translate for small systems such as the two, 76 monomer chains or the eight, 170 monomer chain systems modelled here. Typically, the duration of a dynamics simulations is about 1ns to 10ns. Potential energy over time curves from NVT dynamics simulations of 8, 170 monomer chains of aPP showed that potential energy plateaued even after only about 100ps. Thus, all dynamics durations were 500ps or 1ns long and sample times were either 1ps or 2ps. Figure 2-7 illustrates the relaxation of aPP polymer chains during molecular dynamics for A) 2 chains of 76 monomers and B) 8 chains of 170 monomer.

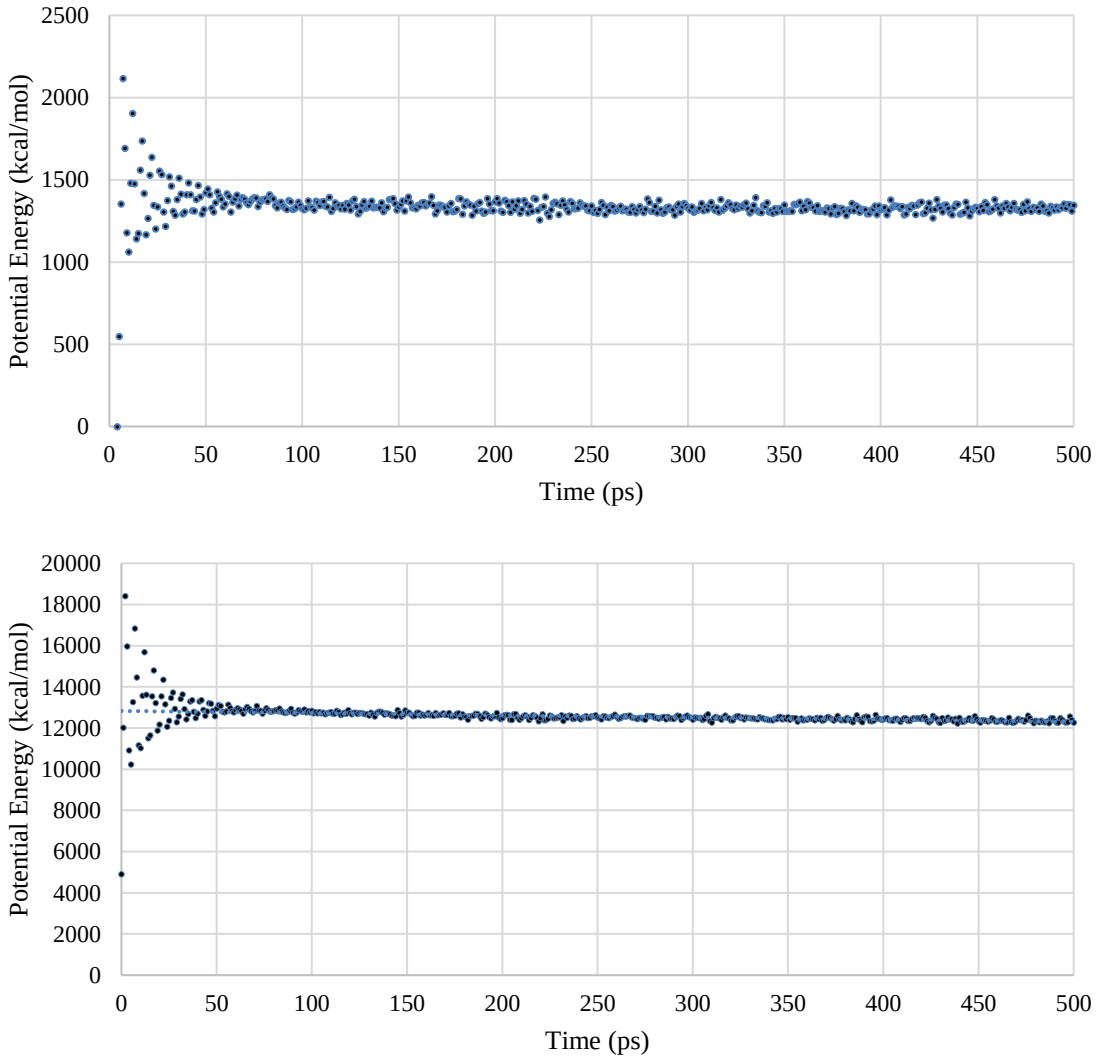

Figure 2-7 NVT Trajectories. A) 2 Chains of 76 Monomers. B) 8 Chains of 170 Monomers

A number of ensemble options are available for molecular dynamics depending on the objectives of the study. Typical ensembles include constant volume and energy (NVE), constant temperature and pressure (NPT), constant temperature and volume (NVT), and constant pressure and enthalpy (NPH). The majority of dynamics simulations conducted in this research were NVT simulations at 300K.[16, 43] For trial runs consisting of 8, 170 monomer chains the volume of the box was 124507.55Å^3 and for trial runs consisting of

2, 76 monomer chains, the volume of the box was 13954.06 $Å^3$. The density utilized was 0.765g/cc. aPS models were composed of 4 chains of 181 monomers in a box with a volume of 125,954.1675 (which is 90% of the experimental density). The aPS models were generated to determine the efficiency of the developed LRC method (Chapter 4) to calculate the electrostatic energy LRC. Some aPP NPT simulations at 300K and 1.01 bar were conducted at 0.765g/cc (100% density) and 0.6885g/cc (90% density) to evaluate the volumetric behavior of the system. We expected that the system would expand, however, at both density options, the instantaneous volume compressed to a density of 0.873g/cc. The Nosé-Hoover-Anderson thermostat equations of motion listed below were utilized for all simulations.

$$\dot{\boldsymbol{q}} = V^{-\frac{1}{3}}\boldsymbol{M^{-1}p} \tag{12}$$

$$\dot{\boldsymbol{p}} = -\nabla \mathrm{U}(\mathbf{r}) - \dot{\chi} + \dot{\eta})\boldsymbol{p} \tag{13}$$

$$\boldsymbol{r} = V^{\frac{1}{3}}\boldsymbol{q} \tag{14}$$

$$\dot{V} = 3V\dot{\chi} \tag{15}$$

$$\ddot{\chi} = \left[\frac{2K - W}{3} - PV\right] Q_P^{-1} \tag{16}$$

$$\ddot{\eta} = (2K - gkT)Q_T^{-1} \tag{17}$$

$$2K = \boldsymbol{p} \cdot \boldsymbol{M^{-1}p} \tag{18}$$

$$W = \boldsymbol{r} \cdot \nabla U(\boldsymbol{r}) \tag{19}$$

CHAPTER 3. PARAMETER AND INITIAL STRUCTURE OPTIMIZATION

Analytical theories of polymers in glass, melt, or solution states are often constructed from models of polymer chains which possess important topological, torsional, and energetic features. These models have been shown to predict polymer properties which compare well with experiment. When a feature from the model results in an accurate prediction (compared with experiment), valuable information on the properties and the features which control these properties can be gathered and used to rationalize effects and behaviors observed in real systems. Accurate parameter acquisition and an understanding of the relationship between a model's features and the resulting properties can facilitate the construction of accurate coarse-grain or mesoscale models—simplified models that represent monomers as unified atoms or 'beads' as an approximation to save computation time. Additionally, many predicted properties can illuminate similarities or differences between polymers of varying chemical structures. Collaborators at University of South Florida have been working in tandem to generate an initial guess for polystyrene, a charged and bulky polymer that is difficult to model. The results from our polypropylene models have provided insight and a solid foundation to model polystyrene, and comparison of these two very different and ubiquitous polymers may provide valuable polymer-specific information for future models such as to model the DSA of BCPs.

3.1 Parameters of Interest

There are a number of conformational, energetic, and diffusive parameters that can be calculated and compared with experiment to verify the accuracy of atomistic models including: cohesive energy density, end-to-end distance, characteristic ratio, radius of gyration, potential energy, mean square displacement during MD, and density.

3.1.1 Shape and Size Parameters

The most frequently calculated model property is the unperturbed mean-square end-to-end distance of a polymer chain, $\langle r^2 \rangle$, which describes the overall size of a polymer chain.[12, 30, 33, 44] The unperturbed mean square end-to-end distance is a measure of the distance between chain ends or the sum of skeletal bond position vectors. It is rapidly calculated using the following equation where n is the number of bonds of length l.

$$\langle r^2 \rangle = \langle \left(\sum_{i=1}^{n} l_i \right)^2 \rangle = 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \langle l_i \cdot l_j \rangle + \sum_{i=1}^{n} \langle l_i \cdot l_i \rangle = 2 \sum_{i=1}^{n-1} \sum_{k=1}^{n-i} \langle l_i \cdot l_{i+k} \rangle + nl^2 \quad (20)$$

The simplest analytical model for polymer chains is the feely-jointed chain where chains have n links of length l that are linearly fashioned and have no bond angle restrictions. The mean square end-to-end distance for such chains would be the product of n and l^2 where subscript f represents free rotation.

$$\langle r_f^2 \rangle = nl^2 \quad (21)$$

The characteristic ratio, C_∞ is a constant that describes the flexibility or stiffness of the polymer chains and depends on the bond angles, torsional states, and foliage of the chains. The characteristic ratio is the ratio between the observed mean square end-to-end distance and the freely jointed chain mean square end-to-end distance. [12, 30, 33, 44] Typical polymer characteristic ratios range from about 5.0 to 10.0. For atactic polypropylene, the characteristic ratio is approximately 6.0.[33, 45]

$$C_\infty = \frac{\langle r^2 \rangle}{\langle r_f^2 \rangle} \tag{22}$$

A Gaussian distribution of end-to-end distance vector probabilities can be obtained with the following equation:

$$\rho_n(r) = \left(\frac{3}{2\pi\langle r^2 \rangle}\right)^{1.5} \exp\left(\frac{-3r^2}{2\langle r^2 \rangle}\right) \tag{23}$$

During a molecular simulation, polymer chains are deformed causing the end-to-end distance vector distribution to assume less probable values, i.e. causing a rubber/elastic response.[12] The mean square radius of gyration, R_g^2, is a metric that is experimentally obtained through elastic scattering methods and can be readily calculated for polymer models. [12, 30, 33, 44] The radius of gyration is the sum of the positional covariances of individual particle masses from the molecule's center of mass. The mean square radius of gyration is not only valuable for molecular simulations because it can be compared with experimental data, but also because it gives information about the coil size and shape of the polymer chain. If the radius of gyration is very small, this would indicate that the polymer chain has collapsed. The radius of gyration can be calculated using Equation 24.

$$\langle R_g^2 \rangle = \frac{1}{(n+1)^2} \langle \sum_{i=0}^{n-1} \sum_{j=i+1}^{n} (\boldsymbol{r}_i - \boldsymbol{r}_j)^2 \rangle \tag{24}$$

For long amorphous chains, it is related to the end-to-end distance, r, by

$$\langle R_g^2 \rangle = \frac{\langle r^2 \rangle}{6} \tag{25}$$

Due to the simplicity of the freely jointed chain model, a parameter we have termed the collapse ratio, C_R, has been formulated which is similar to the characteristic ratio. This parameter is a ratio of the mean square end-to-end distance to the mean square radius of gyration and gives us better insight of the size of the chains.

$$C_R = \frac{\langle r^2 \rangle}{\langle R_g^2 \rangle} \tag{26}$$

Appendix B. R_g^2, LRC, and CED Calculation Code depicts the specific code and formulas utilized to calculate the mean square end-to-end distance, mean square radius of gyration, characteristic ratio, and collapse ratio.

A significant challenge for molecular simulations is capturing the accurate structural components of a polymeric system. Previous simulations have had difficulties producing chains with accurate radii of gyration, and it is a problem that we intended to tackle in this research. Table 3-1 lists average radii of gyration values for 60 isolated chain initial guesses at varying degrees of polymerization. Chains were built in 1000Å×1000Å×1000Å periodic boxes. Radii of gyration were divided by the molecular

weight to compare between chain averages with differing degrees of polymerization. The purpose of generating these isolated chains was to determine if the error associated with the mean square radius of gyration was due to the RIS model or due to system density or forcefield. Using the aPP RIS model reported by Biskup, Cantow, et al., a percent error of about 11% was associated with the radii of gyration averages and an error of about 29% is associated with the characteristic ratios. Increasing the trans proportion (represented by η) from 0.9 to 1.4 and decreasing the gauche minus proportion (represented by τ) from 0.5 to 0.4 resulted in improved isolated chain mean square radius of gyration and characteristic ratio values as shown by Table 3-1.

Table 3-1 Isolated aPP Chain Structural Values 60 Initial Guesses were Averaged at Varying Degrees of Polymerization.

	Experiment	Biskup, Cantow RIS Model			Initial Guess, Modified RIS Model		
Degree Polymerization	555	76	170	555	76	170	555
Molecular Weight (g/mol)	23300	3214.2	7169.8	23371	3214.2	7169.8	23371
R_g^2 Whole Chain (Å^2)	2672.92	274.24	632.86	1982.34	318.9	662.24	2219.07
$R_g/(MW)^{0.5}$	0.3387	0.2921	0.2971	0.2912	0.315	0.3039	0.3081
Percent Error (%)	-	11.2641	9.3916	11.592	3.1799	6.9375	5.4718
First and Second Half p-Values	-	0.0948	0.3846	0.8401	0.6751	0.8814	0.0423
Characteristic Ratio ($<r^2>/nl^2$)	6.1	4.49	4.19	4.34	5.84	5.02	5.70
Collapse Ratio ($<r^2>/<Rg^2>$)	6.02	5.75	5.2	5.13	6.17	5.56	6.02

3.1.2 Density and Molecular Weight

Density is an important parameter to control and monitor since it can significantly impact the resulting structures. Density is one of the few parameters that can be altered to control the properties of an initial guess. For NVT simulations, the density is controlled by setting the volume of the periodic box to achieve a final density equal to the experimental density. Most of the molecular simulations including fifteen of the seventeen methods depicted from Figure 3-3 to Figure 3-13. The length and quantity of chains also significantly affects the overall shape, size, and entanglement of the system—generally, heavier and a greater number of chains results in a system that is less homogeneously entangled and more difficult to relax.

3.1.3 Diffusion and Mean Square Displacement

Self-diffusivities, D, are rates that gauge the relative average diffusion and the entanglement of chains. A highly entangled polymer system will diffuse less during molecular dynamics, and good entanglement is a desirable quality for accurate model systems.[42, 46] Though, the duration of all our molecular dynamics was only between 0.5-1.0ns such that diffusivities could not be compared with experiment, self-diffusivities of the various models proposed could be compared and shed light on the relative entanglement a particular model provided. Self-diffusivities were calculated from the mean-square displacement (MSD) of chains after MD using the equations below.

$$MSD \equiv \langle |x(t) - x(0)|^2 \rangle = \frac{1}{N} \sum_{i=1}^{N} |x_i(t) - x_i(0)|^2 \tag{27}$$

$$D = \frac{MSD}{2nt} \tag{28}$$

The n in Equation 28 represents the number of dimensions in Euclidean space. For mean square displacement calculations, the 3-dimensional positions of the atoms of each chain after a given minimization protocol (t=0) were subtracted from the 3-dimensional positions of the same atoms after 500ps of molecular dynamics. These differences were squared, summed, and divided by the number of atoms, N, for each dimension (x, y, z) to obtain the mean-square displacement.

3.1.4 Energetic Parameters

The cohesive energy, E_{coh}, is a parameter that represents the energy required to remove a single chain from the molecular system and make it a completely isolated chain.[12, 21, 30, 33, 44, 47] In our model systems, a single chain was surrounded by other chains, and the cohesive energy was the non-bonded energy between each chain and the surrounding chains excluding itself. For aPP, the cohesive energy was the energy of non-bonded interactions and was obtained by subtracting the total energy of the system of chains, U_{tot}, from the sum of the energy of the individual or isolated chains, $U_{tot,isolated}$.

$$E_{coh} = U_{tot,isolated} - U_{tot} \tag{29}$$

The total isolated chain energy was calculated by iteratively selecting chains and setting all other chain energies to 'inert', then calculating that chain's potential energy while cutoff was disabled. Each isolated chain energy was summed to give the total isolated chain energy.

The cohesive energy density, CED, is the cohesive energy per unit volume of mols (e.g. J/cm^3), and the Hildebrand solubility is simply the cohesive energy density squared.

$$\delta = \left(\frac{E_{coh}}{V}\right)^2 \tag{30}$$

We postulated that the cohesive energy density would be related to the entanglement of the system, since the more entangled a system, the more each chain interacts with its surroundings and increases in cohesive energy. Figure 3-1 describes the relationship between 3-dimensional mean square displacement—which has been used as a metric for entanglement in previous experimental research—and cohesive energy density for twenty-nine aPP structures created using the 'Original' method, which is the normative method described in Chapter 2.[42, 46-47] If this postulate were true, we would expect that there would be a downward trend; that is, as mean square displacement of the chains increases, cohesive energy density would decrease. As shown in Figure 3-1, there is indeed a slight downward trend as mean square displacement increases for each dimension indicating that with greater cohesive energy density a system is more entangled.

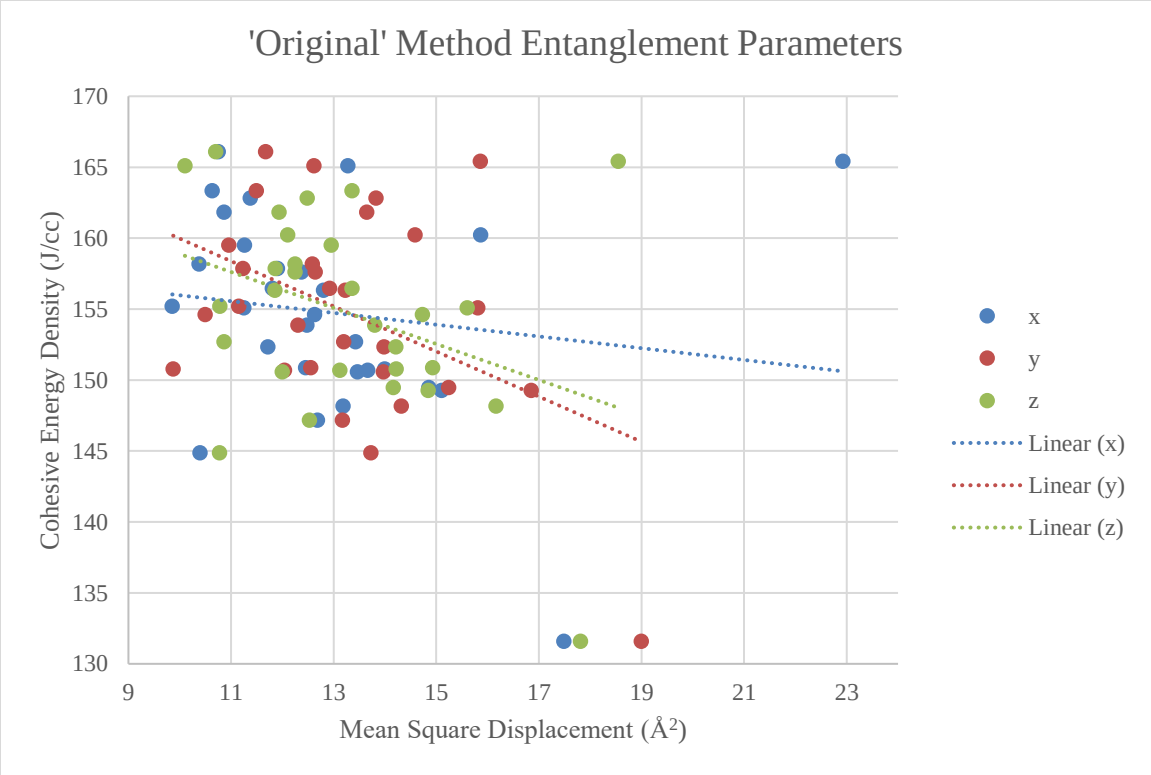

Figure 3-1 CED as a Metric for Entanglement

Given that cohesive energy density was inversely related to mean square displacement and directly related to entanglement, we also postulated that the cohesive energy density would be related to the mean square radius of gyration because the more entangled the chain, the more extended we predicted the chains to be rather than collapsing into themselves. If this postulate were true, we would expect a positive correlation between cohesive energy density and mean square radius of gyration. Figure 3-2 depicts plots relating the mean square radius of gyration and cohesive energy density values of twenty-nine aPP structures generated with the 'Original' initial guess method and subjected to MM and MD procedures described in Chapter 2. Figure 3-2A) has data points for every chain in every structure. Since only a single cohesive energy is associated with each structure and each structure has eight chains, groups of eight data points share the same cohesive

energy density. The mean square radius of gyration for the eight chains of a given structure were averaged and plotted against the structure's corresponding CED in Figure 3-2B). In both plots, there seems to be a slight upward trend. However, after further statistical analysis, this trend is not significant. Ergo, CED is not directly related with mean square radius of gyration. This correlation is a function of the chain morphology. For example, a crystalline chain is likely to have a more significant correlation between these variables because an elongated chain can maximize CED in a crystalline morphology. However, this correlation is not useful for the amorphous chains studied here.

A high value of total potential energy of a system after energy minimization and 1.0ns of NVT MD may indicate that the structure is unfavorable and may be stuck in an unfavorable energy well. Potential energy should be monitored throughout the simulations.

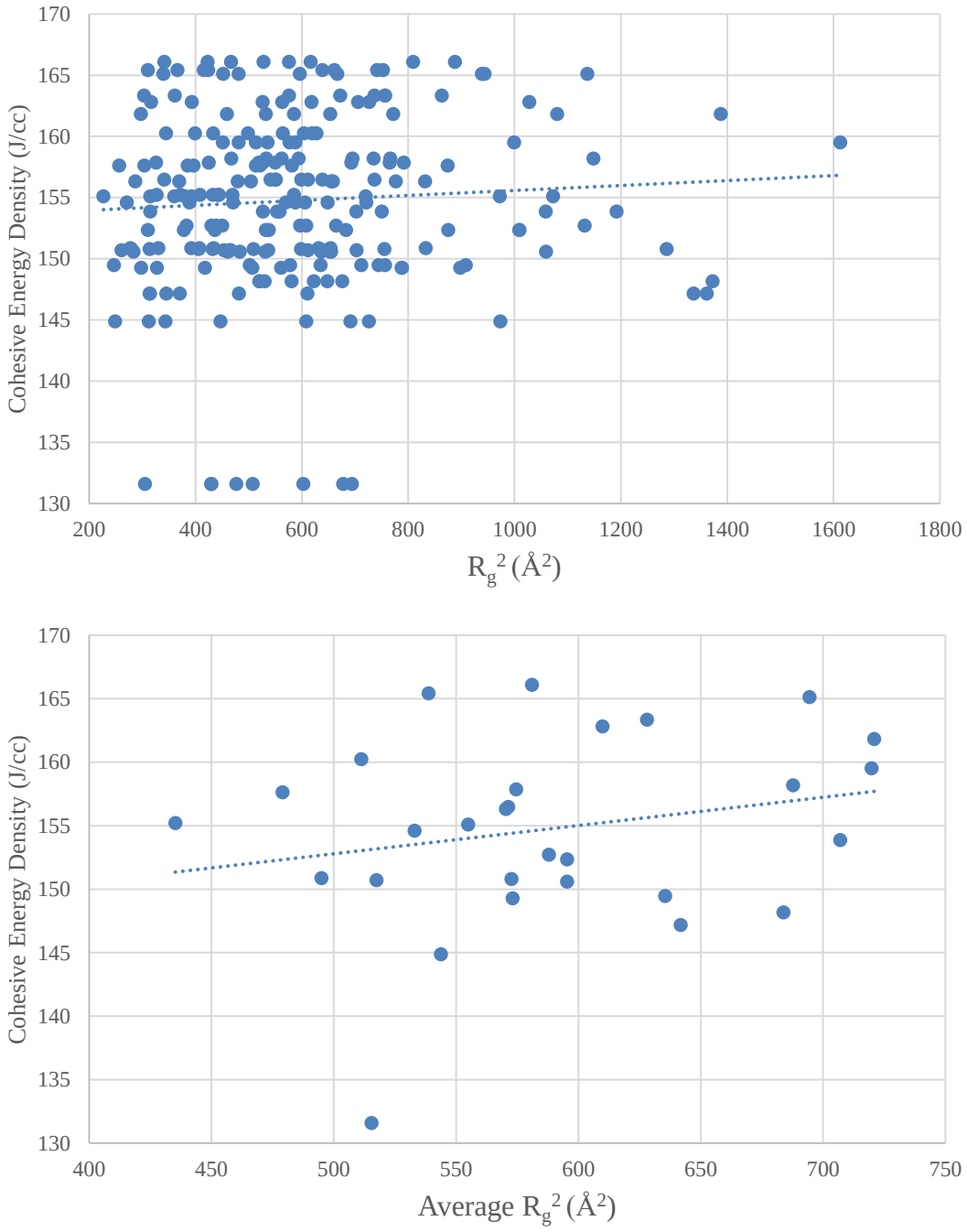

Figure 3-2 Relationship between R_g^2 and CED

3.2 Initial Guess Optimization

Generating an accurate initial guess is vital for molecular simulations because, as discussed in the previous chapter, minimization drastically affects the energetic components of the structure but only marginally affects the conformational components of the structure.[24] Thus, generating an initial guess with as accurate as possible structural components was critical for further polymer property acquisition and for final models to be accurate. Six initial guess generation methods were designed and compared to the 'Original' method described in Chapter 2 and experiment to define an optimal initial guess generator. These initial guess generation methods included reducing the density, applying minimization during initial guess generation, multi-step minimization, and updating the density during polymerization (Constant Density method). Figure 3-3 to Figure 3-8 depict parameters obtained after applying these initial guess methods, energy minimization, and NVT molecular dynamics. Structure filtration and forcefield optimization methods were also investigated and resulted in seventeen total candidates to compare as shown from Figure 3-9 to Figure 3-13; these depict parameters obtained after energy minimization since MD did not significantly improve parameters (Table 3-2). The difference between the parameters obtained after MM and MD for the first four methods utilized was very small (less than 4%, Table 3-2), and the computational cost of running MD is much greater. Additionally, the trends observed in the parameters were the same before and after dynamics. Therefore, the structures generated using the remaining two methods were only subjected to minimization and compared to post-MM structures associated with the other initial guess methods. The mean square displacement (and diffusion) was only compared amongst the first four methods, since these parameters can only be evaluated after MD.

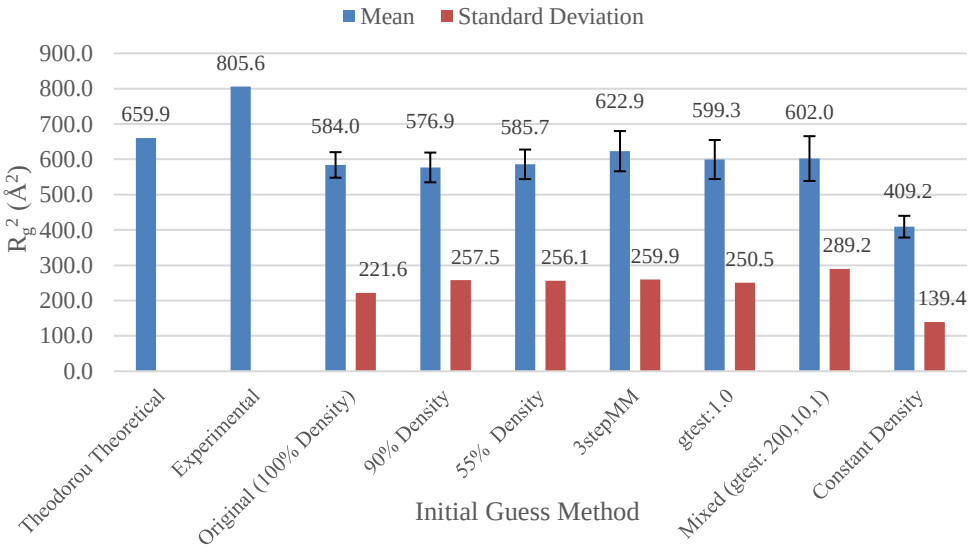

Figure 3-3 Post-Dynamics R_g^2 Comparison for Various Initial Guess Methods, Average over all Structures and Chains.

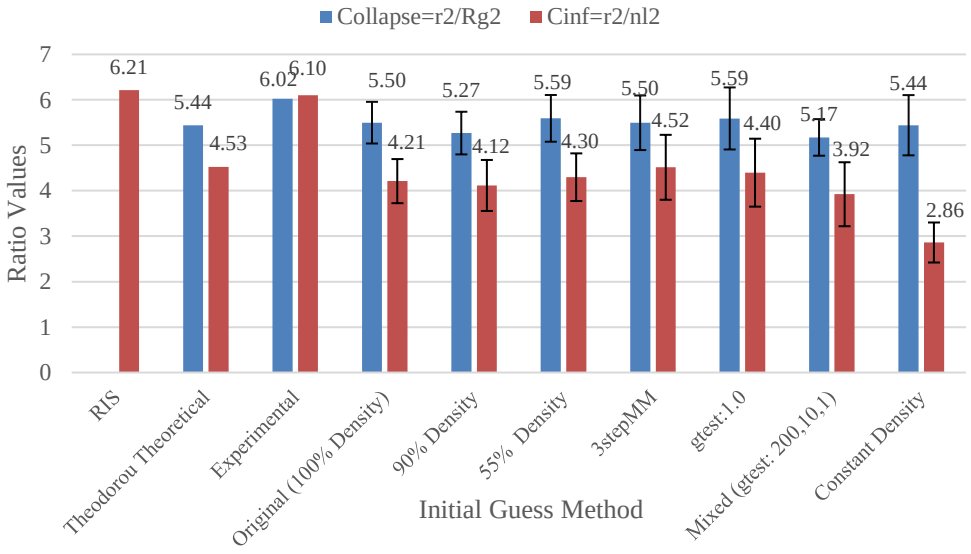

Figure 3-4 Post-Dynamics Collapse and Characteristic Ratio Comparison for Various Initial Guess Methods

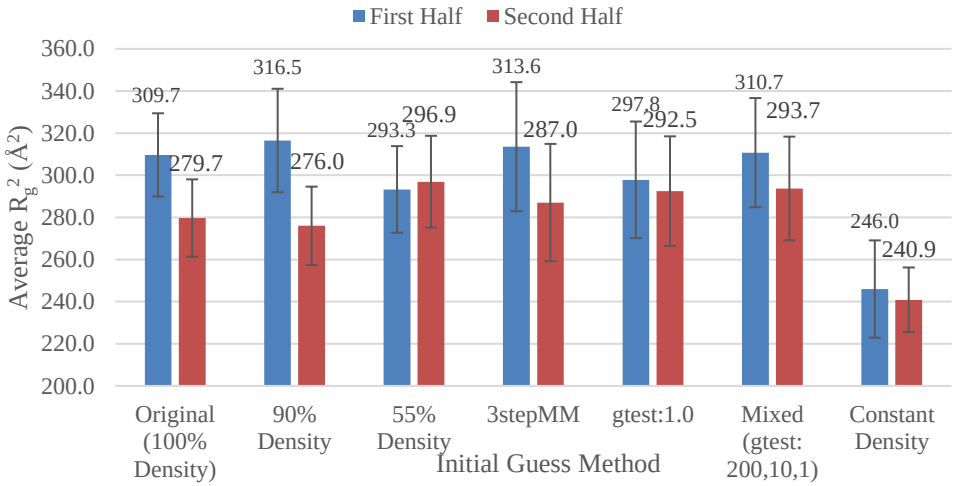

Figure 3-5 Post-Dynamics Chain Half R_g^2 Comparison for Various Initial Guess Methods

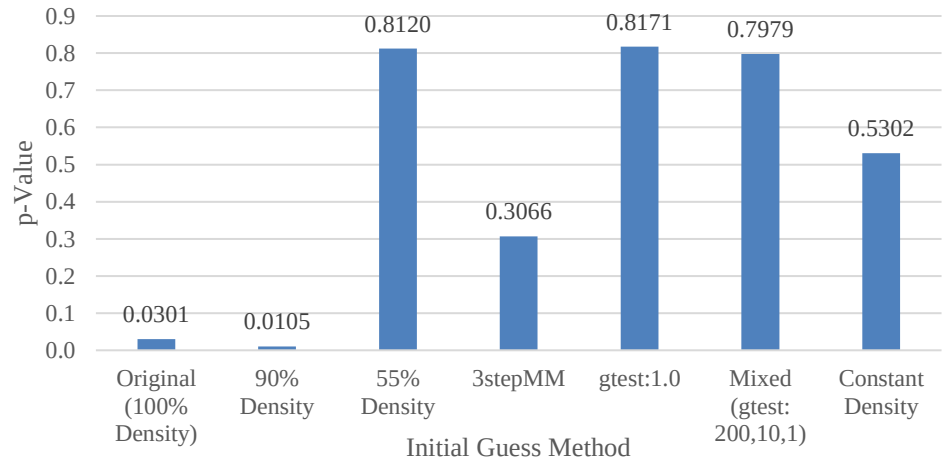

Figure 3-6 Post-Dynamics Chain Half R_g^2 p-Value Comparison for Various Initial Guess Methods

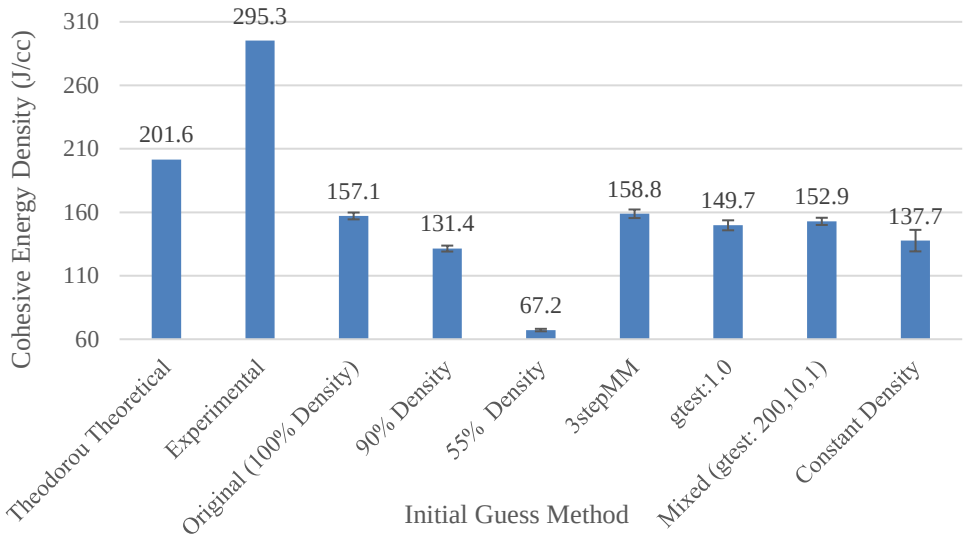

Figure 3-7 Post-Dynamics CED Comparison for Various Initial Guess Methods

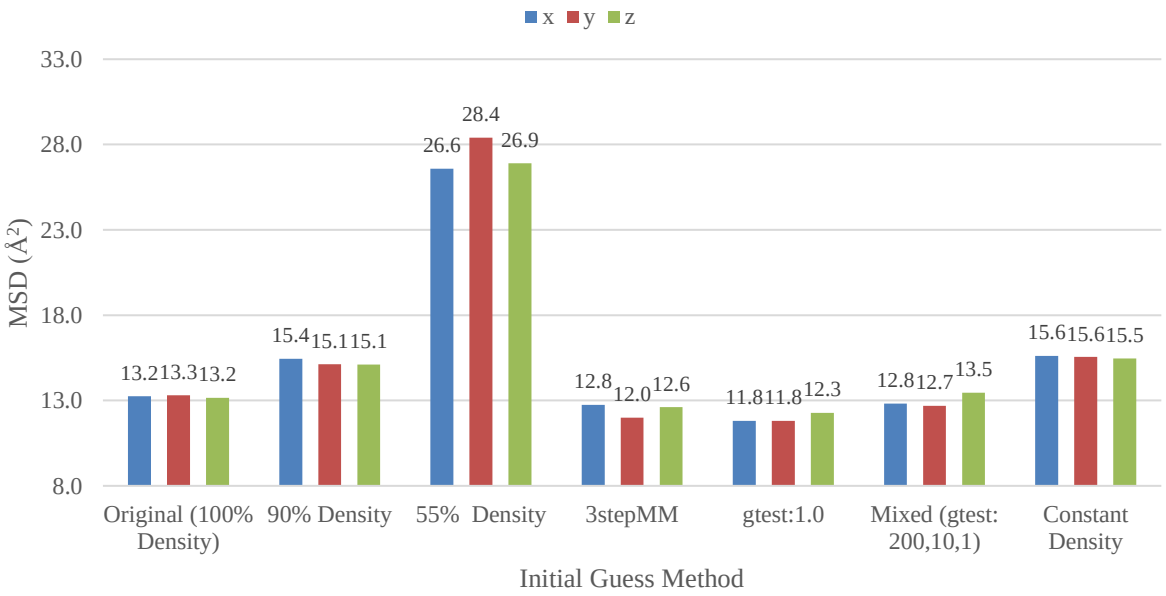

Figure 3-8 Mean Square Displacement After 500ps of Dynamics for Various Initial Guess Methods

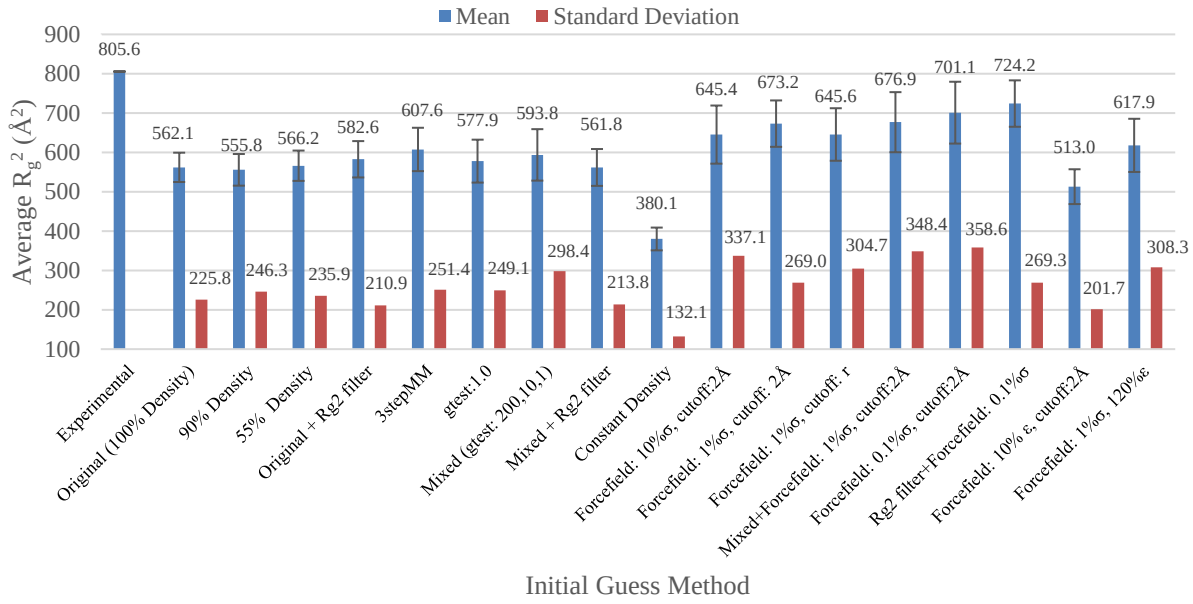

Figure 3-9 Post-Minimization R_g^2 Comparison for Various Initial Guess Methods

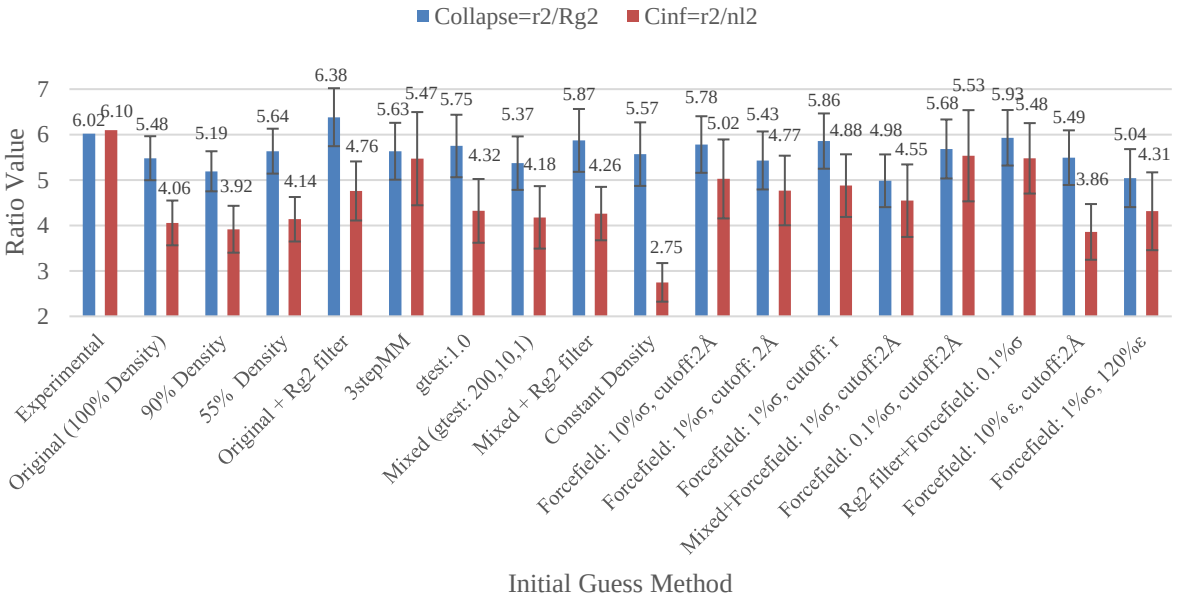

Figure 3-10 Post-Minimization Collapse and Characteristic Ratio Comparison for Various Initial Guess Methods

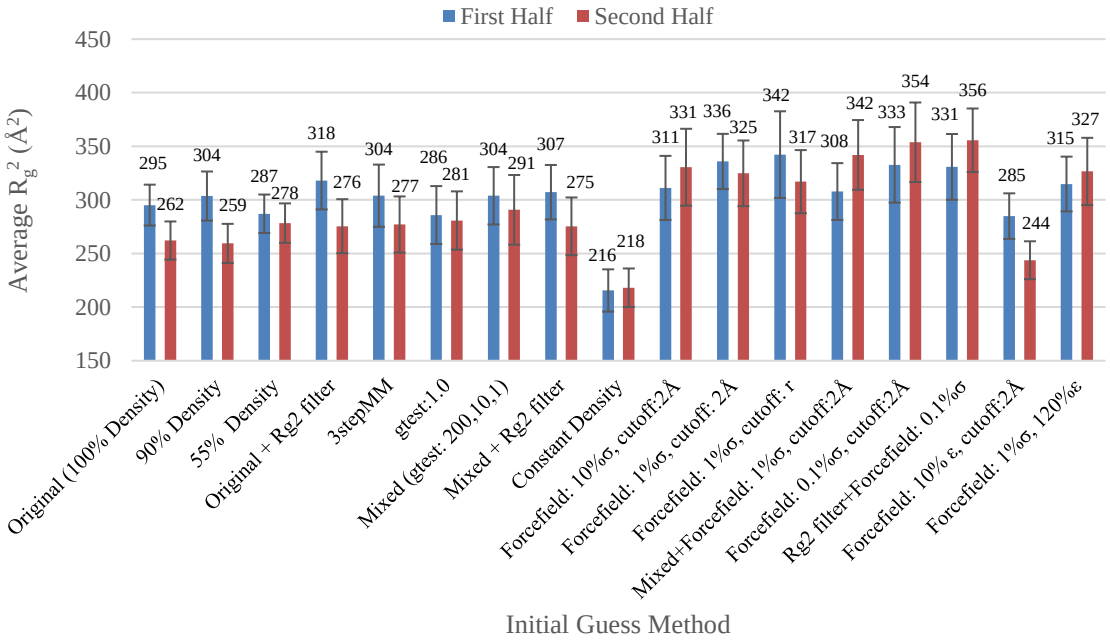

Figure 3-11 Post-Minimization Chain Half R_g^2 Comparison for Various Initial Guess Methods.

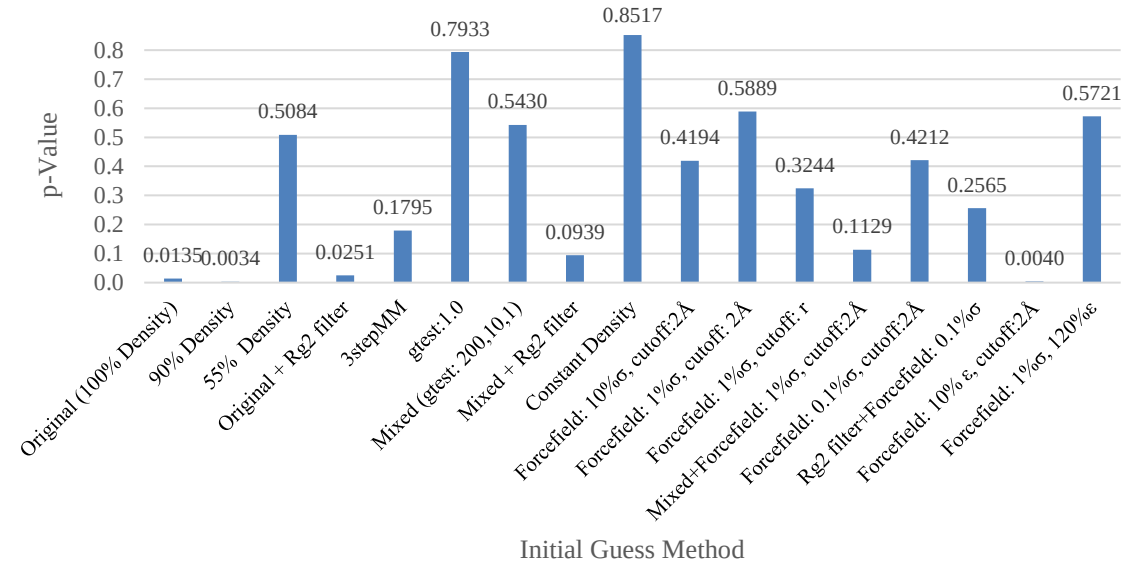

Figure 3-12 Post-Minimization Chain Half R_g^2 p-Value Comparison for Various Initial Guess Methods.

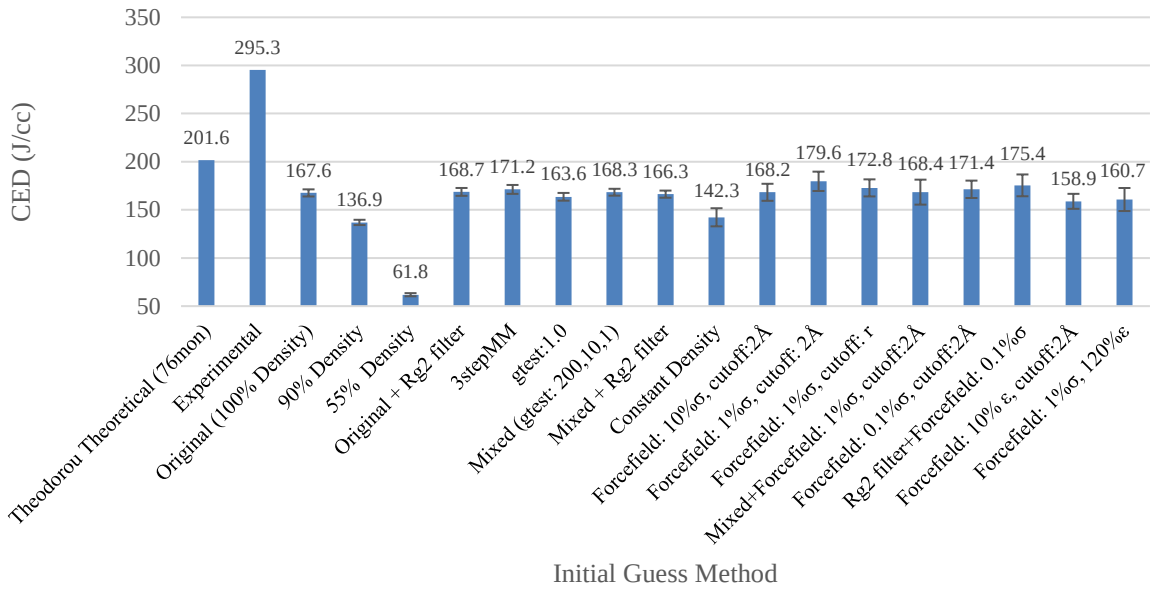

Figure 3-13 Post-Minimization Cohesive Energy Density Comparison for Various Initial Guess Methods.

Table 3-2 Percent Difference in Parameter Values after Molecular Mechanics and Dynamics Using Six Different Methods.

	R_g^2, whole	R_g^2, first half	R_g^2, second half	CED	C_∞	C_R
Original (100%Density)	0.039	0.049	0.0667	-0.0626	0.0032	0.0384
90% Density	0.038	0.0424	0.0638	-0.0405	0.0147	0.051
55% Density	0.0344	0.0216	0.0665	0.0869	-0.0075	0.039
3stepMM	0.0252	0.0316	0.036	-0.0726	-0.0244	-0.1744
gtest:1.0	0.0371	0.0416	0.0415	-0.0845	-0.0276	0.0178
Mixed (gtest:200,10,1)	0.0139	0.0225	0.0101	-0.0919	-0.0369	-0.0611
Constant Density	0.0766	0.1412	0.1044	-0.0324	-0.023	0.0413

3.2.1 3-step Minimization

Previous research conducted by Theodorou and Suter reported that a 3-step minimization was required after the generation of the initial guess.[14, 21, 23] Later research in polystyrene even recommended a 41-step minimization process. [37]The 3-step minimization process involves application of three modified forcefields and the steps/forcefields are described below:

1. Dispersion parameters, ε, for hydrogen atoms were set to zero and particle size parameters, σ, for all carbons were set to one-half their normal value. The cutoff was set to the smallest σ value associated with a carbon atom, and the system was minimized to a gradient magnitude of 0.001 kcal/mol·Å. Effectively, chains were 'phantom-like' and able to pass through each other, relaxing out the worst conformation states.
2. Same as Step 1 except that σ values for carbons were restored to their normal values.
3. Apply the normal forcefield and minimize at a Euclidean norm gradient of 0.0001.

This method provided a small improvement in R_g^2, C_∞, C_R, CED, and MSD compared with the 'Original' method.

3.2.2 Density and Asymmetry of Chains

A critical component of initial guess generation is that all the chains be polymerized synchronously such that chains are sufficiently interwoven and proper entanglement can

be realized. However, a significant issue arises when the chains are about 75% polymerized. Since the periodic cube volume is held constant throughout the duration of the polymerization and the molecular weight of the chains increases to reach the final desired density at the end of the initial guess generation, this means that the free volume has decreased. At about 75% polymerization, the ends of the chains [the remaining 25%], consequently assemble into compact conformations.[25] This would result in the first portion of each chain being longer than the last portion, and chains growing asymmetrically. This can be proven by comparing the mean square radius of gyration of each half of the chains and quantifying this chain half difference with a p-value.

The chain end collapse phenomenon can be attributed to the self-avoiding nature of the chains and forcefield.[25] If the self-avoiding feature of the forcefield is removed by setting the dispersion energy parameter, ε, to zero, 'phantom' chains would result which have the ability to intersect or superpose other chains. The collapse of chain ends may occur to some extent in real polymerizing systems; however, polymerization in real systems occurs at a much slower rate and ends have more time to relax as they grow. Additionally, the observed mean square radius of gyration values were significantly lower than experiment, especially at higher densities. As such, extension of the second halves of the chains (by improving chain symmetry) may improve the mean square radius of gyration of the whole chain.

The molecular weight of the chains may have a significant impact on mean square radius of gyration and symmetry of chains. Previous research reported by Theodorou on aPP only included very small models (2 chains of 76 monomers), and the mean square radius of gyration was about 20% lower than the experimental mean square radius of

gyration (Theodorou: 295.84 Å^2, Experimental: 368.73). [14-15, 21-23, 29] Molecular weights above the molecular entanglement weight such as the 8 chains of 170 monomers systems constructed in this modelling research resulted in even greater difficulties.

To avoid this collapse of chain ends, USF collaborators proposed building the aPP initial guess at very low density (such as 10% of experimental density) such that enough free volume would be available for the chain ends to grow, then shrinking the periodic box to reach experimental density using NPT molecular dynamics. However, as shown qualitatively in Figure 3-14, this method resulted in minimal entanglement, since the act of building the chains at experimental density promoted entanglement, and chains built at too low of a density collapsed into themselves before being brought together at experimental density. As a result, parameters such as potential energy, cohesive energy, radius of gyration, and characteristic ratio were far from experimental values which indicated a very unstable and inaccurate final structure.

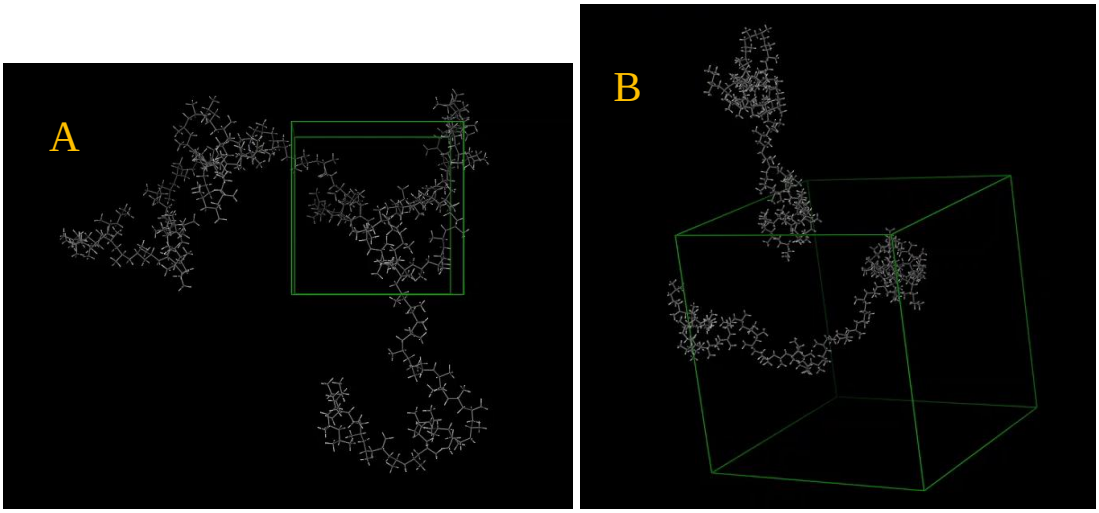

Figure 3-14 Shrinking Periodic Cube

Figure 3-14A depicts an initial guess (2 chains of 76 monomers each) that was generated at experimental density, 0.765g/cc. There is clearly some entanglement. Figure 3-14B depicts the same initial guess built at 10% of the experimental density, 0.0765g/cc.

The extent of asymmetry of aPP chains was examined by statistically analyzing the mean square radius of gyration of 29 structures of 8 chains with 170 monomers per chain at 100% density. Thus, 29 structures with 8 chains per structure resulted in 232 mean square radius of gyration data points from each individual chain to compare. Mean square radius of gyration values were calculated for each half of all the chains as well as the mean square radius of gyration of each whole chain. The three columns of Table 3-3 report the statistical values obtained from comparing chain halves after initial guess generation, energy minimization, and NVT molecular dynamics. As suspected, the average of the first half of all the chains was significantly greater than the second half of all the chains, and p-Values were below an alpha confidence of 0.05 indicating that each chain half was significantly different from each other and that the second half of each chain was collapsing into itself. Neither energy minimization nor dynamics ameliorated this effect.

Table 3-3 Lower Density Method Statistical Values. Twenty Structures were Comparison of Varying Densities and Relaxation Stages.

	Initial Guess			After MM			After MD		
	100% Density	**90% Density**	**55% Density**	**100% Density**	**90% Density**	**55% Density**	**100% Density**	**90% Density**	**55% Density**
Whole Chain:									
Mean	540.08	559.74	518.09	566.80	551.00	564.61	586.97	572.89	586.39
stdev	238.31	240.10	242.87	232.19	242.49	237.17	229.30	258.94	261.20
stdev:Mean	0.44	0.43	0.47	0.41	0.44	0.42	0.39	0.45	0.45
Median	525.80	534.50	476.31	534.42	500.98	538.05	572.83	506.09	560.34
Skew	1.12	0.94	1.23	1.19	1.33	0.90	1.17	1.37	0.83
Chain Halves:									
Mean, First	268.60	286.59	280.83	295.72	307.14	289.56	312.13	321.79	296.05
Mean, Second	231.51	269.61	227.62	261.30	254.88	278.52	278.66	272.26	299.57
p-Values	0.0132	0.2576	0.0009	0.0197	0.0014	0.4533	0.0303	0.0052	0.8394
% decrease in length	-0.16	-0.06	-0.23	-0.13	-0.21	-0.04	-0.12	-0.18	0.01

Furthermore, p-Values decreased as the sample size [number of structures compared] increased, which occurs only when the null hypothesis is false. This further validated our hypothesis that initial guess chain halves were significantly different and that the second halves of the chains were collapsing during initial guess generation. Table 3-4 illustrates the decrease in p-values as the number of structures compared (or sample size) increased.

Table 3-4 p-Values and Sample Size

Number of Structures	Initial Guess	After MM	After MD
5	0.95956	0.58593	0.88287
6	0.93925	0.66919	0.90465
7	0.77442	0.93605	0.93437
8	0.63331	0.76385	0.54861
9	0.81039	0.71644	0.87517
10	0.84573	0.70217	0.90719
11	0.54804	0.52653	0.88142
12	0.17347	0.17022	0.34319
13	0.16462	0.15497	0.33881
14	0.15927	0.16873	0.33498
15	0.07552	0.08721	0.17054
16	0.09792	0.11884	0.24457
17	0.06984	0.07862	0.18559
18	0.02529	0.03057	0.07286
19	0.01181	0.01877	0.04553
20	0.00813	0.01299	0.03514
21	0.00336	0.00703	0.01985
22	0.00170	0.00354	0.00969
23	0.00085	0.00150	0.00440
24	0.00122	0.00201	0.00456
25	0.00061	0.00106	0.00252
26	0.00054	0.00118	0.00260
27	0.00029	0.00059	0.00127
28	0.00088	0.00143	0.00378
29	0.00133	0.00185	0.00573

In order to satisfy both initial guess generation requisites (the chain entanglement effect of the ‘Original’ method and the symmetric chain effect of a very low density model), we built the initial guesses at varying densities to determine a moderate and optimal density. The goal was to generate chains that were both entangled and more symmetric by allowing the chain ends to extend at a moderately lower density and as a result also increase the total, whole-chain mean square radius of gyration. 20 initial guesses were generated at

90% density and 55% density. Table 3-3 compares mean square radius of gyration statistical values at 100%, 90%, and 55% density. During the initial guess generation phase, there was a small increase in R_g^2 of the whole chains from 100% density to 90% density and asymmetry decreased as depicted by the p-values. P-values below the α value of 0.05 indicated a significant difference between each half of the chains. Then from 90% to 55%, R_g^2 and p-values worsened. This may indicate that the density reduction was somewhat beneficial for the structural components of the initial guess, but at too low of a density, the entire chain lengths collapse due to no interactions between other chains. Additionally, any positive effect on R_g^2 at 90% in the initial guess is lost after minimization and dynamics. Therefore, these simulations show that density control had a minimal effect on the shape and size of the aPP system, and alternative methods (described in later sections) were required.

The value associated with the 100% density was 0.765 g/cc at 300K and was obtained from Eckstein, et al. This density is lower than the extrapolated density of 0.8682g/cc at 300K obtained from Mark E. James' polymer handbook.[21, 27, 30] Table 3-5 reports the results of NPT simulations at various initial densities. Final densities were all similar to each other and the experimental density which indicated that the forcefield was surprisingly accurate.

Table 3-5 Densities Observed after NPT Simulations.

8 Chains, 170 Monomers

	Initial Density (g/cc)	Final Density (g/cc)	Initial Boxedge (Å)	Final Avg. Boxedge (Å)
100pDensity	0.765	0.8734	49.967	47.7696
90pDensity	0.6885	0.8744	51.719	47.7515

2 Chains, 76 Monomers

	Initial Density (g/cc)	Final Density (g/cc)	Initial Boxedge (Å)	Final Avg. Boxedge (Å)
90pDensity	0.6885	0.8542	24.936	23.2061
100pDensity	0.765	0.8607	24.075	23.1477
114pDensity (experiment)	0.873	0.8697	23.038	23.0673
130pDensity	0.9945	0.8671	22.059	23.0899
140pDenisty	1.071	0.8675	21.521	23.0871

Though the density had virtually no effect on the structural components (i.e. R_g^2, C_∞, and C_R), it did have a large effect on cohesive energy density, CED, and mean square displacement, MSD, as shown by Figure 3-7, Figure 3-8, Figure 3-13, and Table 3-6. Fifteen structures (2 chains of 76 monomers) were generated at various densities and subjected to minimization and NVT dynamics. Table 3-6 below shows the CED values obtained from applying these densities.

Table 3-6 Post-MD (NVT) Atactic Polypropylene Parameters Obtained from 15 Structures of 2 Chains and 76 Monomers at Various Densities.

	Density (g/cc)	CED (J/cc)	Ds (cm^2/s) × 10-7	R_g^2 ($Å^2$)	$R_g/(MW)^{0.5}$	Error	Chain Half p-values
Biskup, Cantow RIS Model:	0.688	145.40±5.07	17.44±1.29	222.95±29.07	0.2634	0.234	0.3033
	0.765	180.70±5.63	11.44±0.89	224.38±27.68	0.2642	0.2301	0.5501
	0.873	249.53±6.30	5.09±0.29	215.51±28.95	0.2589	0.2551	0.2603
	0.995	309.22±9.60	2.24±0.14	231.71±37.57	0.2685	0.2105	0.0957
	1.071	255.08±16.10	1.71±0.15	233.47±31.52	0.2695	0.2059	0.0016
	1.148	187.56±18.21	1.72±0.14	241.71±46.28	0.2742	0.1851	0.0249
Modified RIS:	0.765	152.62±5.98	10.83±0.82	294.24±52.39	0.3026	0.0742	0.0941
	0.873	212.20±4.61	5.00±0.42	226.21±23.46	0.2653	0.2251	0.0994
	0.995	263.52±10.07	2.34±0.15	238.03±32.71	0.2721	0.1943	0.0102

The above data show that the density that offers the best CED values was a density of 0.995g/cc which is well above either referred density (0.765g/cc or 0.873g/cc). Experimental values for CED are between 282.24 J/cc and 353.44 J/cc, therefore utilizing a density of 0.995g/cc would result in a zero CED percent error.[44] Interestingly, even higher densities, such as 1.071 and 1.148g/cc also yield somewhat reasonable CED values. The diffusivity values (which describe chain entanglement) decreased with increasing density, since at higher densities, chains are able to entangle more. Surprisingly, varying the density had no effect on the mean square radius of gyration. p-values generally decreased with increasing density. This result make sense because at high density and low free volume, chain ends will collapse.

3.2.3 Monomer Minimization

Another approach to creating structurally accurate initial guesses was to integrate global minimization into the initial guess generation code that would iteratively minimize the chains after a monomer was dealt to each chain. The idea was to simultaneously choose the lowest energy conformation for the growing chain ends as well as for the chains surrounding the ends using global minimization; as the density increases during polymerization, the previously self-avoiding and collapse effect of the chain ends would be attenuated since the surrounding chains would also be relaxing and creating free volume for the growing chains. Tuning of the minimization to a specified gradient produced initial guesses with varying total potential energies, radii of gyration, self-diffusivities, and cohesive energy densities. Therefore, a preliminary investigation was conducted to choose the optimal gradient for initial guess generation. The methods are differentiated by the target for the Euclidean norm of the energy gradient (in kcal/mole Å) during a single or a multistage minimization after each monomer is added. We use the term gtest to indicate this gradient targe in this work. The following gradients were tested and two initial guesses were generated for each: 0.1 kcal/mol, 1.0 kcal/mol, 10.0 kcal/mol, 100.0 kcal/mol. Additionally, a 'Mixed' gradient method was utilized which included a gradient of 200.0 kcal/mol for every one of the first 34 monomers, 10.0 kcal/mol for every one of the next 68 monomers, and 1.0 kcal/mol for the last 34 monomers. The goal of this minimization-during-polymer growth hybrid model was to show that gradient tuning could produce more symmetric chains or be at least the same as methods that used only a gradient of 0.1kcal/mol, 1.0kcal/mol or 10.0kcal/mol. Large minimization gradients would be applied only to relax out large conformational instabilities and small minimization gradients would

be applied at high densities when the chain ends begin to collapse into themselves. Figure 3-15 shows the parameters obtained for each gradient category. Just based on the comparison of two structures, there was not much difference between the gradients, 'Mixed', 0.1kcal/mol, 1.0kcal/mol, 10.0kcal/mol, but there was a significant difference between 100.0kcal/mol and the others. Therefore, for further investigation the 'Mixed' and 1.0 gradients were employed. Further trials that applied this method were on systems of 8 chains and 170 monomers and the maximum number of iterations was 100.

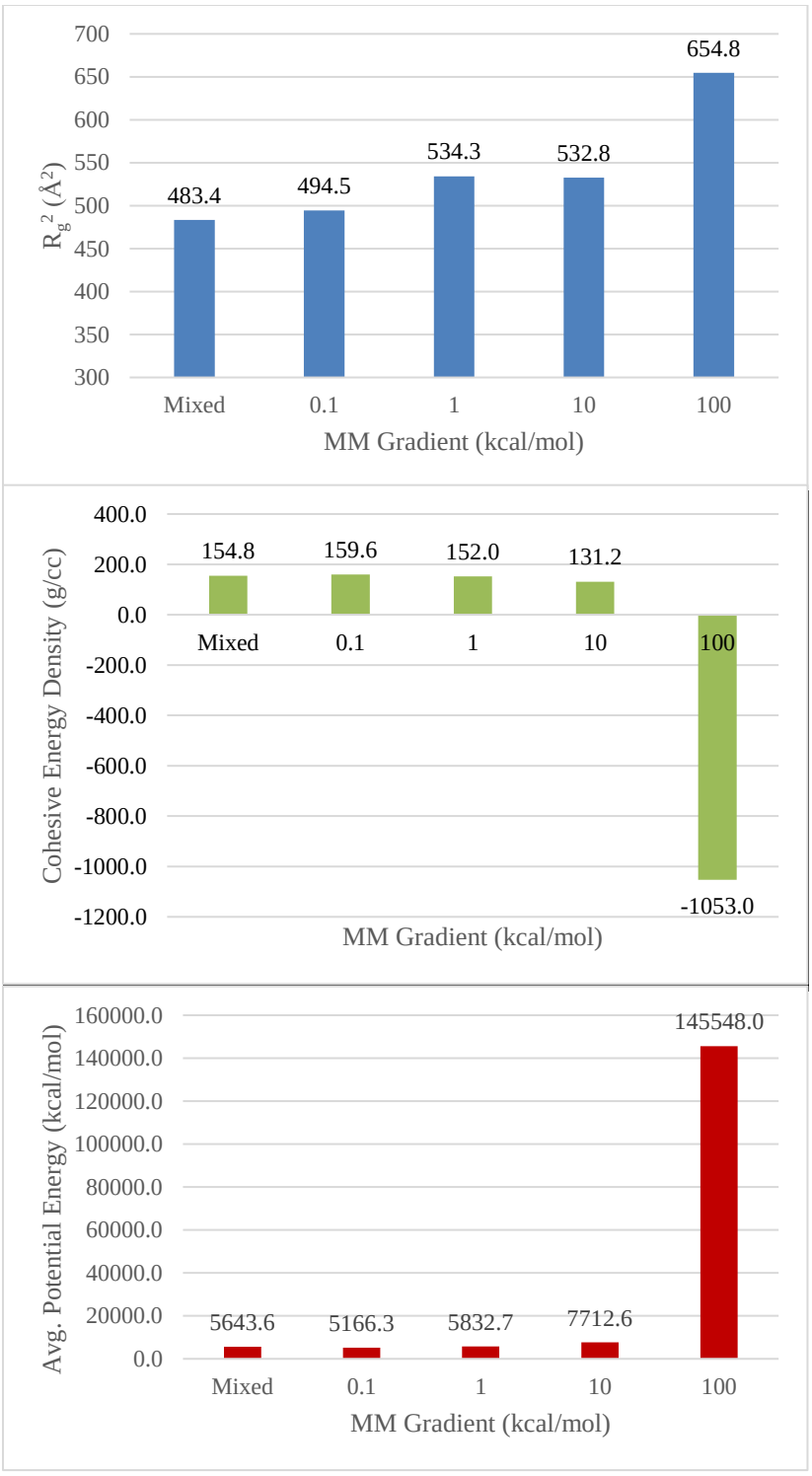

Figure 3-15 Preliminary Gradient Determination for Polymerization-Minimization Hybrid Method. A) Mean Square Radius of Gyration Comparison. B) Cohesive Energy Density Comparison. C) Potential Energy Comparison.

The whole-chain R_g^2, C_∞, C_R, CED, and MSD were about the same for the 'Mixed' and 'gtest:1' (minimization gradient of 1.0kcal/mol) as the 'Original' method. However, there was significant improvements in the symmetries of the chains as shown Figures 3-5, 3-6, 3-11, and 3-12 using these methods. After the constant density method (which did not have acceptable values for all other parameters of interest), the 'gtest:1', 'Mixed', and 'Forcefield: 1%σ, cutoff:2Å' methods offered the best chain symmetry options.

3.2.4 Constant Experimental Density

Due to the asymmetry of the chains and the effect density has on the energetic components of the structure, a method which updates the density throughout the polymerization process was developed. This method would alter the size of the periodic box every time a new monomer was bonded to maintain experimental density throughout the entire polymerization process. The aim was to promote entanglement at the beginning of the chains and expand the box as the chains grew larger and began to self-avoid. As shown in Figures 3-3 and 3-9, the 'Constant Density' method did not have a positive effect on the structural components of the guess since this method produced the lowest R_g^2 and C_R values of all the methods. It also resulted in a relatively low CED and high final potential energy. However, it did provide the best symmetry as represented by the chain half R_g^2 values and p-values depicted in Figures 3-5, 3-6, 3-11, and 3-12. The fact that there was very good symmetry and a low collapse ratio indicates that the self-avoiding problem was exacerbated by applying this constant density method. The main problem with this method was that the initial monomers were concentrated within a smaller box volume at the beginning of the polymerization, resulting in a dense area and less homogeneity throughout the box during the polymerization.

3.2.5 Structure Filtration

Because of the challenges associated with improving the structural components of the initial guess, a filtration method was formulated as a supplemental method to filter out outliers and chains with R_g^2 values that were below a criterion value. Once a chain had grown completely, R_g^2 values were calculated, and chains were accepted if they were within the predetermined range. This range was decided based on the statistical parameters associated with the R_g^2 distribution of the initial guesses generated using a base method, such as skew, median, mean, and standard deviation. In other words, data acquired after the application of a base method underwent a statistical analysis to determine an R_g^2 range that would improve the statistical parameters of the distribution such as the total R_g^2 mean. Figure 3-16 depicts the histograms of the base method data and statistical parameters associated with a particular criterion range. This was used to preemptively decide the filtration range. Figure 3-16A) depicts the R_g^2 distribution of the base method (in this case, the Original Method), 3-16B has a filtration range of 250Å$<R_g^2$, and 3-16C has a range of 260Å$<R_g^2<$1100Å.

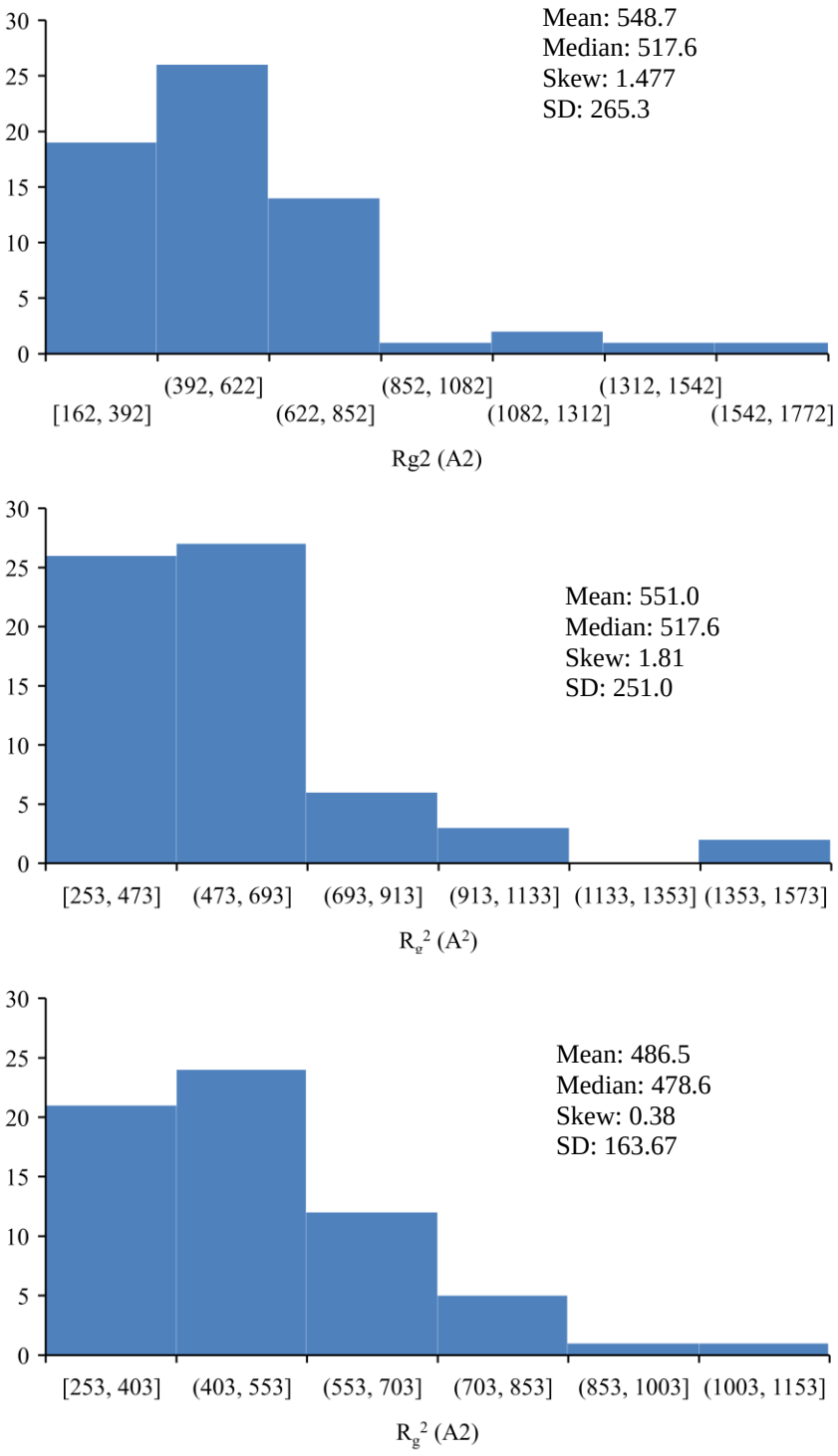

Figure 3-16 Original Method R_g^2 Distribution and Statistical Parameters. Sample size was 64 chains for A, B, and C.

A distribution with a skew below -1.0 or above 1.0 is considered severely skewed, and normative statistical analysis methods cannot be utilized.[48] The base method distribution (Figure 3-16A)) was severely skewed and the standard deviation was high (typically the standard deviation would be about one-third the mean). Discarding values below 250Å, a threshold that was one standard of deviation below the mean of the original data with a sample size of two hundred and thirty-two, resulted in an even worse skew though mean and SD were improved. Filtering out high R_g^2 outliers (Figure 3-16C)) resulted in a worse mean; however, the median is much closer to the mean and skew and SD are great. This is the most normal distribution.

Based on this preliminary analysis, the Filtration method did not look promising for improving the structural parameters. However, as shown in Figure 3-9 and Figure 3-10, application of a filter with a criterion range of 260Å< R_g^2<1100Å resulted in a noticeable improvement in R_g^2, C_∞, and C_R. There was no improvement in CED. For other promising methods, filtration was applied complimentarily. Filtration did not improve the R_g^2 values for initial guesses generated using the 'Mixed' method but did improve R_g^2 for initial guesses generated using the 'Forcefield Modification' method. Importantly, filtration had a much more significant effect on the polystyrene systems investigated by our collaborators. This may be due to the amount of skew already present in the base method data.

3.2.6 Forcefield Modification Methods

Forcefields, which are equations and constants utilized to set molecular geometries and forces between atoms (See Section 2.4), are normally applied to polymer systems

during minimization and dynamics. Furthermore, forcefield modification is typically intended to affect the minimization and dynamics processes. Little research has been conducted to optimize a forcefield for the initial guess generation. We developed an initial guess method that applied a modified forcefield to the initial guess generation process. Specifically, changes were made to the particle size, σ, the dispersion energy parameter, ε, and the periodic boundary cutoff range (a third order spline was applied). These modifications were applied to all carbon and hydrogen atoms and bromine pseudo-atoms. Modifications are tabulated in Table 3-7 below.

Table 3-7 Initial Guess Forcefield Modifications.

	Base and Supplementary Methods	σ^a	ε^a	cutoff	Name
1	Original	10%	-	1Å-2Å	Forcefield: 10%σ, cutoff:2Å
2	Original	1%	-	1Å-2Å	Forcefield: 1%σ, cutoff: 2Å
3	Original	1%	-	carbon $0.9*r^b$-r	Forcefield: 1%σ, cutoff: r
4	Mixed	1%	-	1Å-2Å	Mixed+Forcefield: 1%σ, cutoff:2Å
5	Original	0.1%	-	1Å-2Å	Forcefield: 0.1%σ, cutoff:2Å
6	R_g^2 Filtration	0.10%	-	1Å-2Å	Rg2 filter+Forcefield: 0.1%σ
7	Original	-	10%	1Å-2Å	Forcefield: 10% ε, cutoff:2Å
8	Original	1%	120%	1Å-2Å	Forcefield: 1%σ, 120%ε

[a] percentages were multiplied by the parameter to give the value used in the forcefield method. [b] r is equal to the product of σ and $2^{1/6}/2$. This value is much smaller than σ.

The reduction of the σ parameter had a surprising positive effect on the structural properties of the initial guess. As shown in Figure 3-9, R_g^2 increased as σ was reduced from 10% to 0.1%. A σ parameter of 0.1% with supplementary R_g^2 filtration resulted in the best R_g^2 values of all the candidate methods and the second best collapse ratio. The percent error associated with this method was only 10.1% which shows huge potential when compared with a 30% error of the 'Original' method. The 'Forcefield: 1%σ, cutoff: 2Å' method

provided the best CED of 179.6 J/cc, which is not a huge improvement from the 'Original' method value of 167.6 J/cc. Generally, none of the methods showed a drastic improvement in CED, though density is strongly related with CED. The cutoff range was altered for the 'Forcefield: 1%σ, cutoff: r' method where the cutoff ranged from the value of 90% of the alkyl carbon r parameter (equal to $\sigma*2^{1/6}/2$) to 100% of this parameter. This new cutoff is much smaller than 1Å to 2Å and was compared with the 'Forcefield: 1%σ, cutoff: 2Å' method to determine the effect of only changing the cutoff. There was a small improved in the C_∞ and C_R parameters but a decrease in the R_g^2 parameter. The 'Mixed+Forcefield: 1%σ, cutoff: 2Å' method included application of a reduced σ forcefield on a 'Mixed' base method. As discussed in Section 3.2.3, the 'Mixed' method involved globally after a monomer was polymerized and resulted in improved chain symmetry. Interestingly, the second halves of these chains were longer than the first halves, which is the opposite for chains generated with the 'Original' method. Generally, altering the ε parameter (either increasing to 120% ε or decreasing to 10% ε) resulted in worse structural and energetic resulting properties.

3.2.7 RIS Model Modification

Optimization of the RIS model was conducted to improve the mean square radius of gyration of the isolated chains as shown in Table 3-1. This optimized RIS model—which included increasing the trans proportion, η, from 0.9 to 1.4 and decreasing the gauche minus proportion, τ, from 0.5 to 0.4—was applied to systems of 2 chains of 76 monomers at various densities as shown in Table 3-6. Mean square radius of gyration values did improve significantly. However, CED values and p-values were generally worse. This is

because with a greater trans conformation proportion, chains are not wrapping around each other as much. Therefore, solubility and chain end collapse are worse.

CHAPTER 4. LONG-RANGE CORRECTIONS

4.1 Introduction

Though molecular simulations are a valuable tool for gaining insight into the physical and energetic properties of many materials and systems, they often neglect long-range contributions to energy and pressure calculations.[49] Bulk polymer or biological systems may have hundreds of thousands or even millions of atoms, thus simulators often avoid calculating all pair interactions because of the high computational cost. One way of reducing the number of pairs considered is by applying a cutoff truncation to the atomic energy interaction. This cutoff is required for all periodic systems and cutoff neglects the long-range interactions that contribute to the potential energy and pressure values in the simulation.[12, 43, 49] It "solves" the problem of an expensive calculation, but results in significant inaccuracies. Because electron dispersion (Lennard Jones) energetic interactions are relatively short-range, these can be calculated from an analytical integral over the product of the atomic energy or virial, and the atomic pair distribution function converges to its asymptotic value of one. Unfortunately, electrostatic long-range interactions do not converge and require conditionally convergent methods such as Ewald Sum or Multipole Expansion. Long-range interactions, however, may be significant, and therefore, an efficient and simple method for calculating these corrections would be very valuable.

4.2 Long-Range Energy Corrections

For most molecular simulations, molecules are built inside periodic cells that can be replicated to produce a lattice of mirror images.[11-12, 29, 43] Molecules that are located at or beyond the edges of the periodic box can interact with image molecules of the adjacent virtual box. If a molecule is located along or beyond the face of a lattice, then the molecule will be reflected into the exact opposite face and interact with those molecules to enable the continuity of interactions. A periodic boundary system is convenient for bulk modelling since a periodic cell can be replicated many times and polymerization times can be expedited. Figure 4-1 below illustrates a lattice of mirror image periodic cells that can be consolidated into a single periodic cell to produce a bulk system. Note that the brown molecules extending outside the top of the 3x3 lattice were reflected into the opposite face when the lattice was ‘built’.

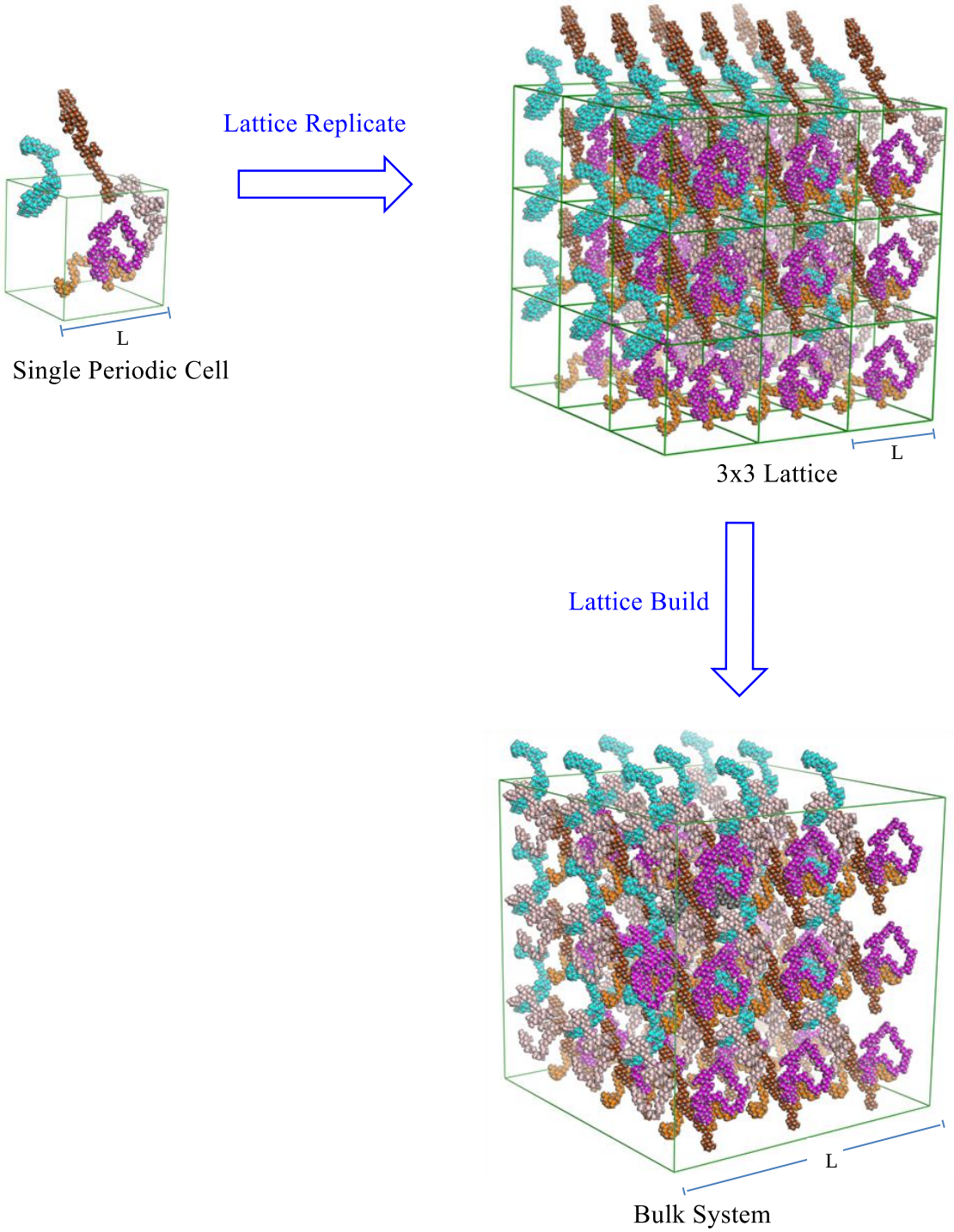

Figure 4-1 Construction of a Bulk Periodic System. The cutoff is equal to half the box side length.

4.2.1 Dispersion Energy and Pressure Corrections

The use of periodic boundary conditions, requires the institution of a cutoff distance equal to one-half the length of a box side.[43] This is to avoid the double-counting of atom pair interactions since, for example, the interaction of atom i with atom j would include all of the periodic images of atom j, and one-half the box side length would be the maximum distance to avoid this false periodicity.[11, 43] Consequently, calculation of the total interaction energy or pressure would require the partition of the short and long-range interactions and the cutoff distance would serve as a truncation point. Often a 'spline'—a third or fifth order polynomial function bounded by a given cutoff range—is utilized over a 'direct cutoff' to gradually taper the interactions examined.[14, 23] Figure 4-2 illustrates an interaction potential curve in blue, and the spline is represented by the red line. The area of the curve beyond the spline represents the long-range interactions that would need to be calculated later.

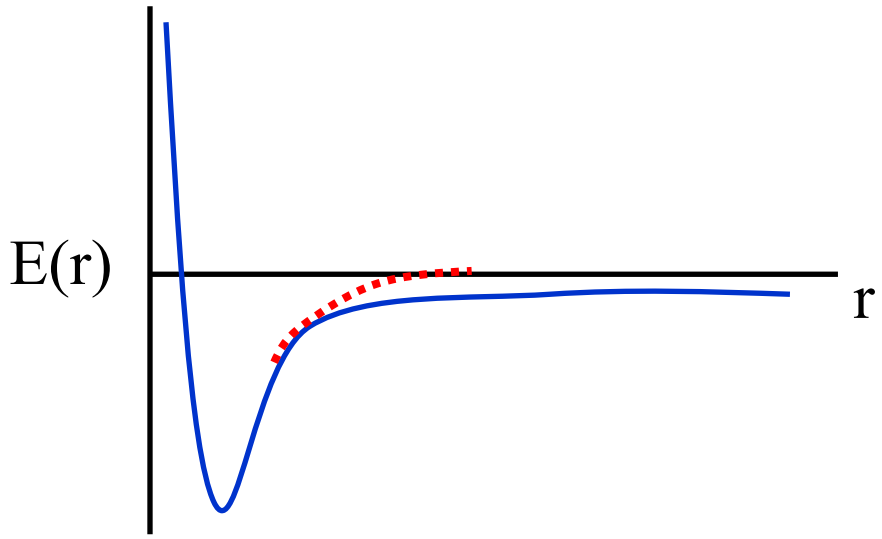

Figure 4-2 van der Waals Interaction Potential Curve with Spline.

Equation 31 is utilized to calculate the total interaction energy for a system with N molecules or atoms, a box side length, L, and a cutoff distance of L/2.[16, 43] u(r) is the pairwise additive intermolecular potential. The first portion of the equation represents the short-range interactions and the second portion of the equation represents the long-range correction.

$$E_{total} = \sum_{i} \sum_{\substack{i>j \\ r_{ij}<L/2}} u(r_{ij}) + 2\pi \frac{N^2}{L^3} \int_{L/2}^{\infty} g(r)u(r)r^2 dr \tag{31}$$

There are a few methods for calculating the long-range correction including: Integral Correction Method, Cell Multipole, and Ewald Summation.[11, 49] The Integral Correction Method involves the direct integration of the second portion of Equation 31, and it is what is typically used to calculate long-range dispersion energy. A pair distribution function, g(r), describing the distribution of radial distances between atom pairs for a given box volume is required to calculate the long-range correction. However, for long-range correction calculations, it is typically assumed that the atom spheres are far enough apart beyond the truncation limit to be mutually independent of each other such that the pair distribution function is approximately unity, simplifying the integral.

Similarly, the contributions pairwise interactions make to the pressure is given by

$$P = \frac{Nk_BT}{V} - 2\pi \left(\frac{N}{V}\right)^2 \int_0^{\infty} g(r) \left(r \frac{du(r)}{dr}\right) r^2 dr \tag{32}$$

The first term on the right of Equation 32 is the kinetic pressure due to the impact of molecules in free motion, the second term is the internal pressure, and the quantity $r\frac{du(r)}{dr}$ is the virial or the product of the force acting between atoms i and j and the distance between the atoms.[16] Even though the dispersive contribution to the long-range energy and pressure can be calculated analytically, this calculation requires detailed knowledge of the forcefield parameters, the use of the spline or cutoff, and the convergence of the pair distribution function. These details vary with different forcefields that may be used, so the proposed technique is much simpler to carry out in a generalized fashion even if no electrostatic charges are employed in the system.

4.2.2 Electrostatic Energy Corrections

For large and charged system the long-range correction calculation diverges unlike for dispersion energy calculations as shown in the following equation for the determination of Coulombic interactions:

$$E_{Coulombic} = 2\frac{N^2}{V}\int_{R_c}^{\infty}\frac{q_i q_j}{4\pi\epsilon_0\epsilon(r)r}r^2 dr \tag{33}$$

To solve this problem, the Ewald Sum Method was established which utilizes Fourier Transforms to calculate the long-range interactions in k-space and the short-range interaction in real-space.[11-12, 49-51] The Ewald Sum method is conditionally convergent because it depends on the order of summation. The Fourier part of the Ewald Sum and the correction due to self-interaction can be calculated using Equations 34 and 35 below

$$U_F = \frac{1}{2V}\sum_{k\neq 0}\frac{4\pi}{k^2}|\rho(\boldsymbol{k})|^2 \exp\left(-\frac{k^2}{4\alpha}\right), \qquad \rho(\boldsymbol{k}) \equiv \sum_{i=1}^{N} q_i \exp\left(i\boldsymbol{k}\cdot\boldsymbol{r_i}\right) \tag{34}$$

$$U_{self} = \left(\frac{\alpha}{\pi}\right)^{0.5}\sum_{i=1}^{N} q_i^2 \tag{35}$$

where **k** = (2π/L)**l, l** = (l_x, l_y, l_z), k is the number of charges, and **r** is the point charge position. The gaussian has a width of $\sqrt{2/\alpha}$. The real-space sum is the short-range contribution and can be calculated with

$$U_{short-range} = \frac{1}{2}\sum_{i\neq j}^{N} q_i q_j erfc(\sqrt{\alpha} r_{ij})/r_{ij} \tag{36}$$

These contributions can be summed to obtain the total electrostatic energy using

$$U_{Coul} = \frac{1}{2V}\sum \frac{4\pi}{k^2}|\rho(\boldsymbol{k})|^2 \exp\left(-\frac{k^2}{4\alpha}\right) - \left(\frac{\alpha}{\pi}\right)^{0.5}\sum_{i=1}^{N} q_i^2 + \frac{1}{2}\sum_{i\neq j}^{N}\frac{q_i q_j erfc(\sqrt{\alpha} r_{ij})}{r_{ij}} \tag{37}$$

The Cell Multipole Expansion Method was subsequently developed which involves a Taylor series (or expansion) that sums the interaction energies a set of charges have with a particular point in space such that the electrostatic potential at this point can be calculated.[11-12, 49, 51] A single parent cell is subdivided into children cells multiple times until a maximum level, R is reached (this subdivision is called the octal tree with 8^R total cells). A multipole expansion about the center of the box is used to describe the charge

distribution in each cell. Then, multipole expansion of the children cells is translated to the center of the parent cell iteratively until level zero is reached. A set of conditions are used to calculate the potential of the cells. This expansion can be truncated depending on the number of harmonic spheres (atoms) in the system. The number of interactions is on the order of N^2 and the time required to perform this method scales as O(N). The time required to conduct multipole expansion is twice that of the Ewald Sum Method. Equation 38 below describes the potential at a point **r'** (θ', φ') sufficiently far away using multipole expansion

$$\varphi(\boldsymbol{r}') = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \frac{\sum_{i=1}^{k} q_i r_i^n Y_n^{-m}(\theta_i, \varphi_i)}{r^{n+1}} Y_n^m(\theta', \varphi'), \tag{38}$$

where k is the number of charges with charge q at a location $\mathbf{r_i}$ (r_i, θ_i, φ_i) and Y_n^{-m} are the Legendre functions. The close and far potential energy is calculated with the following equations

$$U_{close} = \sum_{close} (q_i q_j)/r_{ij} \tag{39}$$

$$U_{far} = \sum_{i} U_{local(R)}(\boldsymbol{r_i}) \tag{40}$$

where U_{local} is calculated from the local expansion.

4.2.3 Proposed Extrapolation Method

Though convergent, the direct analytical integration of dispersion interactions can be complex and, for this reason, is often neglected from potential energy calculations.

Moreover, as previously described, brute-force integration of the Coulombic interaction is not viable because of its divergent nature and requires reformulation of the charge distribution with a multipole expansion or the use of the conditionally convergent Ewald Sum Method.

In this dissertation, a simple and empirical method has been developed to rapidly obtain both the energy and pressure long-range corrections. The idea is that within a periodic system, molecular interactions are a function of the size of the box. As the lattice or total size of the periodic system—and consequently the cutoff—increase, the potential also gradually increases to account for long-range interactions between atoms that are further and further away. The following steps were taken to construct a converging potential curve for the extrapolation of the total interaction energy (refer to Figure 4-1).

1. Construct the molecular system of interest within a 1x1x1 lattice. The cutoff should be one-half the length of the box side.
2. Increase the dimensions of the lattice by an iterative index number.
3. Build or consolidate the lattice to a single periodic cell and modify the cutoff to one-half the side length of the new periodic cell.
4. Obtain the total potential and repeat steps 2 through 4 until the potential has converged.

Figure 4-3 illustrates a converging potential energy curve obtained for an aPP system of A) 2 chains with 76 monomers and a system of B) 8 chains with 170 monomers. Note that the convergence of the smaller system requires a greater number of periodic cubes than for the larger periodic system.

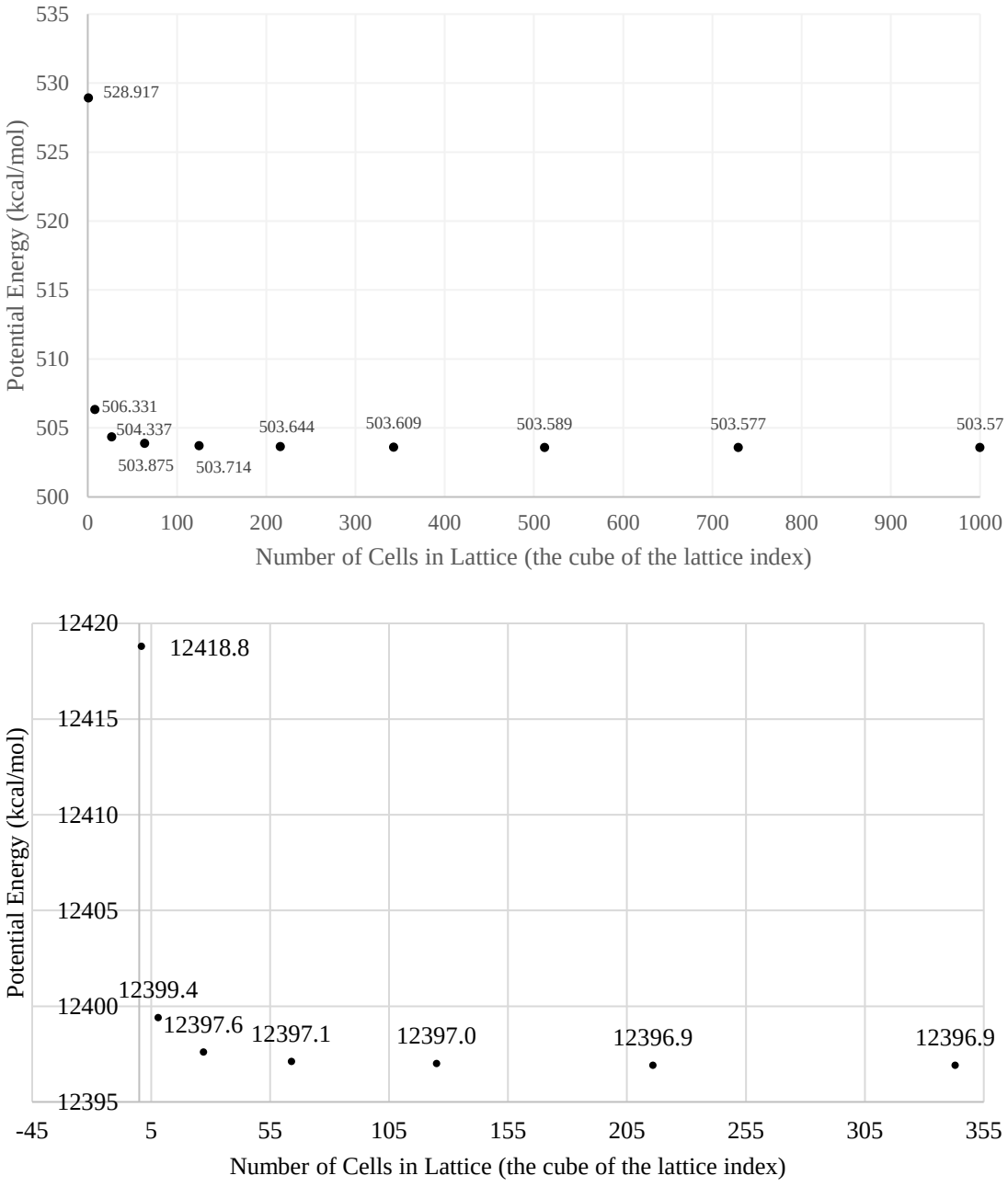

Figure 4-3 Potential Energy Curve for aPP Long-Range Energy Correction Calculation. A) 2 Chains of 76 Monomers, B) 8 Chains of 170 Monomers.

Our extrapolation method may also be utilized to calculate the Hildebrand solubility, δ, parameter described in Chapter 3. In light of the following equations,

$$E_{coh} = U_{tot,isolated} - U_{tot} \tag{41}$$

$$\delta = \left(\frac{E_{coh}}{V}\right)^2 \tag{42}$$

the cohesive energy is dependent only on the total potential energy since the isolated chain potential is a constant value. Therefore, the cohesive energy and the solubility are a function of the number of cells in a lattice and converge as well (Figure 4-4). All cohesive energy density calculations were conducted using our extrapolation method and, thus, include long-range corrections.

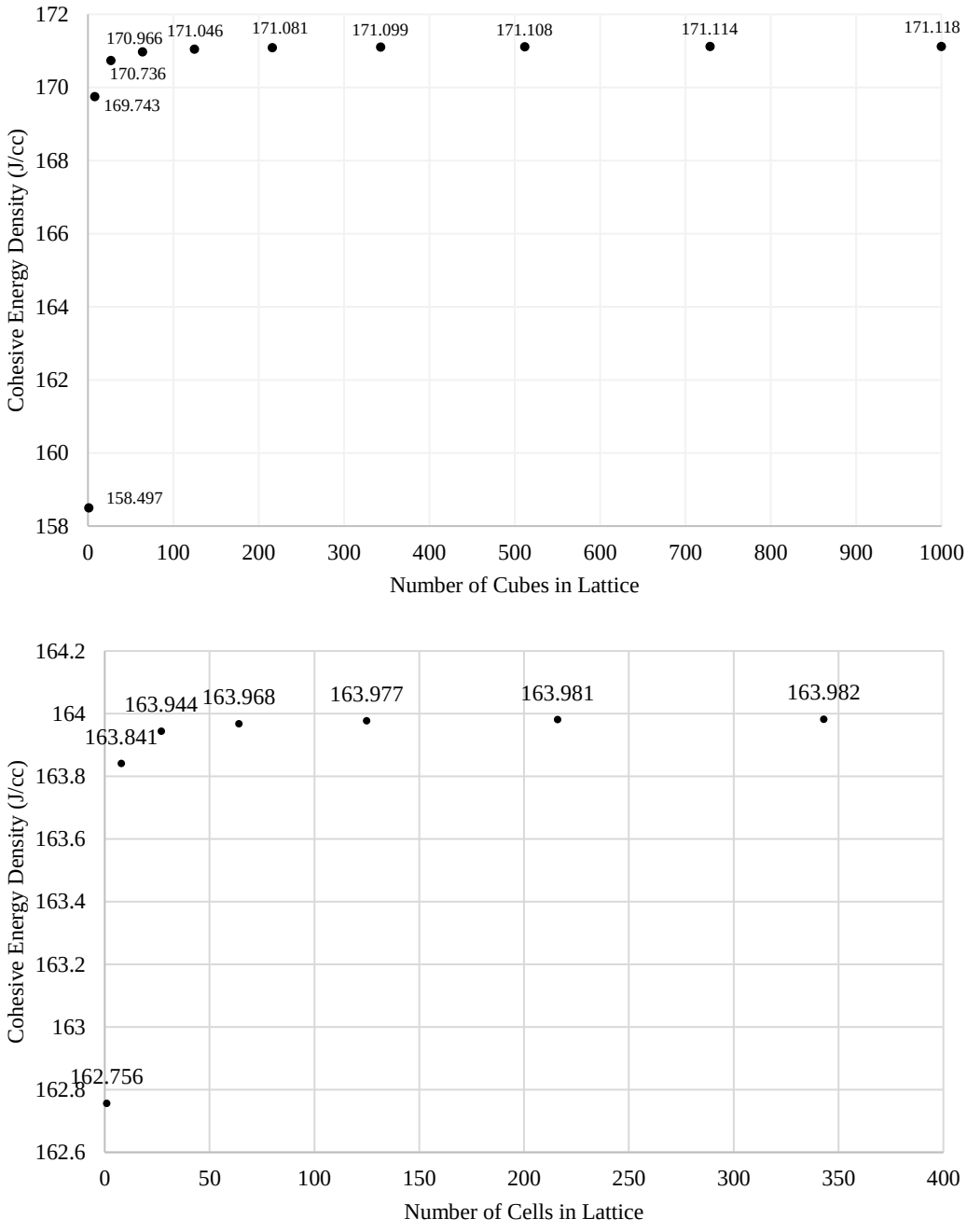

Figure 4-4 aPP Cohesive Energy Density Convergence of A) 2 Chain, 76 Monomer System, B) 8 Chain, 170 Monomer System.

An atactic polystyrene system was generated with 4 chains of 181 monomers at a density of 0.9948g/cc (experimental density is about 1.07g/cc) to evaluate the efficacy of this proposed method for calculating not only the dispersion energy long-range interactions but also the electrostatic long-range interactions.[30] Figure 4-4A) depicts the van der Waals potential energy convergence and Figure 4-4B) depicts the electrostatic potential energy convergence with an increasing number of cells in the lattice. Naturally, the van der Waals potential converged faster than the electrostatic potential. To reach a percent change of less than 1%, a 2x2x2 lattice was required for van der Waals potential convergence, and a 6x6x6 lattice was required for an electrostatic potential convergence. However, building a 6x6x6 lattice was still relatively quick and reasonable and results in accurate acquisition of the long-range electrostatic correction without use of conditionally convergent or more complex methods.

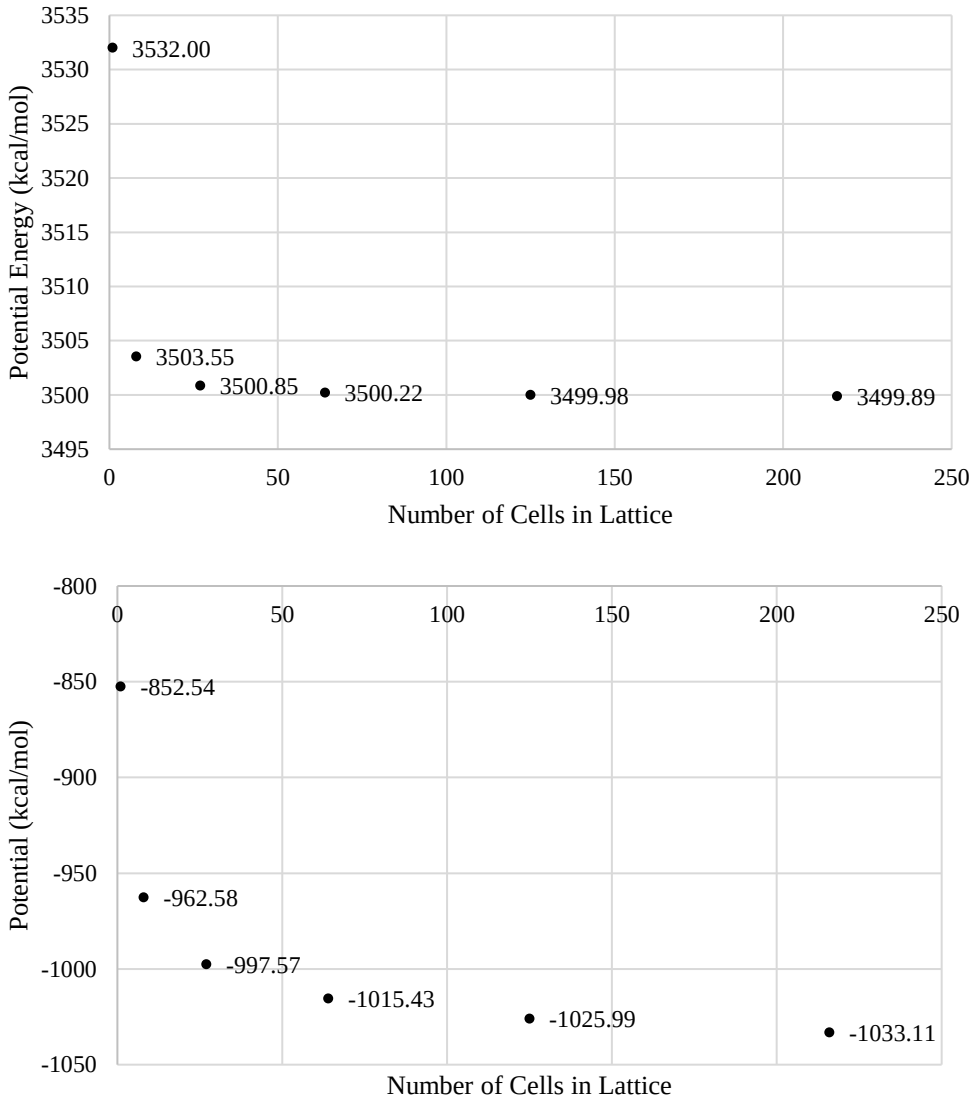

Figure 4-5 Atactic Polystyrene A) Van der Waals and B) Electrostatic Convergence Curves for a System of 4 Chains of 181 Monomers.

Long-range pressure corrections were also calculated for an aPP system of 8 chains of 170 monomers as shown in Figure 4-6 (pressure units are in kPa). The pressure decreases with increasing number of lattice cells since as the number of molecules in the system increases, the cohesion of molecules or the attraction of the molecules for each other increases. The instantaneous pressure calculation included the sum of the virial and the

virial due to bond constraints. The instantaneous pressure is below 100kPa probably because of the randomness of chain generation.

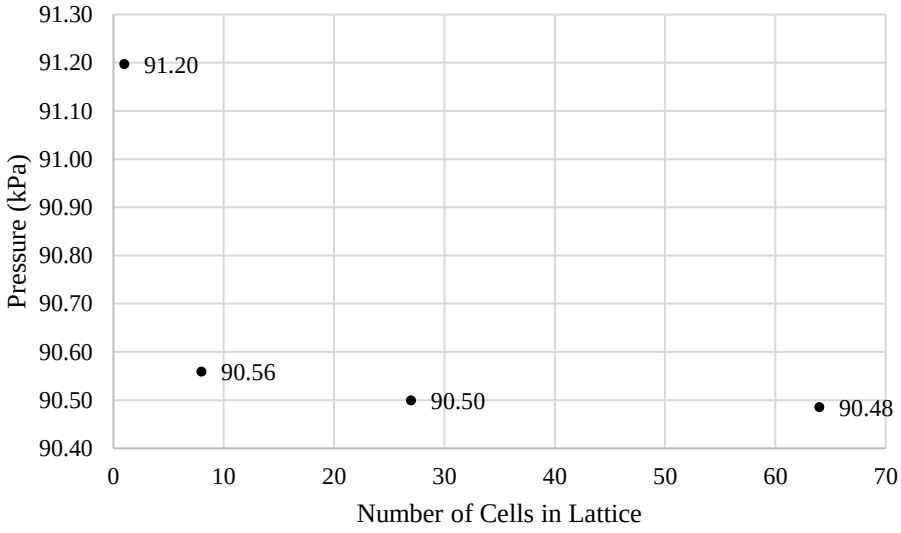

Figure 4-6 aPP Pressure Convergence Curve for a System of 8 Chains of 170 Monomers.

CHAPTER 5. FINAL REMARKS

5.1 Conclusions

- **Density changes had only marginal effects on shape and size parameters such as R_g^2, C_∞, and C_R, and any small improvements to these parameters were lost during minimization and dynamics.** It was initially hypothesized that the density of the system would be the main factor that controlled the shape and size of the polymer chains since at very low density, the whole chains would collapse and not entangle with each other and at high density, chain asymmetry would prevail because of the lack of free volume and the self-avoiding nature of the chains. However, the results described in Chapter 3 show that the shape and size parameters are mainly affected by the RIS model and/or the forcefield.
- **CED and MSD were highly sensitive to density changes**. It was initially hypothesized that moderately lower densities would improve not only the shape and size parameters, but also the cohesive energy density. This was based on the observation that lower density greatly improved atactic polystyrene models developed by collaborators at USF. Interestingly, densities that were *above* the experimental density greatly improved the cohesive energy density of atactic polypropylene such that it matched the cohesive energy density reported by experimental studies.[44] This may be due to less bulky nature of polypropylene compared with polystyrene. For future work with polymers of varying chemical structures, density optimization would be advised.
- **Modifying the RIS model resulted in improved R_g^2 values, but worse chain symmetry and CED values.** An effort was made to extend the chains and improve shape and size parameters by increasing the trans and lowering the gauche minus proportions of the RIS model. This resulted in improved R_g^2 values, however CED decreased and chain symmetry worsened. Increasing the trans proportion and lowering the gauche minus proportion did not only affect the ultimate parts of the chains but also the beginning parts. Thus, as polymerization progressed, the extended chains ran into each other and therefore chain ends collapsed more. Additionally, chains would not entangle with each other as much with these changes to the RIS model.
- **Integrating minimization into the polymerization process reduced the asymmetry of chains.** Global minimization was conducted after the addition of every monomer during initial guess generation. This did not significantly improve energetic parameters of aPP systems, however there were significant improvements to the chain symmetry; with global minimization, all chains were able to shift providing some free volume towards the end of the polymerization process. An advantage of this method is that it can be applied to different segments of the chains, e.g. the ends of the chains, and it can be controlled by altering the Euclidean norm gradient. A lower gradient will result in finer energy minimization.

- **Multi-step minimization after the initial guess was generated using various forcefields for each step resulted in slightly improved structural and energetic parameters.** The initial steps of the multi-step minimization included lowering the σ parameter of the forcefield for the carbon and hydrogen atoms. This had the effect of bringing the chains closer to each other such that chains could slip past each other into more stable conformations. Lowering this parameter may not, however, work for all polymers. Therefore, the number of minimization steps and the forcefields used for each step must be customized and optimized for a particular polymer. A greater number of steps may take longer but result in more accurate structures.
- **R_g^2 filtration improved structural parameters significantly and should be used in future models as a supplementary method.** Filtration of structures with low R_g^2 values is similar to a Markov Chain Monte Carlo sampling method and has proven to be useful in improving the structural components of the chains especially for difficult and bulky polymers such as polystyrene.
- **Modification of the σ parameter in the forcefield during initial guess generation resulted in the best structural and energetic results.** Due to its unbulky nature, aPP benefited significantly from a lower particle size, or σ, parameter in the forcefield since the chains would not be too sterically hindered by passing near each other. Additionally, aPP chain extension is enabled when chains can pass by each other.
- **Brute-force calculation of the long-range energy and pressure of polymer systems can be achieved by plotting the energy or pressure against an increasing system periodicity.** Long-range energy and pressure calculations involve difficult integrals that can be computationally expensive and are often neglected. The method described in Chapter 4 resulted in relatively quick and simple corrected energy and pressure calculation. Moreover, this method can be applied to calculate electrostatic energy contributions which typically require more complex, conditionally convergent methods to calculate. Electrostatic energy converged slower than van der Waals energy, however it still converged relatively quickly.

5.2 Recommendations

- **Increase the number of steps of energy minimization and explore various σ forcefield parameter modifications for each minimization step.** Decreasing the σ parameter was beneficial for a system such as polypropylene, but the opposite may be true for more bulky polymers such as polystyrene. One investigation conducted by Theodorou et al. on polystyrene involved a 41-step minimization method.[37] Further research into minimization steps should be conducted.
- **Polymerize chains bidirectionally.** Significant effort was exerted to ameliorate the asymmetry of chains. The chains generated in this research were built unidirectionally, i.e. chains grew from one end to the other. Another approach could be to begin polymerization from the middle of the chain and bond monomers on both ends such that chains grow bidirectionally. The effects of such a method could

improve structural properties of the chains and may even be useful in generating block copolymers

- **Apply Protracted Colored Noise Dynamics (PCND) to atomistic simulations.** PCND is a technique which has been shown to reduce the relaxation time of coarse-grained copolymer systems by a magnitude of 10^5-10^7 to cope with the long timescales associated with large glassy polymer systems. One of the greatest challenges for this research was the long computational times associated with dynamics. This amount of time was expounded greatly for bulky, charged polymers such as polystyrene.
- **Generate initial guess structures that have a distribution of molecular weights and tacticities.** Real polymer systems are not composed of 'parent' chains, or chains that all have the same molecular weight and tacticity distributions. Generating structures with a distribution of molecular weights and tacticities may significantly affect the entanglement, shape, and size of chains.
- **Obtain the actual conformation distribution of the chains.** This distribution will illuminate whether the trans proportion at the ends of the chains is much lower due to repulsive intermolecular interactions.
- I**nvestigate LowModeMD and determine whether it can be applied locally.** LowModeMD was applied to our polymer systems to search for the lowest energy conformations and configurations; however, this method was only really effective for small molecules. Modification of this method to accommodate large systems would be very valuable research. Additionally, application of LowModeMD to specific areas of the chains would be useful.
- **Apply the techniques and protocols to generate more polymer systems with various chemical structures and compare results with those of aPP.** The research conducted here showed that there some techniques worked very well for aPP and others did not. However, work done on aPS by USF collaborators showed that some techniques that worked for aPP did not work for aPS and vice versa. Understanding the structural and energetic differences and property relationships for a variety of polymers may provide valuable insight about the initial conditions and techniques that should be used. Additionally, this information may be useful for modelling block copolymers.

APPENDIX A. POLYBUILD SVL CODE

```
#svl

#set main 'polybuildpp300'

function polybuildpp300 [dbname,degpoly,nchains,tacticity,boxedge,forcefield]
// polybuildpp300 builds an amorphous polymer chain. This initial guess generator can be easily
//adapted for polymers with various side groups by simply modifying the RIS model section.

//dbname: name of the monomer database; it should have a single monomer entry
//degpoly: degree of polymerization
//nchains: number of chains
//tacticity: tacticity vector generated by the Matlab code in Appendix C
//boxedge: length of periodic box side in Angstroms
//forcefield: forcefield file (.ff extension)

//sample inputs:
//tacticity=[0,1,0,0,0,1,1,0,0,0,1,1,0,0,0,1,1,0,0,1,1,0,0,0,1,0,0,1,1,1,1,0,0,0,1,0,0,0,0,0,0,1,1,0,0,1,
1,1,1,1,0,1,1,1,0,1,0,0,0,0,1,0,0,0,1,1,0,0,0,0,1,0,1,1,0,0,1,1,0,0,0,0,0,0,0,1,1,0,1,1,1,1,1,1,0,0,0,0,
0,0,0,1,1,1,1,1,1,1,0,0,0,0,1,0,0,1,1,0,1,1,1,1,0,1,1,0,0,1,1,1,1,1,0,1,0,1,1,0,0,0,0,1,1,1,0,1,1,0,1,1,0,
0,1,1,1,1,1,1,0,0,0,1,1,1,0,1,1,0,0,1,1,0]
//polybuildpp300['ppmonomer3.mdb',170,8,tacticity,49.9342536857098]

local ckey1,rkey1,ckey2,rkey2,torsionatoms,eval;
local energy,gradient,virial,W,torsionprob,randnum,torstate,torsionang;
local cumprob,i,j,k,l,m,interstate,intrastate;
local nstates=3,angle1,angle2;
local angincrement=360/nstates;
local torsioninc=360/nstates,chosenstate;
local temp=300,test,test2,test3,test4;
local U,Um,Ur,ang0,ang1,angm0,angm1,angr0,angr1,tval;
local risprob,lrprob,angvalues,EnergyLR,interdiadangle,intradiadangle;
local risstate,totprob,testatoms,rkeys,ckeys,chainkeys,reskeys,imon,ichain;
local anglechoices,states,torsionangle,sysenergy;

//Relevant functions being called:
//-----------------------------------------
//monomeradd1 bonds the new monomeric unit in the correct
//orientation to the growing chain and deletes the and atoms
//which are Br atoms assigned the parameters of a unified methyl
//-----------------------------------------
//risparams enters the relevant parameters for the RIS model
//that describes the torsional states
//-----------------------------------------
//choosestate radomly chooses a state based on the
//normalized probabilities that are input
//-----------------------------------------
```

```
//residuebackbone extracts the backbone from a monomer to
//determine the backbone atoms in the torsion angles
//----------------------------------------
//box start randomly places and orients the first monomers in
//the periodic cell
//----------------------------------------
//torsionanglechoice uses both the RIS model and the long-
//range energy to determine the backbone rotations angle
//----------------------------------------
//LREnergyCalc calculates the LR energy for a particular
//monomer at the trans, gauche, and gauche- states

//declare functions
function monomeradd1,risparams,choosestate,residuebackbone,boxstart;
function torsionanglechoice,LREnergyCalc;

//Set the cutoff range, disable electrostatic energy calculation, load the forcefield file
pot_Setup[cutoffEnable:1,cutoffOn:(0.5*boxedge)-
2,cutoffOff:(0.5*boxedge),eleEnable:0,resEnable:0,oopEnable:0,stbEnable:0,solEnable:0];
pot_Load forcefield;

//Open the database of interest
local dbkey=db_Open[dbname];

//Determine that there is one monomer
local nmonomers=db_nEntries dbkey;

//Read the monomer structures from the monomer database
local monomer=cat db_ReadColumn [dbkey,'mol'];

//Randomly place the beginning of the growing chains in the periodic cube
[chainkeys,reskeys]=boxstart [dbname,nchains,tacticity,boxedge];
reskeys=[reskeys];

//Calculate the system energy at the addition of the first monomer. This will be placed in sysenergy
//vector in the output
  test=Potential[W:0,dX:0];
  sysenergy(1)=test(1);

//Assemble Residue Keys (reskeys) in the format reskeys(chain)(residue in chain)
//First initialize all residue keys to zero so we can tell if a residue key is not assigned
reskeys=rep[0,degpoly,nchains];

//Now assign residue keys for the first residue and for all chains
test=Residues[];
for i=1,nchains loop
   reskeys(i)(1)=test(i);
endloop

//input the RIS model parameters
local risparamvalues=risparams[];
```

```
U=risparamvalues(1);
ang0=risparamvalues(4);
ang1=risparamvalues(5);

//Initialize states tensor to keep track of torsional states
//Format of states is that the highest level is for chains, the next level is for residues (or monomers)
//For example, three chains of degree of polymerization four should look like:
//[[[1,2],[2,3],[2,1],[1,3]],[[1,2],[2,2],[2,3],[1,2]],[[1,1],[3,1],[2,3],[1,2]]] such that states(2)(3)(2)
//is a vector of vectors describing the inter and intradiad states(2)(3)(2) for chain 2, monomer 3
//which in this case will be [[1,2],[2,2],[2,3],[1,2]]. states(2,3) describes the inter and intradiad
//states for chain 2 and monomer 3 which is [2,3]. Finally, states(2,3,2) describes the intradiad states
//(the first one is the interdiad state) for chain 2, monomer 3, which is 3 or the gauche minus highest
//energy state.
states=rep[0,2,degpoly,nchains];

//Assign the torsional states of the first monomers:

//Automatically assign first interdiad state to trans(1)
//Set the intradiad angle for the first monomer, note that the interdiad angle in the first monomer is
//not defined, but is set to trans for administrative purposes
for i=1,nchains loop
  states(i)(1)=[1,choosestate[U(1)]];
  angle1=residuebackbone[reskeys(i)(1)];
  aSetDihedralCWDeg[angle1,ang0(1)(states(i)(1)(2))];
endloop

//Deactivate periodicity if boxedge is set equal to zero
if boxedge == 0 then
  CellEnable(0);
endif

//Main loop over all the monomers in the chains
for imon=2,degpoly loop
  pr 'Currently processing monomer';
  pr imon;
  for ichain=1,nchains loop
    ckey2=mol_Create(monomer);
    rkey2=cResidues ckey2;

//Add the next monomer to the growing chain.
//Use the monomeradd1 function to bond the new monomer to the chain
    torsionatoms=monomeradd1[chainkeys(ichain),reskeys(ichain)(imon-
    1),ckey2,rkey2,tacticity(imon)];

//update the residue keys
test=reskeys;
for i=1,nchains loop
  test=poke[test,i,cat cResidues chainkeys(i)];
endloop
reskeys=test;
```

```
//rotate both the interdiad and intradiad angles for the newly added monomer based on both the RIS
//probabilities and the LR probabilities
states=cat torsionanglechoice[ichain,imon,torsionatoms,risparamvalues,tacticity,states,temp];
  endloop

//Calculate the system energy at the addition of each monomer
  test=Potential[W:0,dX:0];
  sysenergy(imon)=test(1);

endloop

//Change the first unified atom to a methyl, and the last one to a hydrogen
for i=1,nchains loop
  test=indexof[aElement cat rAtoms reskeys(i)(1),'Br'];
  test=mget[cat rAtoms reskeys(i)(1),test];
  aSetElement[test,'C'];
  Add_H test;
  test=indexof[aElement cat rAtoms reskeys(i)(degpoly),'Br'];
  test=mget[cat rAtoms reskeys(i)(degpoly),test];
  aSetElement[test,'H'];

endloop

//Calculate and print out the system density
local MW=add aMass Atoms[];
pr 'MW in g/mole';
pr MW;
test= MW/6.02E23;
pr 'MW in actual grams';
pr test;
test= cube(boxedge);
pr 'boxedge cubed';
pr test;
test= test*1E-30;
pr 'box volume';
pr test;
local density=(MW/6.02E23)/(cube(boxedge)*1.0E-24);
pr 'Cell Density in grams/cc';
pr density;

//If desired, set the atomic charges to zero; for charged systems such as aPS, omit this command
aSetCharge[Atoms[],0];  //set charges to zero

return[sysenergy,states];

//================================================
local function monomeradd1 [ckey1,rkey1,ckey2,rkey2,tval]

//Find end atom B on residue 1 and adjacent backbone atoms
//Find end atom A on residue 2 and adjacent backbone atoms
local akey1=cat rAtoms[rkey1];
```

```
local akey2=cat rAtoms[rkey2];
local test;

//find B and A atoms on residues 1 and 2 respectively
local Bkey1=akey1(indexof['B',cat aName[akey1]]);
local Akey2=akey2(indexof['A',cat aName[akey2]]);
local C2key=aBonds[Bkey1];

// Find backbone atoms adjacent to B and A
//Atoms go in order from residue 1 to residue 2: C1,C2,C3,C4
//Atoms C1-C4 define the torsion angle between residues 1 and 2
//C2 is the backbone atom on residue 1 bonded to reside 2
//C1 is the backbone atoms on res 1 bonded to C2
//C3 is the backbone atom on res 2 bonded to C1
//C4 is the backbone atoms on res 2 bonded to C3

        local C2atoms=cat aBonds C2key;
        local C1key=C2atoms(indexof[1,aBackbone C2atoms]);
        local C3key=aBonds[Akey2];
        local C3atoms=cat aBonds C3key;
        local C4key=C3atoms(indexof[1,aBackbone C3atoms]);
        local C4atoms=cat aBonds C4key;
        local C5key=C4atoms(indexof['B',cat aName C4atoms]);
        local backbone=[C1key,C2key,C3key,C4key];

//Modify the orientation of the side group based on tacticity input
    if tval == 1 then
        aInvertChirality C3key;
    endif

//Move new monomer to end of chain
        local transvect=aPos(C2key)-aPos(Akey2);
        local newcoords=aPos(akey2)+transvect;
        aSetPos[akey2,newcoords];

//Rotate Monomer into bonding alignment

//Find rotation angle (rotateangle) and axis of rotation (rotatevect)
        local nattachvect1 = aPos(Bkey1)-aPos(C2key);
        nattachvect1=nattachvect1/(norm nattachvect1);
        local nattachvect2 = aPos(C3key)-aPos(Akey2);
        nattachvect2=nattachvect2/(norm nattachvect2);
        local rotateangle=rot3d_vAngle[nattachvect1,nattachvect2];
        local rotatevect=rot3d_vCross[nattachvect1,nattachvect2];

//Translate monomer 2 to the origin, rotate then translate back
        transvect=aPos(Akey2);
        newcoords=aPos(akey2)-transvect;
        aSetPos[akey2,newcoords];
        local rotatematrix=rot3d_Rotation[-rotateangle,rotatevect];
        local rotatedpos=rot3d_mvMul[rotatematrix,newcoords];
```

```
        aSetPos[akey2,rotatedpos+transvect];


//Adjust local residues and chains by deleting marker atoms and rebonding new residue to the chain
oDestroy Bkey1;
oDestroy Akey2;

Bond[C2key,C3key];
aSetDist [[C2key,C3key],1.54];
oReparent[rkey2,ckey1];
oDestroy[ckey2];

//Return the atom keys that allow for the rotation of the backbone torsion angles
//C1-C2-C3-C4 is the tacticity dependent intermonomer bond
//C2-C3-C4-C5 is the tacticity independent intramonomer bond

//Return the atom keys for the monomer backbone
//C1key-C4key is the interdiad torsion angle
//C2key-C5key is the intradiad torsion angle
return [C1key,C2key,C3key,C4key,C5key];
endfunction


//=========================================================
local function choosestate [normprobs];
//this uses uniform random numbers to choose the RIS state
local i,state,n,randnum,test;

state=0;
n=length(normprobs);

randnum=randU[1];
for i=1,n-1 loop
test=add normprobs[igen i];
  if randnum < test then
    state=i;
    break;
  endif
endloop

if state == 0 then
  state=n;
endif;

return state;

endfunction

//=========================================
local function risparams [];
//this function assigns all the RIS model parameters
```

```
//Assign state probability matrices
//U, Um, Ur are the 2x3 matrices containing the probabilities for the t, g, and g- states
//They are entered initially in a non-normalized state and then normalized below

//Assign RIS state probability matrices in order of t,g,g-
U(1)=[1,1,0.5];
U(2)=[1,0.05,0.5];
U(3)=[1,1,0.025];
Um(1)=[0.05,1,0.025];
Um(2)=[1,0.05,0.025];
Um(3)=[0.05,0.05,0.00125];
Ur(1)=[1, 0.05,0.025];
Ur(2)=[0.05,1,0.025];
Ur(3)=[0.05,0.05,0.00125];

//Assign angle values for states
//ang(i)(j)=three angles (t,g,g-) for diad with monomer j preceded by monomer i

//assign angles for t, g, and g- states for intradiad when that diad has the side chain in the 0 position
ang0(1)=[180,-60,60];
ang0(2)=[180,-60,60];
ang0(3)=[180,-60,60];
//assign angles for t, g, and g- states for intradiad when that diad has the side chain in the 1 position
ang1(1)=[180,60,-60];
ang1(2)=[180,60,-60];
ang1(3)=[180,60,-60];
//assign angles for t, g, and g- states for the meso interdiad when the side chain is in the 0 position
angm0(1)=[180,60,-60];
angm0(2)=[180,60,-60];
angm0(3)=[180,60,-60];
//assign angles for t, g, and g- states for the meso interdiad when the side chain is in the 1 position
angm1(1)=[180,-60,60];
angm1(2)=[180,-60,60];
angm1(3)=[180,-60,60];
//assign angles for t, g, and g- states for the racemic interdiad when the side chain is in the 0 position
angr0(1)=[180,60,-60];
angr0(2)=[180,60,-60];
angr0(3)=[180,60,-60];
//assign angles for t, g, and g- states for the racemic interdiad when the side chain is in the 1 position
angr1(1)=[180,-60,60];
angr1(2)=[180,-60,60];
angr1(3)=[180,-60,60];

//normalize the probability matrices
U(1)=U(1)/(add U(1));
U(2)=U(2)/(add U(2));
U(3)=U(3)/(add U(3));
Um(1)=Um(1)/(add Um(1));
Um(2)=Um(2)/(add Um(2));
Um(3)=Um(3)/(add Um(3));
```

```
Ur(1)=Ur(1)/(add Ur(1));
Ur(2)=Ur(2)/(add Ur(2));
Ur(3)=Ur(3)/(add Ur(3));

return [U,Um,Ur,ang0,ang1,angm0,angm1,angr0,angr1];

endfunction

//=======================================
local function residuebackbone rkey;
//this extracts the backbone carbon atom keys from a currently processing monomer

local residueatomkeys = cat rAtoms rkey;

local C1key=residueatomkeys(cat indicesof['A', cat aName residueatomkeys]);
local C4key=residueatomkeys(cat indicesof['B', cat aName residueatomkeys]);
local C2key=aBonds C1key;
local C3key=aBonds C4key;

return [C1key,C2key,C3key,C4key];

endfunction

//************************************************************

function boxstart [dbname,nchains,tacticity,boxedge]
//boxstart randomly places and rotates the initial monomers of the chains in the periodic box


local i,reskeys,atomkeys,chainkeys,C1key,C2key,C2Pos,axisRot,angleRot,rotationMatrix,rotPos;
local randRotAngles,MX,rotPosX,MY,rotPosY,MZ,rotPosZ,newcoords,transvect,molCenter;

//Create the Periodic Cube unless boxedge is set equal to zero
if boxedge == 0 then
  CellEnable(0);
else
  CellParameters['P1',[boxedge,boxedge,boxedge],[90,90,90]];
  CellEnable(1);
endif

//Open the database of interest
local dbkey=db_Open[dbname];

//Read the monomer structures from the monomer database
local monomer=db_ReadColumn [dbkey,'mol'];

//loop over the number of chains, and create that many initial monomers
for i=1,nchains loop

//Place the new monomer in the periodic box
        chainkeys(i)=mol_Create(monomer(1));
```

```
        reskeys(i)(1)=cResidues chainkeys(i);
        atomkeys(i)=cat cAtoms chainkeys(i);

//Move monomer to origin such that the bond b/w C1 and C2 is aligned with x axis.
        C1key=mget[atomkeys(i),(aName atomkeys(i))=='C1'];
        C2key=mget[atomkeys(i),(aName atomkeys(i))=='C2'];
        aSetPos[atomkeys(i),(aPos atomkeys(i))-(aPos C1key)];
        C2Pos=aPos C2key; //C2 is the one you plan on rotating so that it aligns with x-axis
        axisRot=rot3d_vCross[C2Pos,[aDist[C1key,C2key],0,0]];
        angleRot=rot3d_vAngle[C2Pos,[aDist[C2key,C1key],0,0]];
        rotationMatrix=rot3d_Rotation[angleRot,axisRot];
        rotPos=rot3d_mvMul [rotationMatrix, aPos atomkeys(i)];
        aSetPos [atomkeys(i), rotPos];

//randomly rotate in 3D
        randRotAngles=randU[2*PI,2*PI,2*PI];

        MX=rot3d_XRotation randRotAngles(1);
        rotPosX=rot3d_mvMul [MX, aPos atomkeys(i)];
        aSetPos [atomkeys(i), rotPosX];

        MY=rot3d_YRotation randRotAngles(2);
        rotPosY=rot3d_mvMul [MY, aPos atomkeys(i)];
        aSetPos [atomkeys(i), rotPosY];

        MZ=rot3d_ZRotation randRotAngles(3);
        rotPosZ=rot3d_mvMul [MZ, aPos atomkeys(i)];
        aSetPos [atomkeys(i), rotPosZ];

//Move monomer to a random position inside the box
        newcoords=randU[boxedge,boxedge,boxedge];

//Just for aesthetics, physically reflect all starting monomers inside the box
        newcoords=newcoords-(boxedge/2);

//randlomly translate monomer in periodic box
        molCenter=oCentroid chainkeys(i);
        aSetPos[atomkeys(i),(aPos atomkeys(i))-molCenter+newcoords];

endloop


return [chainkeys,reskeys];//Do not change this. If need to add another output you can do so.

endfunction
//================================================

local function torsionanglechoice [ichain,imon,torsionatoms,risparamvalues,tacticity,states,temp];
//This is the torsionangle function that chooses the torsion angles in the monomeric unit based on
//RIS probabilities and the Long Range Energy interaction
```

```
local anglechoices,risang,risinterdiadnormprobs,state;
local risintradiadnormprobs,LRinterdiadnormprobs,LRintradiadnormprobs;
local normprobsinterdiad,normprobsintradiad,LREnergyinterdiad,LREnergyintradiad;
local nstates=length risang;

//Assign RIS parameters
[U,Um,Ur,ang0,ang1,angm0,angm1,angr0,angr1]=risparamvalues;

//Choose the state of the interdiad angle
if tacticity(imon)==0 and tacticity(imon-1)==0 then
  risinterdiadnormprobs=Um(states(ichain)(imon-1)(2));
  risang=angm0(states(ichain)(imon-1)(2));

elseif tacticity(imon)==0 and tacticity(imon-1)==1 then
  risinterdiadnormprobs=Ur(states(ichain)(imon-1)(2));
  risang=angr0(states(ichain)(imon-1)(2));

elseif tacticity(imon)==1 and tacticity(imon-1)==0 then
  risinterdiadnormprobs=Ur(states(ichain)(imon-1)(2));
  risang=angr1(states(ichain)(imon-1)(2));

elseif tacticity(imon)==1 and tacticity(imon-1)==1 then
  risinterdiadnormprobs=Um(states(ichain)(imon-1)(2));
  risang=angm1(states(ichain)(imon-1)(2));

else
pr 'Tacticity Error - only values of 0 and 1 permitted for tacticity vector';

endif

angle1=[torsionatoms(1),torsionatoms(2),torsionatoms(3),torsionatoms(4)];

LREnergyinterdiad=LREnergyCalc[torsionatoms,risang,1];
LREnergyinterdiad=cat LREnergyinterdiad;
LRinterdiadnormprobs=exp(-LREnergyinterdiad/(KBOLTZ*temp));
LRinterdiadnormprobs=cat LRinterdiadnormprobs;

//Calculate the final normalized state probability for interdiad angles, and choose state(1) for the
//interdiad angle
normprobsinterdiad=risinterdiadnormprobs*LRinterdiadnormprobs;
normprobsinterdiad=normprobsinterdiad/(add normprobsinterdiad);
state(1)=choosestate [normprobsinterdiad];
aSetDihedralCWDeg[angle1,risang(state(1))];

//Choose the state of the intradiad angle
if tacticity(imon)==0  then
  risintradiadnormprobs=U(state(1));
  risang=ang0(states(ichain)(imon-1)(2));

elseif tacticity(imon)==1 then
  risintradiadnormprobs=U(state(1));
```

```
   risang=ang1(states(ichain)(imon-1)(2));

else
pr 'Tacticity Error';

endif

[LREnergyintradiad]=LREnergyCalc[torsionatoms,risang,2];
LRintradiadnormprobs=exp(-LREnergyintradiad/(KBOLTZ*temp));

//Calculate the final normalized state probability for interdiad angles
normprobsintradiad=risintradiadnormprobs*LRintradiadnormprobs;
normprobsintradiad=normprobsintradiad/(add normprobsintradiad);
state(2)=choosestate [normprobsintradiad];
anglechoices(2)=risang(state(2));
angle1=get[torsionatoms,[2,3,4,5]]; //atoms in the interdiad angle

aSetDihedralCWDeg[angle1,risang(state(1))];
aSetDihedralCWDeg[angle1,anglechoices(2)];

//Update the states for this chain and monomer
//'state' contains the states of the two current torsional states
//'states' contains all states for all chains and monomers

states(ichain)(imon)=[state(1),state(2)];

//This code forces the torsional states to a specific value for testing purposes
//setting all states to trans (state=1) makes it easy to see tacticity and structure

return[states];

endfunction


//*************************************************

local function LREnergyCalc[torsionatoms,risang,anglepair]
// LREnergyCalc calculates the LR energy for the RIS states

local LREnergy,anglestate;
local nstates=length risang;

//gather appropriate chain and residue keys
local chainkeys=aChain torsionatoms(1);
local prevresidue=aResidue torsionatoms(1);
local currresidue=aResidue torsionatoms(4);
local nresidues=length cResidues chainkeys(1);
local test,temp1,temp2;

if anglepair==1 then
//Determine the LR energy values for the various interdiad angles
```

```
  angle1=get[torsionatoms,[1,2,3,4]]; //atoms in the interdiad angle
//Set inert atoms for interdiad angle
  aSetInert[rAtoms prevresidue,1];
  aSetInert[aBonds torsionatoms(4),1];
  aSetInert[torsionatoms(3),1];

//Calculate the energy for the system at all 3 torsional states
  for anglestate=1,3 loop
    aSetDihedralCWDeg[angle1,risang(anglestate)];
    test=Potential[W:0,dX:0];
    LREnergy(anglestate)=test(1);
  endloop

//Make all atoms active
  aSetInert[Atoms[],0];

elseif anglepair==2 then
//Determine the LR energy values for the various intradiad angles
  angle1=get[torsionatoms,[2,3,4,5]]; //atoms in the interdiad angle
//Set inert atoms for interdiad angle
  aSetInert[rAtoms prevresidue,1];
  aSetInert[rAtoms currresidue,1];
  aSetInert[aBonds torsionatoms(4),0];
  aSetInert[torsionatoms(3),1];


//Calculate the energy for the system at all 3 torsional states
  for anglestate=1,3 loop
    aSetDihedralCWDeg[angle1,risang(anglestate)];
    test=Potential[W:0,dX:0];
    LREnergy(anglestate)=test(1);
  endloop
else
pr 'anglepair variable out of range';
endif

//Normalize the energy so the lowest value is zero
LREnergy=LREnergy-(min LREnergy);
aSetInert[Atoms[],0];

return[LREnergy];

endfunction

//*****************************************************
endfunction //This closes the main function, polybuildpp300
```

APPENDIX B. R_G^2, LRC, AND CED CALCULATION CODE

```
#svl

function radgyrandlrc[cells,boxedge,fffile,mdb,o]

//radgyrandlrc calculates structural parameters such as <Rg2> for chain halves and whole chains,
//chain end-to-end distance, characteristic ratio, and collapse ratio.
//It also calculates the long-range energy (including VDW and Coulombic) and instantaneous
//pressure (and the virial) corrections using 'CellParamPanel1_moebatch' and 'Dynamics' external
//functions, respectively.
//'CellParamPanel1_moebatch' and 'Dynamics' were adapted from MOE source code functions,
//'cell' and 'md' respectively, for our long-range correction calculations.
//Lastly, radgyrandlrc calculates the cohesive energy density by subtracting the total potential from
//the total isolated chain potential.

//cells: number of cells in one dimension of the lattice, e.g. for a final 3x3x3 lattice, 'cells' would
//be 3. The greater this number is, the more accurate the long-range correction.
//boxedge: length of periodic box side
//fffile: forcefield file name
//mdb: the path and name of the database where the structure of interest is located
//o: the entry number for the structure of interest (see mol_Create(structures(o)))

//samples:
//radgyrandlrc[4,23.03824009,'pptest2.ff','c:/users/ngarrido6/moefiles/dynamics_OPLS_mod/76m
//on2chains_48pMeso/0873density/Trial7/dynamics76mon_0873densityt7.c.500.mdb',1//01]
//radgyrandlrc[4,49.9342536857098,'pptest2.ff','c:/users/ngarrido6/moefiles/dynamics_OPLS-
//mod/170mon8chains_56pMeso/100pDensity/Trial7/dynamics170m56t7.c.500.mdb',501]

local temp,nchains,chainkeys,q,residuekeylist,atomkeylist,xyzPos,lambda,V,radgyr2val,Akey,
local Bkey,sqrendtoenddist,residuekeylist1,residuekeylist2,atomkeylist1,atomkeylist2,xyzPos1,
local xyzPos2,lambda1,lambda2,V1,V2,radgyr2val1,radgyr2val2;
local i,atomkeys,j,potentialEnergy,isolatedenergy,isolatedenergyvector,totPotSeries,xSeries;
local virialP,totalpressure,VDWU,eleU,ivar;

//set the directory to where the forcefield file and external functions (CellParamPanel1_moebatch
//and Dynamics) are located
cd 'c:/users/ngarrido6/moefiles/';

//load the forcefield file to the system and disable restraints, electic, out-of-plane, strech-bend, and
//cutoff contributions from energetic calculations
pot_Load fffile;
pot_Setup[resEnable:0,oopEnable:0,stbEnable:0,cutoffEnable:0,eleEnable:0];

//Set the temperature to 300K (or desired temperature)
```

```
temp=300;

//define chain keys and number of chains in the system and disable periodic boundaries
nchains=length Chains[];
chainkeys=Chains[];
CellEnable(0);

//Calculate <Rg2> of chain halves and whole chains, the end-to-end distance of the chains, and the
//isolated energy of the chains
for q=1,nchains loop
//obtain atom keys of first and second halves of all chains and atom keys of whole chains
        if even length cResidueCount chainkeys(q) then
                residuekeylist1=keep[cat cResidues chainkeys(q),div[cResidueCount
chainkeys(q),2]];
                residuekeylist2=keep[cat cResidues chainkeys(q),neg div[cResidueCount
chainkeys(q),2]];
        else
                residuekeylist1=keep[cat cResidues chainkeys(q),div[sub[cResidueCount
chainkeys(q),1],2]];
                residuekeylist2=keep[cat cResidues chainkeys(q),neg div[sub[cResidueCount
chainkeys(q),1],2]];
        endif
        atomkeylist=cat cAtoms chainkeys(q);
        atomkeylist1=cat rAtoms residuekeylist1;
        atomkeylist2=cat rAtoms residuekeylist2;
//obtain [x,y,z] vector positions of first half, second half, and whole chains
        xyzPos=aPos atomkeylist;
        xyzPos1=aPos atomkeylist1;
        xyzPos2=aPos atomkeylist2;
//Calculate the Covariance from previous position matrices, i.e. calculate eigenvalues and vectors
//sum eigenvalues to obtain radius of gyration values
        [lambda,V]=rot3d_CovarianceEigenSystem xyzPos;
        [lambda1,V1]=rot3d_CovarianceEigenSystem xyzPos1;
        [lambda2,V2]=rot3d_CovarianceEigenSystem xyzPos2;
        radgyr2val(q)=add lambda;
        radgyr2val1(q)=add lambda1;
        radgyr2val2(q)=add lambda2;
//calculate the end-to-end distance, i.e. the last skeletal atom position vector minus the first skeletal
//atom position vector
        Akey=mget[cat cAtoms chainkeys(q),(aName cat cAtoms chainkeys(q))=='A'];
        Bkey=mget[cat cAtoms chainkeys(q),(aName cat cAtoms chainkeys(q))=='B'];
        sqrendtoenddist(q)=sqr aDist[Akey,Bkey];
//Calculate the potential energy of each isolated chain
        aSetInert[Atoms[],1];
        aSetInert[atomkeylist,0];
        potentialEnergy=Potential[W:0,dX:0];
        isolatedenergyvector(q)=potentialEnergy(1);
endloop


//Average the mean square of the radius of gyration values for chain halves and whole chains
```

```
local Rg2Avg=(add radgyr2val)/nchains;
local firstRg2Avg=(add radgyr2val1)/nchains;
local secondRg2Avg=(add radgyr2val2)/nchains;

pr 'Square End-to-End Distance';
pr sqrendtoenddist;

//calculate the collapse ratio
local collapseratio=sqrendtoenddist/radgyr2val;

//Calculate the characteristic ratio
//l is skeletal bond length, 1.52943
local l2=sqr(1.52943);
//n is number of skeletal atoms
local n=((length mget[cat cAtoms chainkeys(1),(aName cat cAtoms chainkeys(1))=='C1'])+(length
mget[cat cAtoms chainkeys(1),(aName cat cAtoms chainkeys(1))=='C2'])+1);
local characteristicratio=sqrendtoenddist/(n*l2);

//Calculate the total chain isolated energy
isolatedenergy=add isolatedenergyvector;
pr 'isolatedenergy';
pr isolatedenergy;

//Enable cutoff and disable atom inertness from previous loop.
//If the structure has partial charges, enable Coulombic interactions (i.e. instead use:
//pot_Setup[cutoffEnable:1,cutoffOn:(0.5*boxedge-2),cutoffOff:(0.5*boxedge),eleEnable:1];)
aSetInert[Atoms[],0];
pot_Setup[cutoffEnable:1,cutoffOn:(0.5*boxedge-2),cutoffOff:(0.5*boxedge)];

//Enable periodic cube with 'boxedge' dimensions and determine its volume
CellParameters['P1',[boxedge,boxedge,boxedge],[90,90,90]];
CellEnable(1);
local boxvolume=cube boxedge;
pr 'Periodic Cell Volume(cubic Angstroms)';
pr boxvolume;

//Calculate long-range energy for a single periodic cube:
local VDWEnergy = pot_vdwEnergy Atoms[];
local eleEnergy = pot_eleEnergy Atoms[];
VDWU(1)=VDWEnergy(1);
eleU(1)=eleEnergy(1);
local v=Potential[dX:0,W:1];
totPotSeries(1)=v(1);

//Pressure and virial long-range correction calculation for a single periodic cube:
//load adapted MD svl file containing adapted 'Dynamics' function for virial and instantaneous
//Pressure calculation
load 'md_cons.svl';

//Set the md options for md and .mdb initialization. The time options in md_opt will not matter, but
//we set to a short trajectory anyway.
```

```
local
md_opt=[startTime:0,checkpointTime:1,sample:1,wrapWater:0,savePosition:1,saveVelocity:0,bo
ndConstraints:'light',verbose:1,algorithm:'NHA',opendbv:0];
//Set the directory to where you want the dummy dynamics trajectory to be saved (this database
//will have no entries since the trajectory duration will be zero ps)
cd 'c:/users/ngarrido6/moefiles/dynamics_OPLS-mod/76mon2chains_48pMeso/test';
//Initialize a dummy dynamics database
call['MD_Initialize',['pressureTest',"equil { ps=1 T=300 }",md_opt]];

//Call 'Dynamics' function which will return the virial and instantaneous pressure (kPa).
//Note that the second element of the 'Dynamics' function input is 0 indicating a trajectory time of
//zero ps.
ivar=call['Dynamics',['pressuretest.a.0.mdb',0,300,[],[],md_opt]];
virialP(1)=ivar.W;
totalpressure(1)=ivar.P;

//Loop to calculate the energy, pressure, and virial with increasing lattice dimensions
for j=2,cells loop
        cd 'c:/users/ngarrido6/moefiles';
        pr 'Current no. cells';
        pr j;
//Destroy previous structure
        oDestroy Chains[];
        CellEnable(0);
//rebuild original structure
        local structures=db_ReadColumn[mdb,'mol'];
        local chains=mol_Create(structures(o));
        CellParameters['P1', [boxedge,boxedge,boxedge], [90,90,90]];
        CellEnable(1);
//Generate a lattice of periodic images
        CellLatticeReplicate[j,j,j];
//BUILD THE CRYSTAL WITH j DIMENSIONS
        run 'CellParamPanel1_moebatch.svl';
//Update the cutoff to half the length of the periodic box side
        pot_Setup[cutoffOn:(0.5*j*boxedge-2),cutoffOff:(0.5*j*boxedge)];
//Calculate the VDW energy, Coulombic energy, and the total energy
        VDWEnergy = pot_vdwEnergy Atoms[];
        eleEnergy = pot_eleEnergy Atoms[];
        VDWU(j)=VDWEnergy(1);
        eleU(j)=eleEnergy(1);
        v=Potential[W:1,dX:0];
        totPotSeries(j)=div[v(1),cube j];
//Obtain pressure and virial using Dynamics function
        cd 'c:/users/ngarrido6/moefiles/dynamics_OPLS-mod/76mon2chains_48pMeso/test';
        call['MD_Initialize',['pressureTest',"equil { ps=1 T=300 }",md_opt]];
        ivar=call['Dynamics',['pressuretest.a.0.mdb',0,300,[],[],md_opt]];
        virialP(j)=ivar.W;
        totalpressure(j)=ivar.P;
//print the time for which an iteration completes
        pr asctime[];
        pr [totPotSeries,virialP,totalpressure];
```

```
endloop

//Calculate cohesive energy density (J/cc)
local CED=4184*((isolatedenergy-totPotSeries)/AVOGADRO)/((boxvolume)/(1E24));

//Calculate the final potential
pot_Setup[cutoffEnable:1,cutoffOn:0.5*boxedge-2,cutoffOff:0.5*boxedge];
local finalpotential=Potential[W:0,dX:0];

pr
'Rg2whole,Rg2first,Rg2second,CED,finalPotential,CollapseRatio,CharacteristicRatio,totPotSerie
s,totalpressure,virial,VDWU,eleU';
return[radgyr2val,radgyr2val1,radgyr2val2,CED,finalpotential(1),collapseratio,characteristicratio,
totPotSeries,totalpressure,virialP,VDWU,eleU];

endfunction
```

APPENDIX C. TACTICITY GENERATOR MATLAB CODE

```
function [monadVectorString,monadVector,diadVector] =
randomNumberGenerator5(nMonomers,prop0)
%   The Input is the number of monomers, the proportion of zeros.
%   The goal is to get a diad vector with 0.48 and 0.52 m:r
%   proportions (76 monomers) and 0.4793 and 0.5207 m:r (170
%   monomers) and then convert it to a monad vector.
%   Sample:
%   [monadVector,diadVector] = randomNumberGenerator5(76,0.48)
%   [monadVector,diadVector] = randomNumberGenerator5(170,81/169)
%   PS: 181monomers,180diads, 8 chains, 0.4500 or 81/180 meso

nDiads=nMonomers-1;
 for i=1:100000
    v=randsrc(1,nDiads,[0,1;prop0,1-prop0]);
    if round(sum(v==0)/length(v),4)==round(prop0,4)
        diadVector=v;
        break
    end
 end
 monadVector=convertToMonomerVector(diadVector);
 monadVectorString= ['[' sprintf(repmat('%d,', 1,
numel(monadVector)-1), monadVector(1:end-1)) sprintf('%d]',
monadVector(end))];% MOE format
end


function [monadVector]=convertToMonomerVector(diadVector)
nMonomers=1+length(diadVector);
monadVector(1)=0;
for i=2:nMonomers
    if diadVector(i-1)==0
        monadVector(i)=monadVector(i-1);
    else
        monadVector(i)=abs(monadVector(i-1)-1);
    end
end
end
```

APPENDIX D. MODIFIED FORCEFIELD

Modified OPLS-AA forcefield:

```
#moe:forcefield 2012.07
#
#       pptest.ff           OPLS molecular mechanics parameters and
#            as a unified methyl group
#
#       02-jul-2012 (pl) new format
#       29-oct-2004 (pl) added description
#       09-feb-2004 (pl) fixed torsions to new format
#       28-jun-2001 (pl) created
#
# NOTES:
#
#  *    The charge parameters contained herein are an approximation to
#       the OPLS dictionary charges (on functional groups).  To get the
#       exact electrostatics, a special charge assignment must be made.
#       In particular, HIS charges resemble engh_huber charges.
#
# REFERENCES
#
#       W. L. Jorgensen, D. S. Maxwell and J. Tirado-Rives, "Development
#       and Testing of the OPLS All-Atom Force Field on Conformational
#       Energetics and Properties of Organic Liquids", J. Am. Chem. Soc.,
#       117, 11225-11236 (1996)
#
#       D. S. Maxwell, J. Tirado-Rives and W. L. Jorgensen, "A Comprehensive
#       Study of the Rotational Energy Profiles of Organic Systems
#       by Ab Initio MO Theory, Forming a Basis for Peptide Torsional
#       Parameters", J.Comput Chem, 16, 984-1010 (1995)
#
#       W. L. Jorgensen and N. A. McDonald, "Development of an All-Atom Force
#       Field for Heterocycles. Properties of Liquid Pyridine and Diazenes
#       THEOCHEM-J. Mol. Struct., 424, 145-155 (1998)
#
#       N. A. McDonald and W. L. Jorgensen, "Development of an All-Atom
#       Force Field for Heterocycles. Properties of Liquid Pyrrole, Furan,
#       Diazoles, and Oxazoles", J. Phys. Chem. B, 102, 8049-8059 (1998)'
#
#       R. C. Rizzo and W. L. Jorgensen, "OPLS All-Atom Model for Amines:
#       Resolution of the Amine Hydration Problem", J. Am. Chem. Soc.,
#       121, 4827-4836 (1999)

title      'OPLS-mod'
disable    stb oop

info desc 'An all-atom forcefield parameterized for proteins and'
info desc 'some small organic molecules.  Partial charges are'
info desc 'based on bond-charge increments that reproduce'
info desc 'the original dictionary charges. Polar hydrogens have'
```

info desc 'with Br unified methyl.

info ref 'Jorgensen, W.L., et al.; J. Am. Chem. Soc. 117, 1996, 11225-11236'
info ref 'D. S. Maxwell D.S., et al.; J. Comput. Chem 16, 1995, 984-1010'
info ref 'Jorgensen, W.L., et al.; -. Mol. Struct. 424, 1998, 145-155'
info ref 'McDonald, N.A., et al.; J. Phys. Chem. B 102, 1998, 8049-8059'
info ref 'Rizzo, R.C et al.; J. Am. Chem. Soc. 121, 1999, 4827-4836'

type	CT	C	'C sp3 alkane'	# [1]
type	HC	H	'H on C alkane'	# [2]
type	CM	C	'C sp2 alkene'	# [3]
type	HM	H	'H on C alkene'	# [4]
type	CA	C	'C benzene'	# [5]
type	HA	H	'H on C benzene'	# [6]
type	OH	O	'OH alcohol'	# [7]
type	HO	H	'H on oxygen'	# [8]
type	CF3	C	'CF3 trifluoroethanol'	# [9]
type	OHa	O	'OH phenol/OHCCF3/diol/hemiacetal'	# [10]
type	F3	F	'F trifluoroethanol'	# [11]
type	OS	O	'O ether'	# [12]
type	OSa	O	'O acetal'	# [13]
type	CTa	C	'C acetal'	# [14]
type	SH	S	'S thiol'	# [15]
type	SH2	S	'S hydrogen sulfide'	# [16]
type	S	S	'S sulfide/disulfide'	# [17]
type	HS	H	'H on sulfur'	# [18]
type	NT	N	'N sp3 3-coord	# [19]
type	HN	H	'H on H sp3 3-coord'	# [20]
type	C=O	C	'C carbonyl'	# [21]
type	O=C	O	'O carbonyl'	# [22]
type	Nam	N	'N in NC=O'	# [23]
type	Ham	H	'H in NC=O and HID TRP HIP'	# [24]
type	HCi	H	'H in HC(=O)NC=O'	# [25]
type	OHc	O	'H in carboxylic acid'	# [26]
type	CO2	C	'C in carboxylate'	# [27]
type	O2	O	'O in carboxylate'	# [28]
type	HCd	H	'H in aldehyde and HC-C=O'	# [29]
type	N3	N	'N ammonium'	# [30]
type	H3	H	'H in ammonium'	# [31]
type	NGD+	N	'N in guanidinium'	# [32]
type	CGD+	C	'C in guanidinium'	# [33]
type	C*	C	'C TRP CG'	# [34]
type	CB	C	'C TRP CD2'	# [35]
type	CN	C	'C TRP CE2'	# [36]
type	NA	N	'N in aro NH1 in TRP, HIS, HIP'	# [37]
type	CP	C	'C in ncn HIS/CE'	# [38]
type	CV	C	'C (beta) HIS'	# [39]
type	CW	C	'C (alpha) HIS'	# [40]
type	NB	N	'N in aro NH0 in HIS, HIP'	# [41]
type	CPCD	C	'C PRO CD'	# [42]
type	Ot3p	O	'O in TIP3P water'	# [43]
type	HW	H	'H in water'	# [44]
type	Ospc	O	'O in SPC water'	# [45]
type	NH3	N	'N in ammonia'	# [46]
type	OHM	N	'O in hydroxide ion'	# [47]
type	Li+	Li	'Li+ Lithium Ion'	# [48]

```
type    Na+     Na      'Na+ Sodium Ion'                # [49]
type    K+      K       'K+ Potassium Ion'                      # [50]
type    RB+     Rb      'Rb+ Rubidium Ion'                      # [51]
type    Cs+     Cs      'Cs+ Cesium Ion'                # [52]
type    Mg+2    Mg      'Mg+2 Magnesium Ion'                    # [53]
type    Ca+2    Ca      'Ca+2 Calcium Ion'                      # [54]
type    Sr+2    Sr      'Sr+2 Strontium Ion'                    # [55]
type    Ba+2    Ba      'Ba+2 Barium Ion'                       # [56]
type    F-      F       'F- Fluoride Ion'               # [57]
type    Cl-     Cl      'Cl- Chloride Ion'              # [58]
#type   Br-     Br      'Br- Bromide Ion'               # [59]
type    Br      Br      'Br as unified methyl                   # [59]
type    He      He      'Helium Atom'                           # [60]
type    Ne      Ne      'Neon Atom'                             # [61]
type    Ar      Ar      'Argon Atom'                            # [62]
type    Kr      Kr      'Krypton Atom'                          # [63]
type    Xe      Xe      'Xenon Atom'                            # [64]

itor CA C=O Nam CO2 C* CB CN NA CP CV CW

[rules] # type pattern matching rules

   # hydrogen

HC      match '[#1]-[C;X4!i]'                           # [ 2]
HM      match '[#1]-[C;X3i]=C'                          # [ 4]
HA      match '[#1]-c'                                  # [ 6] H on C benzene

HO      match '[#1]-[#8]'               # [ 8]
HS      match '[#1]-[#16]'                              # [18] H on sulfur
HN      match '[#1]-[N;X3!i]'                           # [20] H on N sp3 3-coord
H3      match '[#1]-[NX4]'                              # [31] H in ammonium

Ham     match '[#1]-[N;X3!i]C=[OX1]'                    # [24] H in NC=O
Ham     match '[#1]-n'                                  # [24] H in TRP/HIS

HCi     match '[#1]-C(=[OX1])[N;X3!i]C=[OX1]'           # [25] H in formimide
HCd     match '[#1]-C=[OX1]'                            # [29] H in aldehyde
#HCd    match '[#1]-C-C=[OX1]'                          # [29] H in HC-C=O'

   # carbon

CT      match '[C;X4!i]'                # [ 1]
CTa     match '[C;X4]([OX2!i])[OX2!i]'                  # [14] C acetal
CM      match '[C;X3i]=C'                               # [ 3] C alkene
C=O     match '[C;X3i]=[OX1]'                           # [21] C carbonyl
CO2     match '[C;X3i](=[OX1])[OX1]'                    # [27] C in carboxylate

CA      match 'c'                       # [ 5] C benzene
CF3     match '[C;X4!i](F)(F)(F)'       # [ 9]

CV      match '[cX3r5]1:[a;!i2]:[a;i2]:[a;!i2]:[a;!i2]:1'       # [39] c r5 B
CW      match '[cX3r5]1:[a;i2]:[a;!i2]:[a;!i2]:[a;!i2]:1'       # [40] c r5 A

CP      match '[cX3r5]1:n:c:c:n:1'      # [38] C in ncn HIS
CW      match '[cX3r5]1:c:[nX3]:c:[nX3]:1'# [40] C in HIS+ not ncn not ncn
```

```
CPCD    match '[CX4]1[CX4][CX4][CX4Q3][NQ3]1C=O'      # [42] C PRO CD

   # oxygen

OH      match '[O;X2H1]-[CX4]'                        # [ 7]
OHa     match '[O;X2H1]-[CX4]-C(F)(F)(F)'             # [10] thrifluoroethanol
OHa     match '[O;X2H1]-[CX4]-[O!i]'                  # [10] hemiacetal/diol
OHa     match '[O;X2H1]-[i]'                          # [10] enol/phenol
OHc     match '[O;X2H1]-C=[OX1]'                      # [26] OH in carboxylic acid

OS      match '[O;X2Q2]'                    # [12] O ether
OSa     match '[O;X2Q2]-[CX4]-[O!i]'                  # [13] O acetal

OHM     match '[OX1H1]'                               # [47] O in hydroxide ion

O=C     match '[OX1]=[CX3]'                           # [22] O carbonyl
O2      match '[OX1]=[CX3]-[OX1]'                     # [28] O in carboxylate
O2      match '[OX1]-[CX3]=[OX1]'                     # [28] O in carboxylate

   # nitrogen

NH3     match '[N;X3!iH3]'                            # [46] N in ammonia
NT      match '[N;X3!i!Q0]'                           # [19] N sp3 3-coord
N3      match '[NX4]'                                 # [30] N in ammonium

Nam     match '[N;X3!i]C=O'                           # [23] N in NC=O

NA      match '[nH;r5]'                               # [37] N-H1 (aro)
NB      match '[nX2;i1]'                    # [41] N-H0 (aro)

   # sulfur

SH      match '[S;X2H1]'                    # [15] S thiol
SH2     match '[S;X2H2]'                    # [15] S hydrogen sulfide
S       match '[S;X2Q2]'                    # [17] sulfide/disulfide

   # misc.

F3      match 'FC(F)F'                                # [11] F trifluroethanol

   # metals and ions

Li+     match '[Li+;X0]'          # Li+
Na+     match '[Na+;X0]'          # Na+
K+      match '[K+;X0]'                    # K+
RB+     match '[Rb+;X0]'          # Rb+
Cs+     match '[Cs+;X0]'          # Cs+
Mg+2    match '[Mg+2;X0]'                  # Cs+2
Ca+2    match '[Ca+2;X0]'                  # Cs+2
Sr+2    match '[Sr+2;X0]'         # Cs+2
Ba+2    match '[Ba+2;X0]'                  # Cs+2
F-      match '[F-;X0]'                    # F-
Cl-     match '[Cl-;X0]'          # Cl-
#Br-    match '[Br-;X0]'          # Br-
Br      match     '[Br]'                    # Br as unified methyl
```

```
He      match '[He;X0]'                 # He
Ne      match '[Ne;X0]'                 # Ne
Ar      match '[Ar;X0]'                 # Ar
Kr      match '[Kr;X0]'                 # Kr
Xe      match '[Xe;X0]'                 # Xe

   # water

Ot3p    match   '[O;X2H2]'              # TIP3P water O
HW      match   '[#1]-[O;X2H2]'         # TIP3P water H
#Ospc   match   '[O;X2H2]'              # SPC water O

   # guanidinium special case

CGD+    match '[C+0](=[N+!$(*[-])])[N!i+0]'                     # [33]
CGD+    match '[C+1!i]([NX3!i])[NX3!i]'

NGD+    match '[NX3!i][C+0](=[NX3+!$(*[-])])[NX3!i]'            # [32]
NGD+    match '[NX3+!$(*[-])]=[C+0]([NX3!i])[NX3!i]'
NGD+    match '[NX3!i][C+1!i]([NX3!i])[NX3!i]'
NGD+    match '[NX3!i]-[c;r5]1:[nX3+!$(*[-*])]:a:a:[nX3]1'

H3      match '[#1]-[NX3!i][C+0](=[NX3+!$(*[-])])[NX3!i]'
H3      match '[#1]-[NX3+!$(*[-])]=[C+0]([NX3!i])[NX3!i]'
H3      match '[#1]-[NX3!i][C+1!i]([NX3!i])[NX3!i]'
H3      match '[#1]-[N!i+0]-[c;r5]1:[n+!$(*[-*])]:a:a:[i2]1'

   # TRP: 5-ring carbons

C*      match '[cQ3]1:[cQ3]2:[cQ3](:c:c:c:c:2):n:c:1'
CB      match '[cQ3]1(:[cQ3]2:n:c:[cQ3]:1):c:c:c:c:2'
CN      match '[cQ3]1(:n:c:[cQ3]:[cQ3]2:1):c:c:c:c:2'
CA      match '[cQ2]1:n:[cQ3]2:[cQ3](:c:c:c:c:2):c:1'

[str] # bond stretch parameters

* CT    CT      1.5290  268.0000    0.0000    0.0000  0
* Br    CT      1.5290  268.0000    0.0000    0.0000  0
* CT    HC      1.0900  340.0000    0.0000    0.0000  0.0600
* CT    CA      1.5100  317.0000    0.0000    0.0000 -0.1150
* CT    OH      1.4100  320.0000    0.0000    0.0000 -0.2650
* CT    SH      1.8100  222.0000    0.0000    0.0000 -0.1800
* CT    S       1.8100  222.0000    0.0000    0.0000 -0.2175
* CT    C=O     1.5220  317.0000    0.0000    0.0000  0
* CT    Nam     1.4490  337.0000    0.0000    0.0000  0
* CT    CO2     1.5220  317.0000    0.0000    0.0000  0.1000
* CT    N3      1.4710  367.0000    0.0000    0.0000 -0.3100
* CT    NGD+    1.4630  337.0000    0.0000    0.0000 -0.3400
* CT    C*      1.4950  317.0000    0.0000    0.0000  0
* CT    CV      1.5040  317.0000    0.0000    0.0000 -0.1150
* CT    CW      1.5040  317.0000    0.0000    0.0000 -0.1150
* CT    CPCD    1.5290  268.0000    0.0000    0.0000  0
* HC    CPCD    1.0900  340.0000    0.0000    0.0000 -0.0600
* CA    CA      1.4000  469.0000    0.0000    0.0000  0
* CA    HA      1.0800  367.0000    0.0000    0.0000  0.1150
* CA    OHa     1.3640  450.0000    0.0000    0.0000 -0.1500
```

* CA	C*	1.3520	546.0000	0.0000	0.0000	0
* CA	CB	1.4000	469.0000	0.0000	0.0000	0
* CA	CN	1.4000	469.0000	0.0000	0.0000	0
* CA	NA	1.3810	427.0000	0.0000	0.0000	0
* HA	CP	1.0800	367.0000	0.0000	0.0000	-0.1150
* HA	CV	1.0800	367.0000	0.0000	0.0000	-0.1150
* HA	CW	1.0800	367.0000	0.0000	0.0000	-0.1150
* OH	HO	0.9450	553.0000	0.0000	0.0000	0.4180
* HO	OHa	0.9450	553.0000	0.0000	0.0000	-0.4350
* SH	HS	1.3360	274.0000	0.0000	0.0000	0.2550
* S	S	2.0380	166.0000	0.0000	0.0000	0
* C=O	O=C	1.2290	570.0000	0.0000	0.0000	-0.5000
* C=O	Nam	1.3350	490.0000	0.0000	0.0000	0
* C=O	HCd	1.0900	340.0000	0.0000	0.0000	0
* Nam	Ham	1.0100	434.0000	0.0000	0.0000	0.3000
* Nam	CPCD	1.4490	337.0000	0.0000	0.0000	0.0700
* CO2	O2	1.2500	656.0000	0.0000	0.0000	-0.3000
* N3	H3	1.0100	434.0000	0.0000	0.0000	0.3300
* H3	NGD+	1.0100	434.0000	0.0000	0.0000	-0.4600
* NGD+	CGD+	1.3400	481.0000	0.0000	0.0000	-0.1200
* C*	CB	1.4590	388.0000	0.0000	0.0000	-0.0750
* CB	CN	1.4190	447.0000	0.0000	0.0000	-0.0200
* CN	NA	1.3800	428.0000	0.0000	0.0000	-0.1500
* Ot3p	HW	0.9572	529.6000	0.0000	0.0000	0.4170
* HW	Ospc	1.0000	527.2000	0.0000	0.0000	-0.4100
* Ham	NA	1.0100	434.0000	0.0000	0.0000	-0.2580
* NA	CP	1.3430	477.0000	0.0000	0.0000	0.2000
* NA	CW	1.3810	427.0000	0.0000	0.0000	0.2000
* CP	NB	1.3350	488.0000	0.0000	0.0000	-0.2000
* CV	CW	1.3710	518.0000	0.0000	0.0000	-0.1600
* CV	NB	1.3940	410.0000	0.0000	0.0000	-0.2000
* CW	CW	1.3710	518.0000	0.0000	0.0000	0

[ang] # angle bend parameters

* CT	CT	CT	112.700	58.3500	0.0000	0.0000
* CT	CT	Br	112.700	58.3500	0.0000	0.0000
* CT	CT	HC	110.700	37.5000	0.0000	0.0000
* CT	CT	CA	114.000	63.0000	0.0000	0.0000
* CT	CT	OH	109.500	50.0000	0.0000	0.0000
* CT	CT	SH	108.600	50.0000	0.0000	0.0000
* CT	CT	S	114.700	50.0000	0.0000	0.0000
* CT	CT	C=O	111.100	63.0000	0.0000	0.0000
* CT	CT	Nam	109.700	80.0000	0.0000	0.0000
* CT	CT	CO2	111.100	63.0000	0.0000	0.0000
* CT	CT	N3	111.200	80.0000	0.0000	0.0000
* CT	CT	NGD+	111.200	80.0000	0.0000	0.0000
* CT	CT	C*	115.600	63.0000	0.0000	0.0000
* CT	CT	CV	114.000	63.0000	0.0000	0.0000
* CT	CT	CW	114.000	63.0000	0.0000	0.0000
* CT	CT	CPCD	112.700	58.3500	0.0000	0.0000
* HC	CT	HC	107.800	33.0000	0.0000	0.0000
* HC	CT	CA	109.500	35.0000	0.0000	0.0000
* HC	CT	OH	109.500	35.0000	0.0000	0.0000
* HC	CT	SH	109.500	35.0000	0.0000	0.0000
* HC	CT	S	109.500	35.0000	0.0000	0.0000

* HC	CT	C=O	109.500	35.0000	0.0000	0.0000
* HC	CT	Nam	109.500	35.0000	0.0000	0.0000
* HC	CT	CO2	109.500	35.0000	0.0000	0.0000
* HC	CT	N3	109.500	35.0000	0.0000	0.0000
* HC	CT	NGD+	109.500	35.0000	0.0000	0.0000
* HC	CT	C*	109.500	35.0000	0.0000	0.0000
* HC	CT	CV	109.500	35.0000	0.0000	0.0000
* HC	CT	CW	109.500	35.0000	0.0000	0.0000
* HC	CT	CPCD	110.700	37.5000	0.0000	0.0000
* C=O	CT	Nam	110.100	63.0000	0.0000	0.0000
* C=O	CT	N3	111.200	80.0000	0.0000	0.0000
* Nam	CT	CO2	110.100	63.0000	0.0000	0.0000
* CO2	CT	N3	111.200	80.0000	0.0000	0.0000
* CT	CA	CA	120.000	70.0000	0.0000	0.0000
* CA	CA	CA	120.000	63.0000	0.0000	0.0000
* CA	CA	HA	120.000	35.0000	0.0000	0.0000
* CA	CA	OHa	120.000	70.0000	0.0000	0.0000
* CA	CA	CB	120.000	63.0000	0.0000	0.0000
* CA	CA	CN	120.000	85.0000	0.0000	0.0000
* HA	CA	C*	120.000	35.0000	0.0000	0.0000
* HA	CA	CB	120.000	35.0000	0.0000	0.0000
* HA	CA	CN	120.000	35.0000	0.0000	0.0000
* HA	CA	NA	121.600	35.0000	0.0000	0.0000
* C*	CA	NA	108.700	70.0000	0.0000	0.0000
* CT	OH	HO	108.500	55.0000	0.0000	0.0000
* CA	OHa	HO	113.000	35.0000	0.0000	0.0000
* CT	SH	HS	96.000	44.0000	0.0000	0.0000
* CT	S	CT	98.900	62.0000	0.0000	0.0000
* CT	S	S	103.700	68.0000	0.0000	0.0000
* CT	C=O	O=C	120.400	80.0000	0.0000	0.0000
* CT	C=O	Nam	116.600	70.0000	0.0000	0.0000
* O=C	C=O	Nam	122.900	80.0000	0.0000	0.0000
* O=C	C=O	HCd	120.000	35.0000	0.0000	0.0000
* Nam	C=O	HCd	115.000	40.0000	0.0000	0.0000
* CT	Nam	CT	118.000	50.0000	0.0000	0.0000
* CT	Nam	C=O	121.900	50.0000	0.0000	0.0000
* CT	Nam	Ham	118.400	38.0000	0.0000	0.0000
* CT	Nam	CPCD	118.000	50.0000	0.0000	0.0000
* C=O	Nam	Ham	119.800	35.0000	0.0000	0.0000
* C=O	Nam	CPCD	121.900	50.0000	0.0000	0.0000
* Ham	Nam	Ham	120.000	35.0000	0.0000	0.0000
* CT	CO2	O2	117.000	70.0000	0.0000	0.0000
* O2	CO2	O2	126.000	80.0000	0.0000	0.0000
* CT	N3	CT	113.000	50.0000	0.0000	0.0000
* CT	N3	H3	109.500	35.0000	0.0000	0.0000
* H3	N3	H3	109.500	35.0000	0.0000	0.0000
* CT	NGD+	H3	118.400	35.0000	0.0000	0.0000
* CT	NGD+	CGD+	123.200	50.0000	0.0000	0.0000
* H3	NGD+	H3	120.000	35.0000	0.0000	0.0000
* H3	NGD+	CGD+	120.000	35.0000	0.0000	0.0000
* NGD+	CGD+	NGD+	120.000	70.0000	0.0000	0.0000
* CT	C*	CA	125.000	70.0000	0.0000	0.0000
* CT	C*	CB	128.600	70.0000	0.0000	0.0000
* CA	C*	CB	106.400	85.0000	0.0000	0.0000
* CA	CB	C*	134.900	85.0000	0.0000	0.0000
* CA	CB	CN	116.200	85.0000	0.0000	0.0000

* C*	CB	CN	108.800	85.0000	0.0000	0.0000
* CA	CN	CB	122.700	85.0000	0.0000	0.0000
* CA	CN	NA	132.800	70.0000	0.0000	0.0000
* CB	CN	NA	104.400	70.0000	0.0000	0.0000
* CA	NA	Ham	120.000	35.0000	0.0000	0.0000
* CA	NA	CN	111.600	70.0000	0.0000	0.0000
* Ham	NA	CN	123.100	35.0000	0.0000	0.0000
* Ham	NA	CP	126.350	35.0000	0.0000	0.0000
* Ham	NA	CW	126.350	35.0000	0.0000	0.0000
* CP	NA	CW	107.300	70.0000	0.0000	0.0000
* HA	CP	NA	120.000	35.0000	0.0000	0.0000
* HA	CP	NB	120.000	35.0000	0.0000	0.0000
* NA	CP	NA	110.750	70.0000	0.0000	0.0000
* NA	CP	NB	111.600	70.0000	0.0000	0.0000
* CT	CV	CW	130.700	70.0000	0.0000	0.0000
* HA	CV	CW	128.200	35.0000	0.0000	0.0000
* CT	CV	NB	124.500	70.0000	0.0000	0.0000
* HA	CV	NB	120.000	35.0000	0.0000	0.0000
* CW	CV	NB	108.300	70.0000	0.0000	0.0000
* CT	CW	NA	121.600	70.0000	0.0000	0.0000
* CT	CW	CV	130.700	70.0000	0.0000	0.0000
* CT	CW	CW	130.700	70.0000	0.0000	0.0000
* HA	CW	NA	121.600	35.0000	0.0000	0.0000
* HA	CW	CV	130.700	35.0000	0.0000	0.0000
* HA	CW	CW	130.700	35.0000	0.0000	0.0000
* NA	CW	CV	108.700	70.0000	0.0000	0.0000
* NA	CW	CW	106.300	70.0000	0.0000	0.0000
* CP	NB	CV	105.300	70.0000	0.0000	0.0000
* CT	CPCD	HC	110.700	37.5000	0.0000	0.0000
* CT	CPCD	Nam	109.700	80.0000	0.0000	0.0000
* HC	CPCD	HC	107.800	33.0000	0.0000	0.0000
* HC	CPCD	Nam	109.500	35.0000	0.0000	0.0000
* HW	Ot3p	HW	104.520	34.0500	0.0000	0.0000
* HW	Ospc	HW	109.470	37.9500	0.0000	0.0000

[ptor] # torsion parameters

tor-format V1 V2 P2 V3

* CT	CT	CT	CT	0.870	-0.079	180	0.140
* CT	CT	CT	Br	0.870	-0.079	180	0.140
* CT	CT	CT	HC	0.000	0.000	0	0.183
* CT	CT	CT	S	1.310	-0.310	180	0.129
* CT	CT	CT	C=O	-0.849	-0.228	180	0.292
* CT	CT	CT	Nam	0.422	-0.481	180	0.356
* CT	CT	CT	CO2	-1.593	-0.412	180	0.246
* CT	CT	CT	N3	1.366	-0.115	180	0.242
* CT	CT	CT	NGD+	0.982	0.000	0	0.330
* CT	CT	CT	CPCD	0.870	-0.079	180	0.140
* HC	CT	CT	HC	0.000	0.000	0	0.159
* HC	CT	CT	CA	0.000	0.000	0	0.231
* HC	CT	CT	OH	0.000	0.000	0	0.234
* HC	CT	CT	SH	0.000	0.000	0	0.226
* HC	CT	CT	S	0.000	0.000	0	0.226
* HC	CT	CT	C=O	0.000	0.000	0	-0.038
* HC	CT	CT	Nam	0.000	0.000	0	0.232

* HC	CT	CT	CO2	0.000	0.000	0	-0.113
* HC	CT	CT	N3	0.000	0.000	0	0.192
* HC	CT	CT	NGD+	0.000	0.000	0	0.232
* HC	CT	CT	C*	0.000	0.000	0	0.231
* HC	CT	CT	CV	0.000	0.000	0	0.231
* HC	CT	CT	CW	0.000	0.000	0	0.231
* HC	CT	CT	CPCD	0.000	0.000	0	0.183
* CA	CT	CT	C=O	-0.849	-0.228	180	0.292
* CA	CT	CT	Nam	0.422	-0.481	180	0.356
* CA	CT	CT	CO2	-0.849	-0.228	180	0.292
* CA	CT	CT	N3	0.422	-0.481	180	0.356
* OH	CT	CT	C=O	-3.090	0.000	0	0.000
* OH	CT	CT	Nam	3.140	-0.734	180	1.015
* OH	CT	CT	CO2	-3.090	0.000	0	0.000
* OH	CT	CT	N3	3.140	-0.734	180	1.015
* SH	CT	CT	C=O	-2.107	-1.057	180	0.484
* SH	CT	CT	Nam	0.291	-0.582	180	0.070
* SH	CT	CT	CO2	-2.107	-1.057	180	0.484
* SH	CT	CT	N3	0.291	-0.582	180	0.070
* S	CT	CT	C=O	-2.107	-1.057	180	0.484
* S	CT	CT	Nam	0.291	-0.582	180	0.070
* S	CT	CT	CO2	-2.107	-1.057	180	0.484
* S	CT	CT	N3	0.291	-0.582	180	0.070
* C=O	CT	CT	C=O	-0.849	-0.228	180	0.292
* C=O	CT	CT	Nam	0.422	-0.481	180	0.356
* C=O	CT	CT	CO2	-0.849	-0.228	180	0.292
* C=O	CT	CT	N3	0.422	-0.481	180	0.356
* C=O	CT	CT	C*	-0.849	-0.228	180	0.292
* C=O	CT	CT	CV	-0.849	-0.228	180	0.292
* C=O	CT	CT	CW	-0.849	-0.228	180	0.292
* Nam	CT	CT	CO2	0.422	-0.481	180	0.356
* Nam	CT	CT	C*	0.422	-0.481	180	0.356
* Nam	CT	CT	CV	0.422	-0.481	180	0.356
* Nam	CT	CT	CW	0.422	-0.481	180	0.356
* CO2	CT	CT	CO2	0.422	-0.481	180	0.356
* CO2	CT	CT	N3	0.422	-0.481	180	0.356
* CO2	CT	CT	C*	-0.849	-0.228	180	0.292
* N3	CT	CT	C*	0.422	-0.481	180	0.356
* CT	CT	CA	CA	0.000	0.000	0	0.000
* HC	CT	CA	CA	0.000	0.000	0	0.000
* CT	CT	OH	HO	-0.178	-0.087	180	0.246
* HC	CT	OH	HO	0.000	0.000	0	0.225
* CT	CT	SH	HS	-0.380	-0.141	180	0.301
* HC	CT	SH	HS	0.000	0.000	0	0.226
* CT	CT	S	CT	0.463	-0.288	180	0.339
* CT	CT	S	S	0.971	-0.418	180	0.468
* HC	CT	S	CT	0.000	0.000	0	0.324
* HC	CT	S	S	0.000	0.000	0	0.279
* CT	CT	C=O	O=C	0.000	0.000	0	0.000
* CT	CT	C=O	Nam	0.587	0.095	180	-0.600
* HC	CT	C=O	O=C	0.000	0.000	0	0.000
* HC	CT	C=O	Nam	0.000	0.000	0	0.000
* Nam	CT	C=O	O=C	0.000	0.000	0	0.000
* Nam	CT	C=O	Nam	0.908	0.611	180	0.790
* N3	CT	C=O	O=C	0.000	0.000	0	0.000
* N3	CT	C=O	Nam	0.908	0.611	180	0.790

* CT	CT	Nam	C=O	0.000	0.231	180	0.000
* CT	CT	Nam	Ham	0.000	0.000	0	0.000
* CT	CT	Nam	CPCD	2.377	-0.367	180	0.000
* HC	CT	Nam	CT	0.000	0.000	0	0.000
* HC	CT	Nam	C=O	0.000	0.000	0	0.000
* HC	CT	Nam	Ham	0.000	0.000	0	0.000
* HC	CT	Nam	CPCD	0.000	0.000	0	0.000
* C=O	CT	Nam	C=O	-1.183	0.456	180	-0.425
* C=O	CT	Nam	Ham	0.000	0.000	0	0.000
* C=O	CT	Nam	CPCD	-0.869	0.625	180	-1.750
* CO2	CT	Nam	C=O	-1.183	0.456	180	-0.425
* CO2	CT	Nam	Ham	0.000	0.000	0	0.000
* CO2	CT	Nam	CPCD	-0.869	0.625	180	-1.750
* CT	CT	CO2	O2	0.000	0.410	180	0.000
* HC	CT	CO2	O2	0.000	0.000	0	0.000
* Nam	CT	CO2	O2	0.000	0.000	0	0.000
* N3	CT	CO2	O2	0.000	0.000	0	0.000
* CT	CT	N3	CT	0.870	-0.079	180	0.140
* CT	CT	N3	H3	0.000	0.000	0	0.173
* HC	CT	N3	CT	0.000	0.000	0	0.183
* HC	CT	N3	H3	0.000	0.000	0	0.131
* C=O	CT	N3	CT	-0.849	-0.228	180	0.292
* C=O	CT	N3	H3	0.000	0.000	0	0.173
* CO2	CT	N3	H3	0.000	0.000	0	0.173
* CT	CT	NGD+	H3	0.000	0.000	0	0.000
* CT	CT	NGD+	CGD+	0.914	0.121	180	-0.249
* HC	CT	NGD+	H3	0.000	0.000	0	0.000
* HC	CT	NGD+	CGD+	0.000	0.000	0	0.088
* CT	CT	C*	CA	-0.357	0.000	0	0.000
* CT	CT	C*	CB	0.000	0.000	0	0.000
* HC	CT	C*	CA	0.000	0.000	0	-0.240
* HC	CT	C*	CB	0.000	0.000	0	0.000
* CT	CT	CV	CW	0.000	0.000	0	0.000
* CT	CT	CV	NB	1.183	-0.131	180	0.253
* HC	CT	CV	CW	0.000	0.000	0	0.000
* HC	CT	CV	NB	0.000	0.000	0	0.209
* CT	CT	CW	NA	1.183	-0.131	180	0.253
* CT	CT	CW	CV	0.000	0.000	0	0.000
* CT	CT	CW	CW	0.000	0.000	0	0.000
* HC	CT	CW	NA	0.000	0.000	0	0.209
* HC	CT	CW	CV	0.000	0.000	0	0.000
* HC	CT	CW	CW	0.000	0.000	0	0.000
* CT	CT	CPCD	HC	0.000	0.000	0	0.183
* CT	CT	CPCD	Nam	0.422	-0.481	180	0.356
* HC	CT	CPCD	HC	0.000	0.000	0	0.159
* HC	CT	CPCD	Nam	0.000	0.000	0	0.232
* CT	CA	CA	CA	0.000	3.625	180	0.000
* CT	CA	CA	HA	0.000	3.625	180	0.000
* CA	CA	CA	CA	0.000	3.625	180	0.000
* CA	CA	CA	HA	0.000	3.625	180	0.000
* CA	CA	CA	OHa	0.000	3.625	180	0.000
* CA	CA	CA	CB	0.000	3.625	180	0.000
* CA	CA	CA	CN	0.000	3.625	180	0.000
* HA	CA	CA	HA	0.000	3.625	180	0.000
* HA	CA	CA	OHa	0.000	3.625	180	0.000
* HA	CA	CA	CB	0.000	3.625	180	0.000

* HA	CA	CA	CN	0.000	3.625	180	0.000
* CA	CA	OHa	HO	0.000	0.841	180	0.000
* HA	CA	C*	CT	0.000	6.525	180	0.000
* HA	CA	C*	CB	0.000	6.525	180	0.000
* NA	CA	C*	CT	0.000	6.525	180	0.000
* NA	CA	C*	CB	0.000	6.525	180	0.000
* CA	CA	CB	C*	0.000	3.500	180	0.000
* CA	CA	CB	CN	0.000	3.500	180	0.000
* CA	CA	CN	CB	0.000	3.625	180	0.000
* CA	CA	CN	NA	0.000	3.625	180	0.000
* HA	CA	CB	C*	0.000	3.500	180	0.000
* HA	CA	CB	CN	0.000	3.500	180	0.000
* HA	CA	CN	CB	0.000	3.625	180	0.000
* HA	CA	CN	NA	0.000	3.625	180	0.000
* HA	CA	NA	Ham	0.000	1.500	180	0.000
* HA	CA	NA	CN	0.000	1.500	180	0.000
* C*	CA	NA	Ham	0.000	1.500	180	0.000
* C*	CA	NA	CN	0.000	1.500	180	0.000
* CT	S	S	CT	0.000	-3.707	180	0.853
* CT	C=O	Nam	CT	1.150	3.045	180	0.000
* CT	C=O	Nam	Ham	0.000	2.450	180	0.000
* CT	C=O	Nam	CPCD	1.150	3.045	180	0.000
* O=C	C=O	Nam	CT	0.000	3.045	180	0.000
* O=C	C=O	Nam	Ham	0.000	2.450	180	0.000
* O=C	C=O	Nam	CPCD	0.000	3.045	180	0.000
* HCd	C=O	Nam	CT	0.000	3.045	180	0.000
* HCd	C=O	Nam	Ham	0.000	2.450	180	0.000
* HCd	C=O	Nam	CPCD	0.000	3.045	180	0.000
* CT	Nam	CPCD	CT	1.429	1.029	180	-5.633
* CT	Nam	CPCD	HC	0.000	0.000	0	0.000
* C=O	Nam	CPCD	CT	0.000	0.231	180	0.000
* C=O	Nam	CPCD	HC	0.000	0.000	0	0.000
* CT	NGD+	CGD+	NGD+	0.000	3.968	180	0.000
* H3	NGD+	CGD+	NGD+	0.000	1.950	180	0.000
* CT	C*	CB	CA	0.000	1.675	180	0.000
* CT	C*	CB	CN	0.000	1.675	180	0.000
* CA	C*	CB	CA	0.000	1.675	180	0.000
* CA	C*	CB	CN	0.000	1.675	180	0.000
* CA	CB	CN	CA	0.000	3.000	180	0.000
* CA	CB	CN	NA	0.000	3.000	180	0.000
* C*	CB	CN	CA	0.000	3.000	180	0.000
* C*	CB	CN	NA	0.000	3.000	180	0.000
* CA	CN	NA	CA	0.000	1.525	180	0.000
* CA	CN	NA	Ham	0.000	1.525	180	0.000
* CB	CN	NA	CA	0.000	1.525	180	0.000
* CB	CN	NA	Ham	0.000	1.525	180	0.000
* Ham	NA	CP	HA	0.000	2.325	180	0.000
* Ham	NA	CP	NA	0.000	2.325	180	0.000
* Ham	NA	CP	NB	0.000	2.325	180	0.000
* CW	NA	CP	HA	0.000	2.325	180	0.000
* CW	NA	CP	NA	0.000	2.325	180	0.000
* CW	NA	CP	NB	0.000	2.325	180	0.000
* Ham	NA	CW	CT	0.000	1.400	180	0.000
* Ham	NA	CW	HA	0.000	1.600	180	0.000
* Ham	NA	CW	CV	0.000	1.400	180	0.000
* Ham	NA	CW	CW	0.000	1.400	180	0.000

```
* CP    NA    CW    CT      0.000  1.400 180  0.000
* CP    NA    CW    HA      0.000  1.600 180  0.000
* CP    NA    CW    CV      0.000  1.400 180  0.000
* CP    NA    CW    CW      0.000  1.400 180  0.000
* HA    CP    NB    CV      0.000  5.000 180  0.000
* NA    CP    NB    CV      0.000  5.000 180  0.000
* CT    CV    CW    HA      0.000  5.375 180  0.000
* CT    CV    CW    NA      0.000  5.375 180  0.000
* HA    CV    CW    CT      0.000  5.375 180  0.000
* HA    CV    CW    NA      0.000  5.375 180  0.000
* NB    CV    CW    CT      0.000  5.375 180  0.000
* NB    CV    CW    HA      0.000  5.375 180  0.000
* NB    CV    CW    NA      0.000  5.375 180  0.000
* CT    CV    NB    CP      0.000  2.400 180  0.000
* HA    CV    NB    CP      0.000  2.400 180  0.000
* CW    CV    NB    CP      0.000  2.400 180  0.000
* CT    CW    CW    HA      0.000  5.375 180  0.000
* CT    CW    CW    NA      0.000  5.375 180  0.000
* HA    CW    CW    NA      0.000  5.375 180  0.000
* NA    CW    CW    NA      0.000  5.375 180  0.000

[itor] # improper torsion parameters

* CA    CA    CA    CT      0.000  -2.200   0.000   0.000   0.000
* CA    CA    CA    HA      0.000  -2.200   0.000   0.000   0.000
* CA    CA    CA    OHa     0.000  -2.200   0.000   0.000   0.000
* CA    CN    CA    HA      0.000  -2.200   0.000   0.000   0.000
* CB    CA    CA    HA      0.000  -2.200   0.000   0.000   0.000
* NA    C*    CA    HA      0.000  -2.200   0.000   0.000   0.000
* Nam   CT    C=O   O=C     0.000 -21.000   0.000   0.000   0.000
* Nam   HCd   C=O   O=C     0.000 -21.000   0.000   0.000   0.000
* CT    C=O   Nam   Ham     0.000  -2.000   0.000   0.000   0.000
* CT    C=O   Nam   CPCD    0.000  -2.000   0.000   0.000   0.000
* Ham   C=O   Nam   Ham     0.000  -2.000   0.000   0.000   0.000
* O2    CT    CO2   O2      0.000 -21.000   0.000   0.000   0.000
* CB    CA    C*    CT      0.000  -2.200   0.000   0.000   0.000
* C*    CA    CB    CN      0.000  -2.200   0.000   0.000   0.000
* NA    CB    CN    CA      0.000  -2.200   0.000   0.000   0.000
* CN    CA    NA    Ham     0.000  -2.000   0.000   0.000   0.000
* CW    CP    NA    Ham     0.000  -2.000   0.000   0.000   0.000
* NA    NA    CP    HA      0.000  -2.200   0.000   0.000   0.000
* NA    NB    CP    HA      0.000  -2.200   0.000   0.000   0.000
* CW    NB    CV    CT      0.000  -2.200   0.000   0.000   0.000
* CW    NB    CV    HA      0.000  -2.200   0.000   0.000   0.000
* NA    CV    CW    CT      0.000  -2.200   0.000   0.000   0.000
* NA    CV    CW    HA      0.000  -2.200   0.000   0.000   0.000
* NA    CW    CW    CT      0.000  -2.200   0.000   0.000   0.000
* NA    CW    CW    HA      0.000  -2.200   0.000   0.000   0.000

[nonbonded] # nonbonded parameters

vdw-function   12-6
vdw-mix-radius geometric
vdw-mix-well   geometric
vdw-scale14    0.5
vdw-buffer     0 0
```

```
ele-dielectric 1 distance
ele-scale14    0.5
ele-buffer     0
ele-charge-fcn pot_ChargeBCI

CT        1.964 0.0660 0 0  12.0000 -  0.0000 0 -
#The Br below is a unified methyl that uses
#the CT parameters increased by 10%
Br        2.160 0.1826 0 0  12.0000 -  0.0000 0 -
HC        1.403 0.0300 0 0   1.0080 -  0.0000 0 -
CM        1.992 0.0760 0 0  12.0000 -  0.0000 0 -
HM        1.358 0.0300 0 0   1.0080 -  0.0000 0 -
CA        1.992 0.0700 0 0  12.0000 -  0.0000 0 -
HA        1.358 0.0300 0 0   1.0080 -  0.0000 0 -
OH        1.751 0.1700 0 0  15.9990 -  0.0000 0 -
HO        1.000 1e-7   0 0   1.0080 -  0.0000 0 -
CF3       1.824 0.0620 0 0  12.0000 -  0.0000 0 -
OHa       1.723 0.1700 0 0  15.9990 -  0.0000 0 -
F3        1.650 0.0610 0 0  18.9980 -  0.0000 0 -
OS        1.628 0.1400 0 0  15.9990 -  0.0000 0 -
OSa       1.684 0.1400 0 0  15.9990 -  0.0000 0 -
CTa       1.964 0.0660 0 0  12.0000 -  0.0000 0 -
SH        1.992 0.2500 0 0  32.0660 -  0.0000 0 -
SH2       2.077 0.2500 0 0  32.0660 -  0.0000 0 -
S         1.992 0.2500 0 0  32.0660 -  0.0000 0 -
HS        1.000 1e-7   0 0   1.0080 -  0.0000 0 -
NT        1.824 0.1700 0 0  14.0070 -  0.0000 0 -
HN        1.000 1e-7   0 0   1.0080 -  0.0000 0 -
C=O       2.105 0.1050 0 0  12.0000 -  0.0000 0 -
O=C       1.661 0.2100 0 0  15.9990 -  0.0000 0 -
Nam       1.824 0.1700 0 0  14.0070 -  0.0000 0 -
Ham       0.000 0.0000 0 0   1.0080 -  0.0000 0 -
HCi       1.403 0.0200 0 0   1.0080 -  0.0000 0 -
OHc       1.684 0.1700 0 0  15.9990 -  0.0000 0 -
CO2       2.105 0.1050 0 0  12.0000 -  0.0000 0 -
O2        1.661 0.2100 0 0  15.9990 -  ?      0 -
HCd       1.358 0.0150 0 0   1.0080 -  0.0000 0 -
N3        1.824 0.1700 0 0  14.0070 -  1.0000 0 -
H3        1.000 1e-7   0 0   1.0080 -  0.0000 0 -
NGD+      1.824 0.1700 0 0  14.0070 -  0.0000 0 -
CGD+      1.263 0.0500 0 0  12.0000 -  1.0000 0 -
C*        1.992 0.0700 0 0  12.0000 -  0.0000 0 -
CB        1.992 0.0700 0 0  12.0000 -  0.0000 0 -
CN        1.992 0.0700 0 0  12.0000 -  0.0000 0 -
NA        1.824 0.1700 0 0  14.0070 -  ?      0.25 -
CP        1.992 0.0700 0 0  12.0000 -  0.0000 0 -
CV        1.992 0.0700 0 0  12.0000 -  0.0000 0 -
CW        1.992 0.0700 0 0  12.0000 -  0.0000 0 -
NB        1.824 0.1700 0 0  14.0070 -  0.0000 0 -
CPCD      1.964 0.0660 0 0  12.0000 -  0.0000 0 -
Ot3p      1.768 0.1521 0 0  15.9990 -  0.0000 0 -
HW        1.000 1e-7   0 0   1.0080 -  0.0000 0 -
Ospc      1.777 0.1554 0 0  15.9990 -  0.0000 0 -
NH3       1.886 0.2100 0 0  14.0070 -  0.0000 0 -
OHM       1.796 0.2500 0 0  15.9990 - -1.0000 0 -
```

```
Li+     1.193 0.0183 0 0   6.9410 -  1.0000 0 -
Na+     1.869 0.0028 0 0  22.9900 -  1.0000 0 -
K+      2.769 0.0003 0 0  39.0980 -  1.0000 0 -
RB+     3.155 0.0002 0 0  85.4680 -  1.0000 0 -
Cs+     3.769 0.0001 0 0 132.9050 -  1.0000 0 -
Mg+2    0.923 0.8750 0 0  24.3050 -  2.0000 0 -
Ca+2    1.354 0.4497 0 0  40.0780 -  2.0000 0 -
Sr+2    1.741 0.1182 0 0  87.6200 -  2.0000 0 -
Ba+2    2.142 0.0471 0 0 137.3270 -  2.0000 0 -
F-      1.534 0.7200 0 0  18.9980 - -1.0000 0 -
Cl-     2.479 0.1180 0 0  35.4530 - -1.0000 0 -
#Br-    2.595 0.0900 0 0  79.9040 - -1.0000 0 -
He      1.435 0.0200 0 0   4.0030 -  0.0000 0 -
Ne      1.560 0.0690 0 0  20.1790 -  0.0000 0 -
Ar      1.909 0.2339 0 0  39.9480 -  0.0000 0 -
Kr      2.034 0.3170 0 0  83.8000 -  0.0000 0 -
Xe      2.208 0.4330 0 0 131.3000 -  0.0000 0 -

[eof]
[vdw] # specific van der waals parameters

#T1     T2      rad     eps     m  n
#---    ----    -------  -------  -- --

*       *       -       -       12  6
```

APPENDIX E. JOURNAL LICENSES/PERMISSIONS

Figure App-D.1. License for Tables 2-1 and 2-3.

REFERENCES

1. Delony, J.; Ludovice, P.; Henderson, C., *Understanding and mitigating bridge defects in block copolymer directed self-assembly through computational materials design and optimization*. SPIE: 2020; Vol. 11326.

2. Lawson, R. A.; Ludovice, P. J.; Henderson, C. L. In *Development of realistic potentials for the simulation of directed self-assembly of PS-PMMA di-block copolymers*, Alternative Lithographic Technologies III, International Society for Optics and Photonics: 2011; p 79700N.

3. Lawson, R. A.; Peters, A. J.; Ludovice, P. J.; Henderson, C. L. In *Coarse grained molecular dynamics model of block copolymer directed self-assembly*, Alternative Lithographic Technologies V, International Society for Optics and Photonics: 2013; p 86801Y.

4. Nation, B. D.; Ludovice, P. J.; Henderson, C. L., Block copolymer directed self-assembly using chemoepitaxial guiding underlayers with topography. *Journal of Vacuum Science & Technology B, Nanotechnology and Microelectronics: Materials, Processing, Measurement, and Phenomena* **2017,** *35* (6), 06GI01.

5. Peters, A. J.; Lawson, R. A.; Ludovice, P. J.; Henderson, C. L. In *Effects of block copolymer polydispersity and xN on pattern line edge roughness and line width roughness from directed self-assembly of diblock copolymers*, Alternative Lithographic Technologies V, International Society for Optics and Photonics: 2013; p 868020.

6. Peters, A. J.; Nation, B. D.; Nicoloso, D.; Ludovice, P. J.; Henderson, C. L., Protracted colored noise dynamics applied to linear polymer systems. *Macromolecular Theory and Simulations* **2018,** *27* (2), 1700062.

7. Lawson, R.; Peters, A.; Nation, B.; Ludovice, P.; Henderson, C., Simulation study of the effect of differences in block energy and density on the self-assembly of block copolymers. *Journal of Micro/Nanolithography, MEMS, and MOEMS* **2014,** *13* (3), 031308.

8. Craig, G. S.; Thode, C. J.; Onses, M. S.; Nealey, P. F., The Use of Block Copolymers in Nanoscale Patterning. *Sigma Aldrich* **2011**.

9. Kozuch, D. J.; Zhang, W.; Milner, S. T., Predicting the Flory-Huggins χ Parameter for Polymers with Stiffness Mismatch from Molecular Dynamics Simulations. *Polymers* **2016,** *8* (6), 241.

10. Peters, A. J. Mesoscale Simulation of Block Copolymer Phase Seperation and Directed Self-Assembly Processes: Applications for Semiconductor Manufacturing. Georgia Institute of Technology, Atlanta, GA, 2015.

11. Allen, M. P., Tildesley, D. J., *Computer simulation of liquids*. Oxford Oxfordshire : Clarendon Press

New York : Oxford University Press: 1987.

12. Kotelyanskii, M. T., D. N. , *Simulation Method for Polymers*. 1 ed.; CRC Press: Boca Raton, 2004; p 900.

13. Dong, Q.; Li, W., Effect of Molecular Asymmetry on the Formation of Asymmetric Nanostructures in ABC-Type Block Copolymers. *Macromolecules* **2021,** *54* (1), 203-213.

14. Antoniadis, S. J.; Samara, C. T.; Theodorou, D. N., Molecular Dynamics of Atactic Polypropylene Melts. *Macromolecules* **1998,** *31* (22), 7944-7952.

15. Antoniadis, S. J.; Samara, C. T.; Theodorou, D. N., Effect of Tacticity on the Molecular Dynamics of Polypropylene Melts. *Macromolecules* **1999,** *32* (25), 8635-8644.

16. Peter Atkins, P.; De Paula, J., *Atkins' physical chemistry*. OUP Oxford: 2014.

17. Smith, G. D.; Ludovice, P. J.; Jaffe, R. L.; Yoon, D. Y., Conformations of 2, 4-Dichloropentane and 2, 4, 6-Trichloroheptane and a Force Field for Poly (vinyl chloride) Based upon ab Initio Electronic Structure Calculations. *The Journal of Physical Chemistry* **1995,** *99* (1), 164-172.

18. Greenfield, M. L.; Theodorou, D. N., Geometric analysis of diffusion pathways in glassy and melt atactic polypropylene. *Macromolecules* **1993,** *26* (20), 5461-5472.

19. Logotheti, G. E.; Theodorou, D. N., Segmental and Chain Dynamics of Isotactic Polypropylene Melts. *Macromolecules* **2007,** *40* (6), 2235-2245.

20. Panagiotou, E.; Tzoumanekas, C.; Lambropoulou, S.; Millett, K. C.; Theodorou, D. N., A study of the entanglement in systems with periodic boundary conditions. *Progress of Theoretical Physics Supplement* **2011,** *191*, 172-181.

21. Theodorou, D. N.; Suter, U. W., Detailed molecular structure of a vinyl polymer glass. *Macromolecules* **1985,** *18* (7), 1467-1478.

22. Theodorou, D. N.; Suter, U. W., Shape of unperturbed linear polymers: polypropylene. *Macromolecules* **1985,** *18* (6), 1206-1214.

23. Theodorou, D. N.; Suter, U. W., Atomistic modeling of mechanical properties of polymeric glasses. *Macromolecules* **1986,** *19* (1), 139-154.

24. Pascua, O. G.; Ahumada, O.; Laso, M.; Müller, M., The Effect of the Initial Guess Generator on Molecular Mechanics Calculations. *Molecular Simulation* **2003,** *29* (3), 187-199.

25. Lyons, E. P. Computer Simulation of Poly(ethylene terephthalate) and Derivatives Structure and their Ramifications for Gas Transport. Georgia Institute of Technology, Atlanta, GA, 2000.

26. ULC, C. C. G. *Molecular Operating Environment (MOE) Documentation*, 2019.01; 1010 Sherbrooke St. West, Suite #910, Montreal, QC, Canada H3A2R7, 2019.

27. Eckstein, A.; Suhm, J.; Friedrich, C.; Maier, R. D.; Sassmannshausen, J.; Bochmann, M.; Mülhaupt, R., Determination of Plateau Moduli and Entanglement Molecular Weights of Isotactic, Syndiotactic, and Atactic Polypropylenes Synthesized with Metallocene Catalysts. *Macromolecules* **1998,** *31* (4), 1335-1340.

28. van Meerveld, J., A method to extract the monomer friction coefficient from the linear viscoelastic behavior of linear, entangled polymer melts. *Rheologica Acta* **2004,** *43* (6), 615-623.

29. Theodorou, D. N.; Boone, T. D.; Dodd, L. R.; Mansfield, K. F., Stress tensor in model polymer systems with periodic boundaries. *Macromolecular Theory and Simulations* **1993,** *2* (2), 191-238.

30. Mark, J. E., *Physical properties of polymers handbook*. Second edition.. ed.; New York : Springer: 2007.

31. Huang, C.-L.; Chen, Y.-C.; Hsiao, T.-J.; Tsai, J.-C.; Wang, C., Effect of Tacticity on Viscoelastic Properties of Polystyrene. *Macromolecules* **2011,** *44* (15), 6155-6161.

32. Suter, U. W.; Flory, P. J., Conformational Energy and Configurational Statistics of Polypropylene. *Macromolecules* **1975,** *8* (6), 765-776.

33. Sperling, L. H., *Introduction to physical polymer science*. Fourth edition.. ed.; Hoboken, N.J. : Wiley: 2006.

34. Flory, P. J.; Volkenstein, M., Statistical mechanics of chain molecules. *Biopolymers* **1969,** *8* (5), 699-700.

35. Rehahn, M., Mattice, W.L., Suter, U.W., *Rotational isomeric state models in macromolecular systems*. Berlin

New York : Springer: 1997.

36. Volkenshtein, M. V., *Molecular biophysics / M. V. Volkenstein*. Academic Press: New York, 1977.

37. Rapold, R. F.; Suter, U. W.; Theodorou, D. N., Static atomistic modelling of the structure and ring dynamics of bulk amorphous polystyrene. *Macromolecular Theory and Simulations* **1994,** *3* (1), 19-43.

38. Biskup, U.; Cantow, H. J., Calculations on the Unperturbed Dimensions of Polypropylenes. *Macromolecules* **1972,** *5* (5), 546-550.

39. Flory, P.; Fujiwara, Y., Configurational Statistics and Stereochemical Structure of Vinyl Polymers. *Macromolecules* **1969,** *2* (4), 315-327.

40. Kuhn, B.; Ehrig, M.; Ahlrichs, R., Atomistic Modeling of Amorphous Polymer Bulk Based on an ab Initio Optimized Force Field. *Macromolecules* **1996,** *29* (11), 4051-4059.

41. Ramos, J.; Vega, J. F.; Theodorou, D. N.; Martinez-Salazar, J., Entanglement Relaxation Time in Polyethylene: Simulation versus Experimental Data. *Macromolecules* **2008,** *41* (8), 2959-2962.

42. Tirrell, M., Polymer Self-Diffusion in Entangled Systems. *Rubber Chemistry and Technology* **1984,** *57* (3), 523-556.

43. Sandler, S. I., *An introduction to applied statistical thermodynamics*. John Wiley & Sons: 2010.

44. Krevelen, D. W. v., *Properties of polymers, their estimation and correlation with chemical structure*. Second, completely revised edition.. ed.; Amsterdam

New York : Elsevier Scientific Pub. Co.: 1976.

45. Zirkel, A.; Urban, V.; Richter, D.; Fetters, L. J.; Huang, J. S.; Kampmann, R.; Hadjichristidis, N., Small-angle neutron scattering evaluation of the temperature dependence of atactic polypropylene and poly(1-butene) chain dimensions in the melt. *Macromolecules* **1992,** *25* (23), 6148-6155.

46. Sun, Q.; Faller, R., Crossover from Unentangled to Entangled Dynamics in a Systematically Coarse-Grained Polystyrene Melt. *Macromolecules* **2006,** *39* (2), 812-820.

47. Dee, G. T.; Sauer, B. B., The cohesive energy density of polymers and its relationship to surface tension, bulk thermodynamic properties, and chain structure. *Journal of Applied Polymer Science* **2017,** *134* (5).

48. Barr, C.; Diez, D. M.; Rundel, C., OpenIntro statistics. Opentextbookstore: 2016.

49. Frenkel, D.; Smit, B., *Understanding Molecular Simulation: From Algorithms to Applications*. Elsevier Science: 2001.

50. Ewald, P. P., Ewald summation. *Ann. Phys* **1921,** *369* (253), 1-2.2.

51. Jenkins, J. W. Novel Efficient Simulation Techniques for Use in Molecular Modeling. Georgia Institute of Technology, Atlanta, GA, 2000.

52. Abe, Y.; Tonelli, A.; Flory, P., Optical Anisotropy of Vinyl Polymer Chains. I. Strain Birefringence of Polypropylene and Polystyrene. *Macromolecules* **1970,** *3* (3), 294-303.

53. Allegra, G.; Calligaris, M.; Randaccio, L.; Moraglio, G., Isotactic Polypropylene. Unperturbed Dimensions Calculated Accounting for All Skeletal Rotations. *Macromolecules* **1973,** *6* (3), 397-403.

54. Aoki, H.; Mori, K.; Ito, S., Conformational analysis of single polymer chains in three dimensions by super-resolution fluorescence microscopy. *Soft Matter* **2012,** *8* (16), 4390-4395.

55. Bahar, I.; Mattice, W. L., Dynamics of conformational transitions to isomeric states favoring intramolecular excimer formation in polymeric chains. Application to dimer models of polystyrene. *The Journal of Chemical Physics* **1989,** *90* (11), 6775-6782.

56. Ball, R. C.; Doi, M.; Edwards, S. F.; Warner, M., Elasticity of entangled networks. *Polymer* **1981,** *22* (8), 1010-1018.

57. Chapman, D. V.; Du, H.; Lee, W. Y.; Wiesner, U. B., Optical super-resolution microscopy in polymer science. *Progress in polymer science* **2020,** *111* (C), 101312.

58. Chatzieleftheriou, S.; Anogiannakis, S.; Theodorou, D. N.; Lagaros, N. D., SimNano: A Trust Region Strategy for Large-Scale Molecular Systems Energy Minimization Based on Exact Second-Order Derivative Information. *Journal of Chemical Information and Modeling* **2019,** *59* (1), 190-205.

59. Chen, X.; Yuan, C.; Wong, C. K. Y.; Zhang, G., Molecular modeling of temperature dependence of solubility parameters for amorphous polymers. *J Mol Model* **2012,** *18* (6), 2333-2341.

60. Doi, M.; Edwards, S. F., Dynamics of concentrated polymer systems. Part 1.—Brownian motion in the equilibrium state. *Journal of the Chemical Society, Faraday Transactions 2: Molecular and Chemical Physics* **1978,** *74* (0), 1789-1801.

61. Edwards, S., The theory of rubber elasticity. *British Polymer Journal* **1977,** *9* (2), 140-143.

62. Ferro, D.; Zambelli, A.; Provasoli, A.; Locatelli, P.; Rigamonti, E., Correlation between 13C NMR Chemical Shifts and the Conformation of Polymers. 2. An Improved Method of Calculation. *Macromolecules* **1980,** *13* (1), 179-186.

63. Ferro, D. R.; Ragazzi, M., Analysis of vicinal proton-proton coupling constants in polypropylene chains. *Macromolecules* **1981,** *14* (6), 1830-1831.

64. Florian, S.; Lathova, E.; Lath, D., Unperturbed Characterisitics of Atactic Polypropylene: Some New Theta Solvents Based on Aliphatic Ketones. *Journal of Macromolecular Science, Part A* **1999,** *36* (11), 1771-1781.

65. Flory, P., On the Stereochemical Constitution and Nuclear Magnetic Resonance Spectra of Polypropylenes. *Macromolecules* **1970,** *3* (5), 613-617.

66. Gabrieli, A.; Sant, M.; Demontis, P.; Suffritti, G. B., A combined energy-force fitting procedure to develop DFT-based force fields. *The Journal of Physical Chemistry C* **2016,** *120* (46), 26309-26319.

67. Gill, P. E.; Murray, W.; Wright, M. H., *Practical optimization*. SIAM: 2019.

68. Hearst, J. E., Statistical mechanics of chain molecules. Paul J. Flory. Interscience, New York, 1969 432 pp. $17.50. *Journal of Polymer Science Part B: Polymer Letters* **1970,** *8* (4), 305-306.

69. Heatley, F.; Salovey, R.; Bovey, F., Polymer Nuclear Magnetic Resonance Spectroscopy. XVIII. The Nuclear Magnetic Resonance Spectrum, Dimensions, and Steric Interactions of Isotactic Polypropylene. *Macromolecules* **1969,** *2* (6), 619-623.

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

71. Kotelyanskii, M.; Wagner, N. J.; Paulaitis, M. E., Building Large Amorphous Polymer Structures: Atomistic Simulation of Glassy Polystyrene. *Macromolecules* **1996,** *29* (26), 8497-8506.

72. Kröger, M., Shortest multiple disconnected path for the analysis of entanglements in two-and three-dimensional polymeric systems. *Computer physics communications* **2005,** *168* (3), 209-232.

73. Labute, P., LowModeMD—Implicit Low-Mode Velocity Filtering Applied to Conformational Search of Macrocycles and Protein Loops. *Journal of Chemical Information and Modeling* **2010,** *50* (5), 792-800.

74. Li, J.; Rincon-Delgadillo, P. A.; Suh, H. S.; Mannaert, G.; Nealey, P. F., Understanding Kinetics of Defect Annihilation in Chemoepitaxy-Directed Self-Assembly. *ACS Applied Materials & Interfaces* **2021,** *13* (21), 25357-25364.

75. Mark, J.; Curro, J.; DeBolt, L., Effects of stereochemical structure on distribution functions for short polypropylene and poly (vinyl chloride) chains. *Macromolecules* **1986,** *19* (2), 491-492.

76. Mays, J. W.; Fetters, L. J., Temperature coefficients of unperturbed dimensions for atactic polypropylene and alternating poly (ethylene-propylene). *Macromolecules* **1989,** *22* (2), 921-926.

77. Provasoli, A.; Ferro, D., Correlation between 13C NMR Chemical Shifts and Conformation of Polymers. I. The Methyl Spectrum of 3, 5, 7, 9, 11, 13, 15-Heptamethylheptadecane. *Macromolecules* **1977,** *10* (4), 874-877.

78. Resconi, L.; Jones, R. L.; Rheingold, A. L.; Yap, G. P. A., High-Molecular-Weight Atactic Polypropylene from Metallocene Catalysts. 1. Me2Si(9-Flu)2ZrX2 (X = Cl, Me). *Organometallics* **1996,** *15* (3), 998-1005.

79. Robyr, P.; Tomaselli, M.; Grob-Pisano, C.; Meier, B. H.; Ernst, R. R.; Suter, U. W., Characterization of Local Order in Atactic Polystyrene Using Two-Dimensional Nuclear Magnetic Resonance and Atomistic Simulations. *Macromolecules* **1995,** *28* (15), 5320-5324.

80. Schaefer, D.; Spiess, H.; Suter, U.; Fleming, W., Two-dimensional solid-state NMR studies of ultraslow chain motion: glass transition in atactic poly (propylene) versus helical jumps in isotactic poly (propylene). *Macromolecules* **1990,** *23* (14), 3431-3439.

81. Schilling, F.; Tonelli, A., Carbon-13 nuclear magnetic resonance of atactic polypropylene. *Macromolecules* **1980,** *13* (2), 270-275.

82. Smith, A. A.; Gujrati, P., Solubility in compressible polymers: Beyond the regular solution theory. *European polymer journal* **2007,** *43* (2), 425-459.

83. Spyriouni, T.; Tzoumanekas, C.; Theodorou, D.; Müller-Plathe, F.; Milano, G., Coarse-grained and reverse-mapped united-atom simulations of long-chain atactic polystyrene melts: Structure, thermodynamic properties, chain conformation, and entanglements. *Macromolecules* **2007,** *40* (10), 3876-3885.

84. Suter, U. W.; Flory, P. J., Optical anisotropy of polystyrene and its low molecular analogues. *Journal of the Chemical Society, Faraday Transactions 2: Molecular and Chemical Physics* **1977,** *73* (7), 1521-1537.

85. Theodorou, D. N.; Suter, U. W., Local structure and the mechanism of response to elastic deformation in a glassy polymer. *Macromolecules* **1986,** *19* (2), 379-387.

86. Tonelli, A., Stereosequence-dependent carbon-13 NMR chemical shifts in polystyrene. *Macromolecules* **1983,** *16* (4), 604-607.

87. Tonelli, A.; Abe, Y.; Flory, P., Optical Anisotropy of Vinyl Polymer Chains. II. Depolarized Scattering by Polypropylene and Polystyrene. *Macromolecules* **1970,** *3* (3), 303-308.

88. Tonelli, A. E., Calculated γ effects on the 13C-NMR spectra of 3, 5, 7, 9, 11, 13, 15-heptamethylheptadecane stereoisomers and their implications for the conformational characteristics of polypropylene. *Macromolecules* **1978,** *11* (3), 565-567.

89. Tzoumanekas, C.; Theodorou, D. N., Topological Analysis of Linear Polymer Melts: A Statistical Approach. *Macromolecules* **2006,** *39* (13), 4592-4604.

90. ULC, C. C. G. *SVL Exchange, Scientific Vector Language (SVL) source code*, 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2019.