Multi-scale Modeling of Cellulosic Polymers for Optimal Drug Delivery Properties in Solid Dispersion Formulation

by

Wenjun Huang

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Chemical Engineering)
in the University of Michigan
2017

Doctoral Committee:

Professor Ronald G. Larson, Chair
Professor Sharon C. Glotzer
Associate Professor Angela Violi
Assistant Professor Fei Wen

Wenjun Huang

wenjunh@umich.edu

ORCID ID: 0000-0001-8948-7262

2017

To my Dearest Parents and Xueqing

Acknowledgments

I want to first and foremost thank my advisor Prof. Ronald Larson for taking me on and allowing me to work on such an impactful and challenging doctoral research project. I have had a wonderful research experience in the past five years. The depth and breadth of Ron's knowledge has never ceased to amaze me. I greatly appreciate his work ethic. He holds weekly meetings with every student and guided students through their toughest moments in the research. He has always challenged me to strive for fundamental and comprehensive understanding of a subject before moving on to the next one. Ron has been a great mentor to me, both professionally and personally. He has supported me when I applied for a second degree, pursued side research interest, and took the summer internship to further my professional development. He has shared many personal stories and has invited students to his house every summer, which will remain in my memories in many years to come.

I acknowledge the funding source of this doctoral work, the Dow Chemical Company, and the people I have worked with from Dow, William "Trey" Porter, Valeriy Ginzburg, Jonathan Moore, and Robert Sammler. We had many stimulating and fruitful discussions during both over-the-phone and on-site meetings. I also appreciate the team at Dow to extend the collaboration to University of Minnesota so that I had the opportunities to share my work with the research groups at Minnesota. I want to thank Prof. Timothy Lodge, Prof. Frank Bates, Prof. Theresa Reineke, Prof. Marc Hillmyer, and Jeffery Ting for their suggestions and inputs on the direction of this doctoral work.

I thank Prof. Sharon Glotzer, Prof. Angela Violi, and Prof. Fei Wen to serve on my doctoral committee and to provide feedback throughout the thesis work. Fei has provided many useful inputs on my side research project.

I appreciate the great support from my group members and friends in Ann Arbor, who have added much laughter to my time in Michigan. I want to acknowledge Indranil Dalel, Prateek Jha, Shihu Wang, Ali Salehi, and Taraknath Mandal to help me solve challenging problems throughout my thesis work.

Finally, I thank my parents, Jian Huang and Yunxia Pan, for providing support and resource that have led me to my Ph.D. study at Michigan. You have always taught me to strive for excellence in my work. I also thank my wife Xueqing Ji for her encouragement and support over the years.

Table of Contents

Dedication ii

Acknowledgments iii

List of Figures vi

List of Tables xiv

Abstract xvi

Chapter 1: Introduction 1

1.1 Overview of the Oral Drug Delivery 1

1.2 Review of the Polymeric Excipient 2

1.3 Review of Cellulosic Polymers 3

1.4 Review of Computational Simulation of Polymeric System 5

1.5 Project Overview 7

Chapter 2: Simulation Models of Cellulosic Polymers 8

2.1 Introduction 8

2.2 Atomistic Model and Simulation Details 9

2.3 Atomistic Model Validation 10

2.4 Coarse-Grained Simulation Details 13

2.5 Coarse-Grained Methylcellulose Model 14

2.6 Coarse-Grained Model Validation 27

2.7 Coarse-Grained Hydroxypropyl Methylcellulose Acetate Succinate (HPMCAS) Model 31

2.8 Coarse-Grained Phenytoin Model 36

2.9 Coarse-Grained Phenytoin Model Validation 39

2.10 Concluding Remark 41
Chapter 3: Modeling Commercial Methylcellulose 42
3.1 Introduction 42
3.2 Atomistic Simulation Results 43
3.3 Coarse-Grained Simulation Results 51
3.4 Analytical Semi-Flexible Polymer Model 61
3.5 Analytical Model Prediction and Comparison with Simulation Results 68
3.6 Concluding Remark 71
Chapter 4: Modeling Hydroxypropyl Methylcellulose Acetate Succinate – Phenytoin Solid Dispersion Formulation 73
4.1 Introduction 73
4.2 Simulation Results for HPMCAS 74
4.3 Simulation of HPMCAS and Phenytoin 79
4.4 Inhibition of Drug Aggregation in HPMCAS-Phenytoin Complex 85
4.5 Simulation of Phenytoin Release from Polymer-Drug Complex 90
4.6 Analytical Model for Phenytoin Release 93
4.7 Concluding Remark 98
Chapter 5: Conclusions and Future Directions 100
5.1 Conclusions 100
5.2 Future Directions 102
References 105

List of Figures

Figure 1.1: Typical dissolution profile of drug only (red), drug with ineffective polymeric excipient (blue), and drug with effective polymer excipient (orange) systems. The typical time scale of the release is in hours, and the typical concentration scale is several hundreds of micrograms per milliliter. 3

Figure 1.2: Schematic of cellulose and methylcellulose, where n is even interger. Anhydroglucose unit (AGU) is the monomer unit of cellulose, where each R group is a hydrogen. In monomers of methylcellulose, one or more R groups are methyl groups (-CH_3). Both AGUs and methylcellulose monomers are connected via a β-(1→4) linkage.The primes indicate atoms on the second (right) monomer in any two consecutive monomers in the cellulosic chain, although the two monomers are bonded identically if they are not the terminal monomers (see text for details). O-2, O-3, and O-6 are the three reactive hydroxyl groups (-OH) on each AGU that can be substituted with different functional groups (i.e. methyl groups) so that the AGU becomes a methylcellulose monomer. The blue dashed lines indicate the two most predominant intramolecular hydrogen bonds that are present in cellulose. The AGUs can be also modified into a HPMCAS monomer by adding methyl, hydroxypropyl, acetyl, and succinyl functional groups. 5

Figure 2.1: Average of dihedral angles φ and ψ from simulations of cellobiose in water using GROMOS 45A4, GROMOS 56Acarbo, and AMBER 03 force fields, along with the two pairs of experimentally determined average values <φ> and <ψ> from the literature. (φ = ∠O-5–C-1–O-4'–C-4' and ψ = ∠C-1–O-4'–C-4'–C-5' where the prime denotes the atom on the neighboring monomer).....11

Figure 2.2: Simulated radii of gyration (R_g) of MC chains with the GROMOS 45A4 force field and fits to the Kratky-Porod model, using various persistence lengths lp. 12

Figure 2.3: Schematics of the methylcellulose coarse grained model. Each methylcellulose monomer (DS ranging from 0 to 3) is represented by one bead centered at the monomer center-of-mass (COM). The beads are connected via hard harmonic springs.....15

Figure 2.4: a) Intramolecular atomistic monomer RDFs obtained from atomistic simulations of 10-mer single chain homogenous methylcellulose. The RDF is an average of RDFs obtained from eight single homopolymer chain systems. Each homopolymer consists of a chain of 10 monomers of one of the eight methylcellulose monomer

substitution types. b), c), and d) are zoomed-in view of the second, third, and fourth peak of those shown in a) at 1.0, 1.5, and 2.0nm respectively. 16

Figure 2.5: a) Intermolecular atomistic monomer RDFs for 10 wt% 2,6-MC with different chain lengths at 25℃. To reduce statistical noise, we averaged RDFs of 5 and 10-mers (red), 15 and 20-mers (green), and 25 and 30-mers (blue). b) Same as a) except these RDFs were generated at 50℃.....20

Figure 2.6: Fits of the intermolecular CG RDFs to intermolecular atomistic monomer RDFs for four homo-methylcellulose systems. The atomistic systems are cellulose, 3-MC, 2,6-MC, and 2,3,6-MC. The atomistic RDFs were obtained by computing the intermolecular monomer COM-monomer COM RDF for each 10wt% 15-mer homo-oligomer system. The CG RDFs were computed by the same method were fit to match the position (r_p) and height ($g(r_p)$) of the first peak in atomistic RDFs. 22

Figure 2.7: a) Intermolecular monomer RDF fitting with different cutoff settings applied to the b) truncated and shifted analytical LJ 9-6 potential. We choose Cutoff = 1.8 to be the cutoff distance throughout this study by setting $rc\_ii = rp/0.7$. The σii value is kept constant in all cases, and εii is adjusted to maintain the fitting of first peak and the well depth in the potential.....23

Figure 2.8: Snapshots of aggregation at 50 ℃ of CG chains and their atomistic counterparts. All systems contained 10wt% 20-mer homo-oligomers. The snapshots were obtained after simulation times of 40ns (atomistic) and $6x10^6$ steps (coarse grain). 23

Figure 2.9: a) Extrapolation of the εii values to long chain lengths. Blue and red stars are fitted values obtained from corresponding atomistic simulations. The dashed lines are fits of $\epsilon ii = AN - B$ to data from 15, 20, 25, and 30-mer results; while dotted lines are estimated error bounds from re-computing the ϵii values using A±STE and B±STE (STE is short for "standard error"). b) Extrapolation of the εii value to long chain lengths for C, 3-MC, 2,6-MC, and 2,3,6-MC at 25℃ (blue) and 50℃ (red).....26

Figure 2.10: a) Schematic of the CG pulling simulation. Two 10-mer homo-oligomers are placed parallel to each other in the simulation box. A pulling force is applied vertically on the center-of-mass of the upper chain to pull the two chains apart. The atomistic pulling simulation is set up in a similar fashion b) Potential of mean force (PMF) diagrams with representative error bars of two 2,6-MC chains from atomistic (solid line) and CG simulations (points).....28

Figure 2.11: Polymer self-diffusivity estimated from experiments (red) and simulations (blue) plotted against number of monomers in the chain. 29

Figure 2.12: a) R_g versus chain length of four representative methylcellulose homopolymers (C, 3-MC, 2,6-MC, and 2,3,6-MC) at room temperature, with Kratky-Porod fits giving persistence lengths of 8nm and 11nm. b) same as a) except R_g values were obtained at elevated temperature. 31

Figure 2.13: Molecular structure of atomistic and coarse-grained HPMCAS. Each HPMACS monomer can be either unsubstituted (H) or substituted with methyl (Me), hydroxypropyl acetyl (HPAc), acetyl (Ac), or succinyl (Su). Each HPMCAS monomer backbone, including any attached methyl groups, is represented by one bead (cyan) centered at the backbone atoms center-of-mass (COM). The HPMACS substitution groups, namely HPAc, Ac, and Su are each modeled by one bead (red) centered at the COM of the functional group atoms. 32

Figure 2.14: Intramolecular atomistic monomer radial distribution functions (RDFs) obtained from atomistic simulations of a 10-mer single chain homogenous HPMCAS model oligomer (blue). The bonded parameters were determined so that the intramolecular CG RDFs (red) reach good agreement with their atomistic counterparts. 34

Figure 2.15: Intermolecular atomistic monomer RDFs obtained from atomistic simulations of 15 chains of 20-mer homogenous HPMCAS model oligomers (blue). The non-bonded parameters were determined by fitting the intermolecular CG RDFs (red) to their atomistic counterparts using a modified Iterative Boltzmann Inversion (IBI) technique.....35

Figure 2.16: Chemical structure of phenytoin (left), map of phenytoin CG beads (middle), and the schematics of the phenytoin CG model (right). Each phenytoin molecule was represented by three CG beads connected with stiff springs. The two phenyl groups were labeled as bead type A (blue), while the middle group containing the amide groups end was labeled as bead type B (yellow). Three bonds are used to connect the three phenytoin beads, namely A-A bond and two A-B bonds. 37

Figure 2.17: Intramolecular atomistic monomer RDFs obtained from atomistic simulation of a single phenytoin molecule (blue). The bonded parameters were determined by matching intramolecular CG RDFs (red) to their atomistic counterparts. 37

Figure 2.18: Intermolecular atomistic oligomer-drug monomer RDFs obtained from atomistic simulations of 15 chains of 20-mer homogenous HPMCAS model oligomers and 150 phenytoin molecules (blue). The explicit cross

interaction parameters were determined by fitting the intermolecular CG oligomer-drug RDFs (red) to their atomistic counterparts using a modified IBI technique. 39

Figure 2.19: Snapshots from atomistic and CG simulations of phenytoin-only systems containing 150 phenytoin molecules in cubic simulation boxes of 12nm on the side (~3.5 wt%), after a simulation time of 30ns and 500ns for atomistic and CG simulations respectively. The CG simulations use the aggregation-based CG force field (middle) and the crystal-based CG force field (right), respectively. 40

Figure 2.20: Intermolecular bead-bead RDFs obtained from atomistic and CG simulations of phenytoin aggregates. The inset shows the intermolecular bead-bead RDF obtained from atomistic simulation of a bulk phenytoin crystal. 41

Figure 3.1: Snap-shots after 35ns for simulations of METHOCEL™ A in water at 25℃.....47

Figure 3.2: The same as Figure 3.1, except at 50℃.....47

Figure 3.3: Close-in views of the METHOCEL™ A "40-mer 4wt%" (left) and "20-mer 6wt%" (right) systems at 50℃ show that model METHOCEL™ A oligomers condense parallel to each other. Each chain is colored differently. 48

Figure 3.4: Peak heights of oligomer-oligomer RDFs (at r≈1.2nm) in 4 METHOCEL™ A model oligomer solutions at room temperature (25℃), shown in blue points, and elevated temperature (50℃), shown in red. 48

Figure 3.5: Contact maps averaged over four model METHOCEL™ A systems ("10-mer 4 wt%", "20-mer 3 wt%", "20-mer 6 wt%", and "40-mer 4 wt%") at 25℃(a) and at 50℃(b). Each tabulated value in contact maps is an averaged value over the actual number of contacts of that type obtained from the four model systems, divided by the number of that type that would exist if the same total number of contacts were assigned randomly, based on the fraction of each monomer type present. The corresponding standard errors of the two contact maps are shown in (c) and (d).....51

Figure 3.6: Simulated chain-length dependence of R_g for methylcellulose model oligomers at two temperatures (25℃ and 50℃), along with experimental data provided by Li (2016) and Patel et al. (2008). The inset shows R_g for

methylcellulose model oligomer with intermolecular interaction value (εii) for each monomer type set to 80% of the value at 50℃, corresponding to 43℃ if we assume εii value scales with temperature linearly. 53

Figure 3.7: R_g versus time for model METHOCEL^{TM} A oligomer. a) At elevated temperature the chain transforms from a stretched state (left inset) to a collapsed ring state (right inset). The scale bar for the left inset is 50nm and 5nm for the right inset. b) R_g at room temperature starting from the ring structure formed at high temperature, showing re-expansion of the polymer.....54

Figure 3.8: a) Snapshot of the stable ring structure formed by heterogamous methylcellulose chain at elevated temperature. b) Three eigenvalues of the radius of gyration tensor.....54

Figure 3.9**.** Snapshots from five simulations of chains with DP = 1000 with varying polymer concentrations. a) At the beginning, three chains are randomly placed in a cubic simulation box with periodic boundary conditions. To vary the concentration, we change the box size L = 400, 180, 80, 60, and 45 nm, corresponding to polymer concentrations of 0.002, 0.017, 0.19, 0.46, and 1.09 wt %. b-f) Final snapshots of the simulations corresponding to various MC concentrations. All simulations were performed at T = 50 ℃.....57

Figure 3.10**.** Initial and final snapshots from simulations of five replicated rings self-assembling into tubular structure. The ring is formed from a single heterogeneous MC chain with DP = 1000. All scale bars are 5nm. Three initial configurations were constructed. a) Replicates are placed on top of each other. b-c) Replicates are placed randomly in the simulation box at concentration of 3.9 and 1.3 weight percent respectively. 59

Figure 3.11. a) Two five-chain proto-tubes are brought together and equilibrated, forming a ten-chain tube (b) Cross-section and dimensions of a three-chain proto-tube. All scale bars in the figure are 5nm.....59

Figure 3.12. Snapshots of (a) initial, (b) metastable, and (c) final structures in the fifteen-chain tube structure growth simulations. All scale bars in the figure are 5nm.....60

Figure 3.13**.** Three averaged eigenvalues (λ_x, λ_y, λ_z, in *nm*) of the radius of gyration tensor as a function of number of chains in the ring and tube structures.....60

Figure 3.14**.** Snapshots of ten-chain tube structure at a) 50℃ and b) 43℃. All scale bars in the figure are 5nm. 61

Figure 3.15. Schematics of torus (a) and bundle (b); (c) Schematic of a generic end fold we consider in this work, which has three exposed monomers (colored in gray). The angles in the end fold, $\vartheta 1-3$, are formed among three consecutive monomer beads. For simplicity, these angles are taken to be 120°.....63

Figure 3.16**.** Phase diagrams for polymer chain with dimensionless chain length $L^* = L/\sigma = 600$. Each symbol on the phase diagram marks the parameter values at which a simulation was conducted. Specifically, open circles, crosses, and asterisk symbols are the simulation runs using stiff, harmonic, and cosine bending potentials respectively. A vertical bar connects simulations data producing the same type of structure. The color of the vertical bars, of the dashed lines, and of the regions of the diagram producing this structure coded as follows: blue—random coil (RC), pink—fluctuating between random coil and torus (RC&T), red—torus (T), green—fluctuating between torus and bundle (T&B), yellow—bundle (B), and black- globule (G). Note the pink and blue regions only occur in simulations because we do not model random coil phase. We regard the simulated T&RC phase simply as a torus phase (T) because we do not model the random coil (RC) phase analytically. The dashed lines connecting the vertical bars mark the upper boundaries of the phases obtained from the simulations (e.g., simulation results between yellow and grey dashed lines resulted in a bundle as the final collapsed structure). Exact solutions for the boundaries between the three collapsed phases are shown on the phase diagram (black dash-dotted lines). We have selected representative conformations in each phase, pointed by the arrows. We have calculated the theoretical aspect ratios for these conformations, depicted in the insets I–V. We have also conducted simulations using harmonic bending potential under the same conditions, and the final snapshots from the simulations are supplied in additional insets I–V with subscript "sim". 70

Figure 4.1: Snapshots from atomistic and CG simulations. The monomer substitution types of the homogenous oligomers and polymers simulated are listed as the row titles. Snapshots from the atomistic simulations are shown in the first column, and the snapshot from the CG simulations of 20-mer, 50-mer, and 100-mers are shown in the second, third, and fourth columns respectively. For the 20-mer systems, 15 chains were randomly placed in the cubic simulation box of 12nm on the side, leading to a concentration of 10wt%. For the 50-mer and 100-mer systems, 70 chains and 35 chains, respectively, were placed in the cubic simulation box of 35nm on the side, producing a concentration of 3wt%.....77

Figure 4.2: Intermolecular CG RDFs obtained from CG simulations of multiple 20-mer, 50-mer, and 100-mer homogenous HPMCAS model oligomers. The RDFs for 2,6-Me (black) are presented on the 10-based log scale on the right.....78

Figure 4.3: Snapshots from atomistic and CG simulations of model polymer (blue)-phenytoin (red) systems. The monomer substitution patterns of the homogenous oligomers and polymers simulated are listed as the row titles.

Snapshots from the atomistic simulations are shown in the first column, and the snapshots from the CG simulations of 20-mers and 50-mers using the aggregation-based CG force field (FF), and 50-mers using the crystal-based CG FF, are shown in the second, third, and fourth columns, respectively. For the 20-mer systems, 15 oligomers (10wt%) and 150 drug molecules (3.5wt%) were randomly placed in the cubic simulation box of 12nm on the side. For the 50-mer systems, 70 polymers (3wt%) and 1500 drug molecules (1.5wt%) were placed in the cubic simulation box of 35nm on the side. 81

Figure 4.4: Intermolecular CG polymer-drug RDFs obtained from CG simulations of 70 50-mer homogenous model polymers and 1500 phenytoin molecules. RDFs of the simulations using aggregate-based CG force field and crystal-based CG force field are shown in blue and red respectively. RDFs of the polymer systems with protonated succniyl group (Su) are shown as solid lines and with deprotonated succniyl group (SuDP) as dashed lines. 81

Figure 4.5: Snapshots from CG simulations of model polymer (blue)-phenytoin (red) systems. In the first row, we show the transformation from phenytoin-2,6-Me-3-Su system to phenytoin-2,6-Me-3-SuDP system, mimicking pH shift from 3 to 7. In the second row, we show the reversed transformation. The polymer and phenytoin concentrations are 3wt% and 1.5wt% respectively. 82

Figure 4.6: Snapshots from CG simulations of phenytoin (red) only and polymer (blue)-phenytoin (red) systems. The heterogeneous 50-mer model polymer chains are constructed using the percentages of each monomer substitution type shown in Table I. The polymer and phenytoin concentrations are 3wt% and 1.5wt% respectively. 84

Figure 4.7: The sizes of the largest phenytoin clusters as a function of simulation time in three systems shown in Figure 4.6. The cluster sizes are tracked using the GROMACS utility g_clustersize. 84

Figure 4.8: Number of contacts between phenytoin molecules as a function of simulation time. The 50-mer systems (left) contain 70 polymer chains (~3 wt%) and 1500 phenytoin molecules (~1.5 wt%) for the 33% drug loading, or 140 polymer chains (~6 wt%) and 1500 phenytoin molecules (~1.5 wt%) for 20% drug loading respectively. The 100-mer systems (right) contain half number of polymer chains comparing to their 50-mer counterparts. 87

Figure 4.9: Snapshot of the simulation systems of 50-mer 2,3-Me-6-HPAc (a), 2,6-Me (b), and 2,3-Me-6-SuDP (c). The drug loadings in all three systems are 20%. 88

Figure 4.10: Same as Figure 4.9, except that polymers are 100-mer long. 88

Figure 4.11: Same as Figure 4.8, except for different polymer chemistries. 90

Figure 4.12: Snapshots of the drug release simulations in rectangular (a) and spherical (b) geometries respectively.91

Figure 4.13: Release profiles of the phenytoin molecules in 50-mer 2,3-Me-6-HPAc polymer-drug complex. The system has 20% drug loading. 91

Figure 4.14: Release profiles of the phenytoin molecules in 50-mer polymer-drug clusters (left) and 100-mer polymer-drug clusters (right). The compositions of the clusters are the same as the ones described in Figure 4.8.....92

Figure 4.15: Same as Figure 4.14, except for different polymer chemistries. 93

Figure 4.16: Schematics of the drug release model. The inner region (region 1) contains the polymer matrix, and the outer region (region 2) is the release media, which is taken to be water in our transport model. 94

Figure 4.17: Fitted mutual diffusivity of phenytoin molecules from polymer-drug clusters with various diameters. The system simulated is 2,3-Me-6-HPAc homopolymers with 20% drug loading.....96

Figure 4.18: Fitted mutual diffusivity of phenytoin molecules from polymer-drug clusters of various polymer chemistries. Two drug loadings are studied, namely 20% and 33%. 97

List of Tables

Table 2.1: Summary of the intramolecular parameters in CG model. Here εii, σii, and rc_ii are the input values to the Lennard-Jones 9-6 potential (Equation 2.6).....17

Table 2.2: Summary of the intermolecular parameters of the LJ 9-6 potential used in the CG model, namely εii, σii, and R_c. The ε_{ii} value is determined by an analytical expression of chain length (N), $\varepsilon_{ii} = AN^{-B}$, where the constants A and B are tabulated. 26

Table 2.3: Dissociation free energy of four representative methylcellulose homopolymers (C, 3-MC, 2,6-MC, and 2,3,6-MC) at both room temperature ($25\,C°$) and elevated temperature ($50\,C°$).....28

Table 2.4: Fitted persistence lengths of four representative methylcellulose monomer substitution types (C, 3-MC, 2,6-MC, and 2,3,6-MC) at both room temperature and elevated temperature. The persistence length is fitted using the Kratky-Porod model. 31

Table 2.5: HPMCAS monomers and co-polymer composition modeled in this work. In the first column, the numbers refer to the substitution positions, shown in Figure 1. The second column lists the number of CG bead types used to model each monomer substitution type. The total mole percentages of each monomer substitution type are shown in the third column, which represent an example of commercial HPMCAS polymer. The protonated and deprotonated 2,6-Me-3-Su substitution types are one single monomer substitution type under different pH values in the commercial polymer product, and therefore share the same mole percentages (shown in parenthesis). We assume random substitution of these monomer substitution types when constructing heterogeneous chains in this work.....33

Table 3.1: Average mole fraction of each methylated cellulose monomer in METHOCEL™ A chemistry.....44

Table 3.2: Numbers of hydrogen bonds, per 10 monomers in the chain, with standard error, and percentages of different hydrogen bond types in the METHOCEL™A model systems. The normalization of the hydrogen bond counts per numbers of 10-mer sub-chains, means that numbers of hydrogen bonds in a 20-mer chain and in a 40-mer chain are obtained from the entries in the table by multiplying by 2 and 4 respectively. 49

Table 4.1: Mutual diffusivities obtained from simulation and release time estimated from transport model. The release time is estimated based on 20% drug loading, 2μm diameter particles and the time required to achieve 99% particles release.....98

Abstract

Solid dispersion formulation is a promising method to maintain *in vivo* drug solubility and to improve drug efficacy. However, the exact drug stabilization and release mechanisms of the solid dispersion formulation are unclear. In this doctoral work, we present a multi-scale modeling approach to study the solvation behavior of cellulosic polymers and their interactions with the model drug phenytoin. We compare a number of atomistic force fields and find they give similar predictions for the stiffness of the cellulose chains. We then develop systematic coarse-grained (CG) force fields for two cellulosic polymers, namely methylcellulose and hydroxylpropyl methylcellulose acetate succinate (HPMCAS), based on the radial distribution functions obtained from atomistic simulations. We use the methylcellulose CG model to simulate the self-assembly of multiple 1000 monomers long polymer chains, and find that they spontaneously form ring or tubular structures with outer diameter of 14nm and void fraction of 26%. These structures appear to be precursors to the methylcellulose fibrils, whose diameter and structure are in good agreement with both theoretical and experimental results, and thus shine light on the methylcellulose gelation mechanism. We also present a simplified continuum analytical model to predict a phase map of the collapse conformations of a single self-attractive semiflexible polymer chain in solution into either folded or ring structures depending on the chains bending energy and self-interaction energy. The predicted phase map is in good qualitative agreement with simulation results for these collapsed structures. We use the HPACAS CG model to study the intermolecular interaction modes between 9 functional groups on HPMCAS and model drug phenytoin. We adopt two criteria to quantify the effectiveness of the polymeric excipients, namely 1) the ability to inhibit drug aggregation and 2) the ability to slow down drug release. We find the size of the functional group is more responsible for the former, while the intermolecular interaction strength is more responsible for the later. Therefore, hydroxypropyl acetyl group, which has both bulky size and strong interaction strength, is the most effective functional group, followed by hydroxypropyl and acetyl group, in good agreement with the results from experimental dissolution tests. In addition, we provide continuum models and predict that the drug release time from a typical solid dispersion particle with 2μm diameter ranges from several seconds to less than 10 minutes depending on the functional group. The systematic coarse-graining approach offer molecular level insights that aid the design of high performance polymeric excipients, and can be extended to cellulosic polymers with novel functional groups and additional drug candidates of interest. Thus, our multi-scale modeling approach is of great interest to the pharmaceutical and material design fields.

Chapter 1: Introduction

1.1 Overview of the Oral Drug Delivery

The global pharmaceutical industry had a combined revenue of over 500 billion U.S Dollars in 2016 and is projected to triple in the next decade[1]. Among the dozens of new drug products approved by the Food and Drug Administration (FDA) every year, two most common administration methods are via oral and via intravenous. Between the two approaches, the oral route is the preferred dosage form for many physicians due to the convenience and compliance by patients. However, there are two major physiological challenges involved during this process. First, the intestinal epithelium can severely limit the permeation of the drug molecules into the blood stream, thus preventing sufficient amount of drugs being delivered. Secondly, the low pH level and enzyme in the stomach can cause the drug molecules to degrade. To address the first challenge, Lipinski suggested that a drug candidate should be lipophilic to allow high permeability through the intestinal epithelium, in addition to four other characteristics. These typical characteristics, known as the Lipinski rule of five,[2] have guided the search of the new drug candidates and have revolutionized the pharmaceutical industry. As a result, more than 50% of drugs candidates, or active pharmaceutical ingredients (APIs), that are currently in the research and development pipeline are estimated to be lipophilic.[3,4] Even though lipophilic molecules have high permeability through the intestinal epithelium, they have low solubility in water and crystallize easily. Because the majority of the gastrointestinal fluid is water, drug crystallization occurs frequently, resulting in poor bioavailability. To maintain the API solubility and slow down or prohibit crystallization, many formulation techniques have been explored, including complexation,[5,6] particle size reduction,[7] and amorphous solid dispersion.[8–11] Amorphous solid dispersion, where API molecules in their amorphous form are mixed with polymer excipients, has become a very promising approach to API solubility enhancement due to the following reasons: 1) many hydrophobic drug molecules are readily soluble in organic solvents used to prepare the formulation and therefore remain dispersed in the polymer matrix after spray drying; 2) no chemical bonds are formed between the polymers and drug molecules; 3) the production of the formulation is reproducible, controllable, and scalable.[9]

1.2 Review of the Polymeric Excipient

The performance of the solid dispersion formulation is characterized by the release profiles, shown in Figure 1.1. There are three typical types of release profiles. When the drugs are released without the presence of the polymers, the free drug concentration in the solution quickly plateaus at the saturation concentration. The presence of the polymers generally aids the drug concentration to reach the super-saturation regime (i.e. concentration that is higher than the saturation concentration of the drug). While an average polymeric excipient allows the drug concentration to gradually decrease to its saturation concentration over time (blue curve), an excellent polymeric excipient can maintain the solubility of the drug molecules at super-saturation regime for up to several hours, shown in the orange curve in the release profile diagram. Maintaining supersaturation facilitates sustained drug delivery and reducing the frequency of drug administration are both highly desired qualifications for an effective pharmaceutical product. Nevertheless, how polymeric excipient works in a solid dispersion formulation is still mysterious. Generally, it is believed that polymeric excipient works through either one or both of the following approaches 1) to reduce the diffusion of the drug in order to delay the form of drug crystals, and 2) to interact with drug with special interaction sites on the polymer chain through intermolecular interactions including hydrophobic interaction and hydrogen bonding, thereby increasing the activation barrier for crystallization.[12,13] In addition, an effective polymeric excipient contains hydrophobic functional groups that stabilize the drug in the stomach, and hydrophilic functional groups that allow drug to release from the small intestine.

A number of polymers have been identified as promising polymeric excipient, including polyvinylpyrrolidone (PVP), polyethylene glycol (PEG), hydroxypropyl methycellulose (HPMC), and Hydroxypropyl methylcellulose acetate succinate (HPMCAS),[14] all of which have been approved by FDA to be used in pharmaceutical products.[15] Among these polymers, HPMCAS has been identified as one of the most effective polymer excipients for solid dispersion formulation.[9,16,17] However, because it is still unclear what is the exact mechanism of how polymeric excipient works, the discovery of an effective polymeric excipient for a new drug candidate is largely trial and error based, thus is very labor and time consuming. To make matters even more complex, many polymeric materials, including HPMCAS, are random co-polymers. The huge design space prohibits the use of a systematic experimental approach to fine tune the polymeric excipient for each individual drug candidate. As a result, the formulation of

the drug candidates can be very challenging, and result in potentially missed opportunities of many highly effective drug candidates. We note that, in many cases, there are synthesis methods to fine tune the composition of these polymers. The missing piece is the lack of clear understanding of the interaction mode between the polymers and drug molecules, particularly the role of each functional group on the polymers. If a clear understanding of the role of the polymer can be achieved, a systematic design rule can be subsequently generalized to improve the performance of polymer excipients for individual drug candidate. Moreover, the systematic understanding of polymer drug interaction can guide the design of new polymeric excipient that outperformances the existing ones.[18,19]

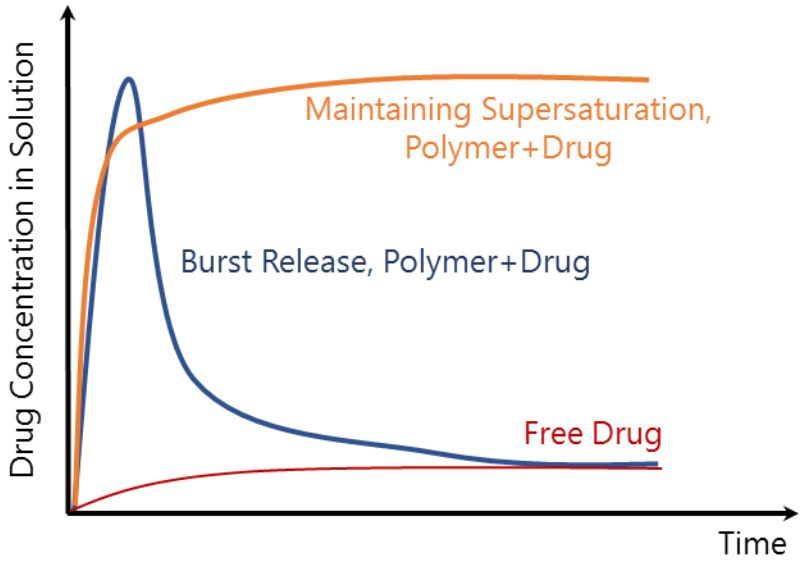


Figure 1.1: Typical dissolution profile of drug only (red), drug with ineffective polymeric excipient (blue), and drug with effective polymer excipient (orange) systems. The typical time scale of the release is in hours, and the typical concentration scale is several hundreds of micrograms per milliliter.

1.3 Review of Cellulosic Polymers

In addition to the applications in pharmaceutical field, various cellulosic polymers have key applications in many fields including food and agriculture[20]. For example, cellulose, being the most abundant organic polymers on earth[21], is largely used in the paper industry. Methylcellulose is a common food additive and a key ingredient in the adhesive material. These cellulosic polymers share the same backbone structure, where D-glucose monomer units are connected by β(1→4) linkages (Figure 1.2). Moreover, the cellulosic polymers offer huge design space and

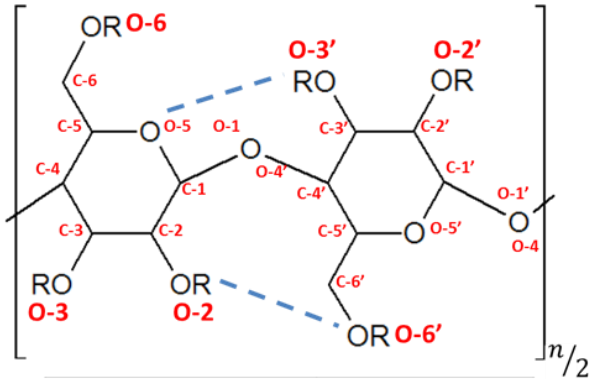
polyfunctionalities. Each monomer unit contains three substitution positions that allow functional groups with various sizes and hydrophobicity levels to be attached. The flexible composition gives the cellulosic polymers unique solvation properties. In the following, we will highlight two of such properties.

Although each AGU has three hydroxyl groups, natural cellulose is insoluble in water because the hydroxyl groups form an extensive intermolecular hydrogen bond network[21]. Methylcellulose with fully substituted hydrophobic functional groups (i.e. methylcellulose with DS=3), on the other hand, are also insoluble in water due to their strong hydrophobic interactions. Interestingly, partially substituted hydrophobically modified methylcellulose with a DS around 2 is water soluble and is therefore of great commercial and scientific value[22]. Although the structural properties of these methylcelluloses in water and other solvents have been extensively characterized over the past few decades[22–26], a complete picture of their solvation behavior has not been revealed. In particular, methylcellulose forms a thermoreversible gel at elevated temperatures. The morphology of the MC gels has recently been identified by Lott *et al.*[27], using cryo-TEM, as a network of fibril structures with a uniform diameter of around 14±2nm above 55℃. However, what is still unclear is the fibril formation mechanism. Early theoretical work[28] hypothesized the MC gel to be bundled structure, but such theory fails to answer why the diameter of a bundled fibril stops increasing beyond 14nm.

HPMCAS contains four major functional groups, namely methyl, hydroxypropyl, acetyl, and succinyl. In general, it is thought that HPMCAS based solid dispersion formulation forms a matrix after being administrated orally. The hydrophobic acetyl group stabilizes the hydrophobic drug molecules in the matrix, while the unsubstituted groups allow hydration of the matrix upon solvation. The succinyl group is pH sensitive. It allows strong interaction with drug molecules at a low pH level of 3 to help stabilizing the drug, and ionizes at pH level of 7 to provide colloidal stability.[9] As a result, the pH sensitive HPMCAS stabilizes the drug, promotes the drug solubility, and allows the drugs to be released at an appropriate rate at the same time, making it one of the best polymeric excipient candidates. In addition, by changing the composition (e.g. actyl/succinyl ratio) of the HPMCAS, a range of solubility and dissolution behavior can be achieved for various drug candidates[29].

If a clear understanding of these solvation properties can be achieved, researchers can

subsequently leverage these properties and design better material for specific applications.



Cellulose	**R**=H
Methylcellulose	**R**=H,CH_3(methyl)
Hydroxylpropyl Methylcellulose Acetate Succinate (HPMCAS)	**R**=H,CH_3,$CH_2CH(OCOCH_3)CH_3$(hydroxypropyl acetyl), $CCOCH_3$(acetyl),$COCH_2CH_2COOH$(succinyl)

Figure 1.2: Schematic of cellulose and methylcellulose, where n is even interger. Anhydroglucose unit (AGU) is the monomer unit of cellulose, where each R group is a hydrogen. In monomers of methylcellulose, one or more R groups are methyl groups (-CH_3). Both AGUs and methylcellulose monomers are connected via a β-(1→4) linkage.The primes indicate atoms on the second (right) monomer in any two consecutive monomers in the cellulosic chain, although the two monomers are bonded identically if they are not the terminal monomers (see text for details). O-2, O-3, and O-6 are the three reactive hydroxyl groups (-OH) on each AGU that can be substituted with different functional groups (i.e. methyl groups) so that the AGU becomes a methylcellulose monomer. The blue dashed lines indicate the two most predominant intramolecular hydrogen bonds that are present in cellulose. The AGUs can be also modified into a HPMCAS monomer by adding methyl, hydroxypropyl, acetyl, and succinyl functional groups.

1.4 Review of Computational Simulation of Polymeric System

Over the past decades, computer simulation techniques have evolved into a powerful research tool to study the molecular systems. In many cases, computer simulations are deployed to help designing new materials for various applications[30]. One of the major advantages of the computational modeling is that the molecular knowledge obtained from computationally affordable simulations allows researchers to gain systematic understanding of the molecular interactions in the system of interest. The simulation results are usually validated against existing data first and then used to provide material property predictions. These predictions, together with the knowledge gained from simulations, help researchers to save experimental effort as well as cost and time. In what follows, we highlight two examples of how computational simulations aid the design of advanced materials.

The ability to predict assembled structures based on the properties of a material's building blocks remains an important goal for material design. In particular, determining the relationship between the shape of the building blocks and final self-assembled shapes is extremely valuable for designing various classes of materials including colloids, nanoparticles, and proteins. A number of computational studies have been conducted to elucidate this relationship[30–32]. For example, Damasceno et al.[32] simulated the assembled structure of over 100 convex polyhedra and demonstrated that these polyhedra self-assemble into four major structural categories, namely liquid crystals, plastic crystals, crystals, and disordered phases. More importantly, a design rule has been generalized for predicting the final assembled structure based on the shape of polyhedra and their local order in the fluid. This knowledge, combined with the breakthroughs in particle synthesis, revolutionize the design of next generation smart materials that are capable of self-assembling into various forms.[33]

Computational simulations have also been used to model commercial polymeric products and offer molecular level insights which are not available through conventional experiments. For example, a multiscale simulation study have been carried to reveal the interactions within the waterborne latex paint.[34–37] Yuan et al.[37] employed atomistic simulations to model the temperature dependent surface energy at the latex binder and water interface. They also employed coarse-grained simulations to extract the free energy of a polymer chain escaping a micellar structure, which was subsequently used to estimate the dynamics within the surfactant molecule and to predict the size distribution of the micelles in the waterborne paint. These simulation studies provide useful tools for researches to optimize a complex commercial product from a different perspective.

HPMCAS based solid dispersion formulation is also a complex commercial polymeric product that lacks fundamental understanding of the molecular level interactions. There is a clear need for a systematic design approach to optimize the composition of HPMCAS in order to achieve the best performance with a drug candidate of interest. This is an excellent opportunity for a computational modeling work. Computational simulations have been used to model many polymeric systems for drug delivery applications. For example, Subashini *et al.*[38]. have used molecular dynamics (MD) simulation technique to study the drug take by polymer and used the information to help design polymer-based drug delivery system Zeng *et al.*[39] have offered a detailed review of multiscale modeling of polymer nanocomposite for various drug delivery

applications. Jha *et al.*[40] simulated HPMCAS model oligomers and their interactions with drug molecules, Xiang *et al.*[41] also used molecular dynamics (MD) to model HPMCAS short oligomer melts, and studied the diffusion of water molecules within this melt. Both atomistic scale works are limited by the size of the system (~10nm) and the simulation time (up to 100ns), and are not readily comparable to any experimental dissolution study. More recently, Srinivas *et al*.[42] developed a solvent-free coarse-grained model for crystalline cellulose. Their simulations show a clear transition between the crystalline cellulose structure and amorphous cellulose structure, which reflects the solvation behavior of cellulose under different solvent environments. In this doctoral work, we want to build on the atomistic model of HPMCAS and the coarse-grained model of cellulose, and develop a systematic coarse-graining approach that can be applied to model the interaction between cellulosic polymers with multiple functional groups and drug candidates of interest.

1.5 Project Overview

In this work, we present a comprehensive multi-scale modeling approach to model the solvation behavior of cellulosic polymers. The multi-scale modeling includes atomistic, coarse-grained, and continuum scales. We first present the molecular dynamics simulation methods in Chapter 2. In particular, we compare the two atomistic force fields that are optimized for cellulosic polymer chains. We then describe the systematic approach to obtain coarse-grained force fields for cellulosic polymers. We highlight two examples, namely the force fields for methylcellulose and HPMCAS. In chapter 3 and 4, we showcase the application of our coarse-grained force field for methylcellulose and HPMCAS. In particular, we focus on the methylcellulose gelation mechanism and the interaction between HPMCAS and model drug molecule phenytoin. Our approach for obtaining the CG force fields is systematic and robust, and can be extended to other cellulose based polymers with various functional groups. In addition, we provide continuum scale modeling to model the collapse behavior or semi-flexible polymer chain and to model the drug release behavior from a polymer matrix.

Chapter 2: Simulation Models of Cellulosic Polymers

Some of the materials in this chapter are results of a collaborative work with Dr. Indranil S. Dalal, Dr. Prateek K. Jha, Rahul Ramesh, and Dr. Taraknath Mandal.

2.1 Introduction

The goal of this thesis work is to develop a computational framework to model the solvation behavior of cellulosic polymers that contain various functional groups and their interactions with small molecules such as drug. Intermolecular interaction between polymer chains, and between polymers and drug molecules, are therefore the key properties to be accurately modeled in our computational framework. Our framework can then be adopted to provide guidance on optimizing the performance of cellulosic polymers for specific purpose. To achieve this goal, the following three major challenges need to be addressed

1) Cellulosic chains adopt beta linkage between the monomers, subsequently resulted in a rather stiff backbone. In addition, monomers adopt different conformations when substituted with different functional groups. Thus, our model needs to capture the persistence length of the cellulosic chains and conformation of monomers to relative accuracy.
2) Each monomer substitution type has different solvation behavior. For example, methylcellulose with fully substituted hydrophobic functional groups (i.e. methylcellulose with DS=3) is insoluble in water due to the strong hydrophobic intermolecular interactions. However, partially substituted methylcellulose (Figure 1.1) with a DS around 2 is water soluble[22]. Therefore, our model should reflect the different solvation behaviors of various cellulosic monomers.
3) Commercial cellulosic polymers are heterogenous and contain hundreds to thousands of monomers. Moreover, many of the events we are interested in (e.g. gelation and drug release) occurs at a time scale of microseconds. As a result, our model needs to be computationally efficient enough to allow simulations of cellulosic chains at large length and time scales, while modeling the unique interactions among different monomer substitution groups.

Although there has been a considerable experimental effort to characterize the solvation behavior of both cellulose and methylcellulose, computer simulations of cellulosic polymers in water have been rare. Until fairly recently, cellobiose has been used as the main model for studying the conformation of cellulose[43–45], while short cellulose oligomers (< 10 monomer units) in solution have also been simulated[46–48]. Simulations up to micro-second scale have been used to study cellulose fibrils and cellulose melts[42,49,50]. However, these studies did not investigate the interaction between multiple cellulosic oligomers in aqueous solution. A number of Coarse-Grained (CG) force fields have been developed for crystalline or amorphous cellulose structures. The MARTINI and the M3B force fields[51,52] both adopted three-site CG models (i.e. wherein each cellulose monomer is represented by three CG beads). Srinivas et al.[42] adopted a one-site CG model and used it to demonstrate the transition of cellulose from a crystalline fibril to an amorphous state. These CG force fields were not suitable for our purpose (i.e. 3 beads are too much).

Here we present a systematic and comprehensive coarse-grained (CG) computational modeling approach to model the cellulosic polymers. We discuss the interactions included in our CG model, and show how to obtain the interaction parameters from atomistic simulations. We first compare the simulation results between three atomistic force fields. We then discuss the CG force field for methylcellulose, and we further extend the CG force field to HPMCAS and drug molecule phenytoin.

2.2 Atomistic Model and Simulation Details

We constructed methylcellulose monomers with all possible combinations of methyl substituents using Materials Studio (version 8.0, Accelrys Inc.). These monomers were then used to build homo-oligomers and random oligomers with user-defined probabilities for incorporation of each monomer. Three different force fields were evaluated, namely GROMOS 56Acarbo[53], which is a more recent force field, the older GROMOS 45A4[54], from which the GROMOS 56Acarbo is derived, and AMBER03 force field[55]. Both GROMOS force fields are specifically optimized for carbohydrates, with the newer one featuring an improved treatment of the ring atoms in the cellulosic repeat unit.

The simulations were carried out in GROMACS (version 4.6.5)[56–58]. The box sizes for the single-chain simulations range from 6 to 22 nm. The simulations of multiple homo-oligomers are

carried out in a 12nm cubic periodic box. Unless otherwise stated, the oligomers were solvated with simple point charge (SPC) water[59] when simulated using GROMOS force fields, and TIP3P water[60] when simulated using AMBER force field. The density of the initial simulation box was approximately equal to the bulk water density, 1g/cc. All systems are subjected to 20,000 steps of energy minimization using a steepest-descent method. A 0.5 ns NVT equilibration followed by a 10ns NPT equilibration was conducted for each system using a time step of 1 fs. A weak temperature coupling using a V-rescale thermostat[61] with relaxation time of 0.2 ps, and a weak pressure coupling using a Berendsen barostat[62] with a relaxation time of 0.5 ps were used for these two equilibration stages, as needed. Production runs were then performed in the NPT ensemble, with temperature coupling using a Nose-Hoover thermostat[63,64] and pressure coupling using a Parrinello-Rahmam barostat[65,66], both with a relaxation time of 0.5 ps. Unless otherwise stated, the temperature was kept at 298K and system pressure at 1 bar. The configuration was constrained by the LINCS algorithm[67], and neighbor lists were updated every 5 time steps in the equilibration runs and every 10 time steps in the production runs. The non-bonded interaction settings were adopted from ref. 25[53]. Specifically, a twin-range cutoff scheme (0.8nm for short range cutoff and 1.4nm for long range cutoff) was used to handle the non-bonded interaction. Beyond 1.4nm, the long-range electrostatics was handled either by the reaction-field method with a dielectric constant set at 61 for SPC water or Particle Mesh Ewald (PME) method for TIP3P water.

2.3 Atomistic Model Validation

We briefly compared the conformational preference of model dimer cellobiose using the three force fields. GROMOS 45A4 has been validated quite extensively in the past based on simulations of various disaccharides and oligosaccharides including cellobiose, amylose fragments, and a selection of methylated cellulose short oligomers[45,46,54,68]. The GROMOS 56Acarbo force field, on the other hand, has not been systematically validated for methylated cellulosic dimers or fragments. AMBER 03 force field is a general purposed force field. Therefore, it is worth validating these force fields before applying them to simulations of longer cellulosic oligomers.

For each of the simulated cellobiose, the averages of the two characteristic dihedral angles that define the β-(1→4) linkage, namely φ (∠O-5–C-1–O-4'–C-4') and ψ (∠C-1–O-4'–C-4'–C-5'), were computed from two 5ns-simulations of one cellobiose molecule using the three force fields.

These values were plotted in Figure 2.1 along with the energy-minimum configuration values derived from X-ray and NMR experiments in the literature. The averaged angles (<φ>, <ψ>) were (-88.9,-129.9), (-60.7,-126.5), and (-90.0,-130.9) for GROMOS 45A4, GROMOS 56Acarbo, and AMBER 03 respectively. For each angle, the average values for the two force fields are comparable to each other were in good agreement with both experimental results and previous simulation results for cellobiose[43,45,46,48,50,68].

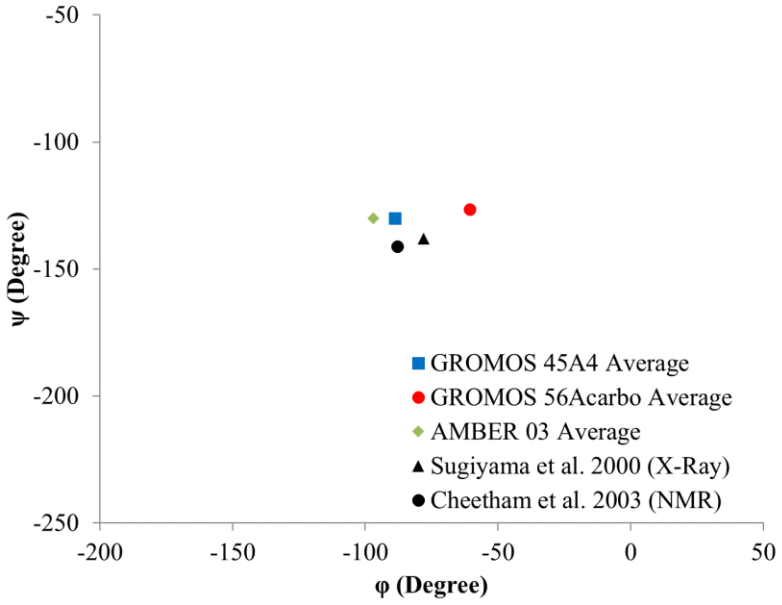


Figure 2.1: Average of dihedral angles φ and ψ from simulations of cellobiose in water using GROMOS 45A4, GROMOS 56Acarbo, and AMBER 03 force fields, along with the two pairs of experimentally determined average values <φ> and <ψ> from the literature. (φ = ∠O-5–C-1–O-4'–C-4' and ψ = ∠C-1–O-4'–C-4'–C-5' where the prime denotes the atom on the neighboring monomer)

The persistence length has been measured experimentally for methylcellulose, but has not been estimated from previous simulations due to limitations on the oligomer chain-length that could be simulated. Figure 2.2 shows our calculations of the radius of gyration R_g as a function of chain length and the fit of the persistence length from the Kratky-Porod worm-like chain model for MC oligomer chains using GROMOS 45A4 force field. Two sets of MC oligomers were simulated, namely 2,6-MC homo-oligomers and random MC co-oligomers. The 2,6-MC homo-oligomers with 10, 20, 35, and 40 monomer units and two random co-oligomers with 18 and 28 monomer units were simulated using GROMOS 45A4. The 2,6-MC homo-oligomer was chosen due to the high abundance of this monomer in the METHOCEL™ A chemistry. The oligomers were relaxed for 10ns before data were gathered from which to compute R_g. The length of each cubic simulation box was set equal to the contour length of the oligomer, which is estimated to

be 0.54n (in units of nm), where n is the number of monomers in the chain, and the factor of 0.54 is reported by Patel *et al.* [69]. The radii of gyration (R_g) were computed using the g_gyrate function in GROMACS and averaged over two 5ns periods. If the averaged R_g values obtained from the two 5ns-intervals differed by more than 10%, the oligomer is simulated for another 5ns to ensure the convergence of the averaged R_g values. The simulated R_g values of all the oligomers using GROMOS 45A4 are plotted against that for a rod-like oligomer in Figure 2.2. R_g value is estimated to be $L/\sqrt{12}$ nm where L is 0.54n nm. The R_g values start to deviate from rod-like behavior beyond 20 monomers length, which roughly corresponds to the persistence length of MC chains.

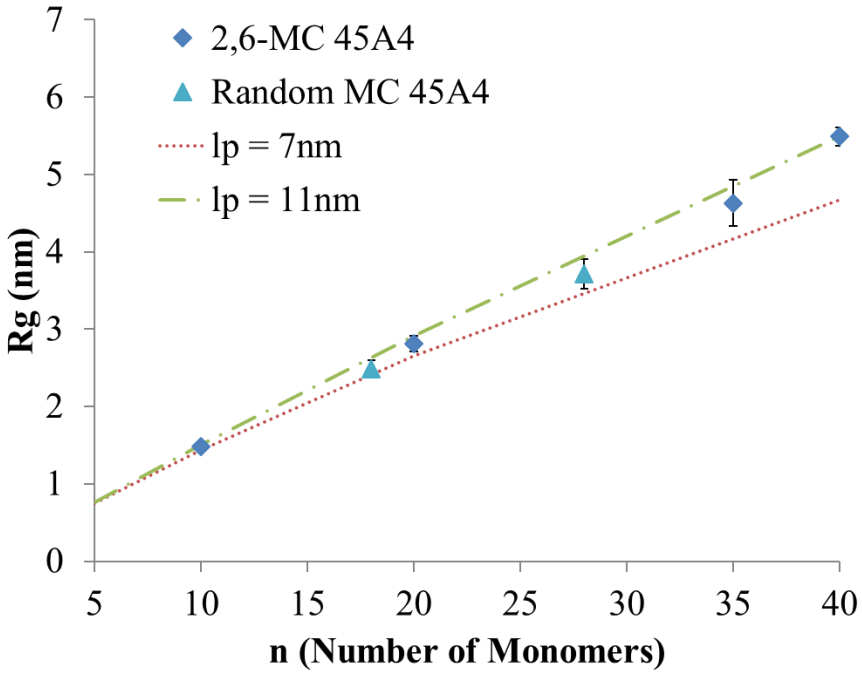


Figure 2.2: Simulated radii of gyration (R_g) of MC chains with the GROMOS 45A4 force field and fits to the Kratky-Porod model, using various persistence lengths l_p.

We further fit the simulated R_g values with the Kratky-Porod model[70] for semiflexible chains, for which:

$$R_g^2 = \frac{1}{3} l_P L - l_p^2 + \frac{2l_p^4}{L^2}\left(e^{-L/l_p} - 1\right) + \frac{2l_p^3}{L} \qquad (2.1)$$

The persistence length (l_p) can be estimated by fitting this model to the simulated R_g values of oligomers with different chain contour lengths (L). The fitting was conducted using the curve-fitting toolbox in MATLAB (MathWorks, R2014b). The persistence length estimated for model

oligomers simulated using the GROMOS 45A4 force field is 11.9 nm, with a 95% confidence interval (CI) of 8.4-15.5 nm. This estimation was in good agreement with the experimental persistence lengths obtained for six different methylcellulose samples by Patel *et al.*, ranging between 12-17 nm with error bars of ± 2 nm for each sample[69]. The estimated persistence length for model oligomers simulated using GROMOS 56Acarbo and AMBER 03 were also in the similar range. We simulated chains that are longer than 40 monomers using Coarse-Grained (CG) force field, which will be discussed in the section 2.6.

2.4 Coarse-Grained Simulation Details

We chose a Brownian Dynamics (BD) simulation technique to perform the CG simulations. The simple BD method used to model the cellulosic polymers, although neglecting long-range hydrodynamic interactions, is a suitable tool for this work because we are primarily interested in the interaction between hydrophobic polymers rather than interaction between polymer and solvent, and the effect of solvent on polymer-polymer interactions can be captured implicitly through the CG effective polymer-polymer interactions. BD simulations were performed using both LAMMPS simulation package[71] (ver. Feb 2014) in an NVE ensemble and GROMACS 4.6.5. In LAMMPS, simulations were set up with dimensionless LJ (Lennard-Jones) units, with the Boltzmann constant (k_b) and the three fundamental units (scales) defined as the mass (m), distance (σ), and energy (ε). The dimensionless values of particle mass, bond length, and temperature are taken equal to unity. The conversion factors between these fundamental dimensionless quantities and their dimensional counterparts have been determined and will be discussed in the next section. Using a Langevin thermostat,[72] the temperature was maintained at 1 dimensionless temperature unit (298K). Note here we kept the dimensionless temperature parameter constant regardless of the "effective temperature" at which we are running the simulation. The "effective temperature" is implicitly captured through the intermolecular interaction parameters parameterized from atomistic simulations conducted at different temperatures. The damping coefficient was set to 10 and a typical time step of 0.001τ, where $\tau = \sqrt{\frac{m\sigma^2}{\epsilon}}$ being the time scale. A typical simulation was carried out for at least 10^7 steps depending on the length of the chain, and the final one third of the data were used to obtain the averaged R_g values for the polymer chain unless the R_g values converged to a specific value earlier than this, which occurred in the case of chain collapse. Because solvent molecules were

implicitly represented, the size of the box was arbitrarily chosen to be 1.5 times the contour length of the longest polymer chain in the box. CG "pulling" simulations were set up with the pulling distance set to be from 1.0 σ to 7.0 σ; the harmonic spring force constant for each histogram was set to 7.0 ε, and a total of 30 simulation windows with 0.2 σ spacing were used to generate PMF. A WHAM calculation code (V2.0.9) developed by Grossfield lab was used in the PMF calculation[73].

The CG simulations performed using GROMACS 4.6.5 employs v-rescale thermostat[61] to maintain the temperature at 298K with the temperature constant (τ_t) set to 0.01ps. The friction force of each bead in the BD simulation was calculated by the mass of each bead divided by the temperature constant, and a typical time step used was 0.025ps, which is comparable to the typical time step used in MARTINI force field[74] often employed to model biological systems. The cutoff distance for Van der Waal's interaction was 2.0nm.

2.5 Coarse-Grained Methylcellulose Model

The CG polymer chains were modeled using beads and stiff springs, with each bead located at the center-of-mass (COM) of a MC monomer, as shown in Figure 2.3. We included both bonded and non-bonded interactions in our bead/stiff-spring model. Because each bead represents one charge-neutral MC monomer, we did not include any explicit electrostatic interaction terms. The complete CG polymer interaction potential was expressed by the following equation

$$U_{CG,polymer} = U_{bond} + U_{angle} + U_{dihedral} + U_{non-bonded} \quad (2.2)$$

The bonded interactions included harmonic bond, angle, and dihedral interactions, which were applied to any two, three, and four consecutive beads on a chain respectively (equation 2.3-2.5).

$$U_{bond} = {}^{1}/_{2}\, K_b\, (l - l_0)^{\,2} \quad (2.3)$$

$$U_{angle} = {}^{1}/_{2}\, K_\theta\, (\theta - \theta_0)^{\,2} \quad (2.4)$$

$$U_{dihedral} = K_\varphi[1 + dcos(n\varphi)] \quad (2.5)$$

Here l_0 and θ_0 are the equilibrium bond length and angle, and K_b and K_θ are the corresponding bond and angle force constants, respectively. In the dihedral expression, d and n are the phase constants, and K_φ is the dihedral force constant. These parameters were determined from a single 10-mer chain atomistic simulation by mapping the intramolecular CG bead-bead radial

distribution function (RDF) onto the corresponding atomistic single chain intramolecular monomer COM-monomer COM RDF, referred to hereafter as the "intramolecular atomistic monomer RDF." A typical fit of the CG RDF to the intramolecular atomistic monomer RDF is shown in Figure 2.4. The equilibrium bond length (l_0) and angle (θ_0) were determined by matching the peak position (r value) of the first (~0.5nm) and the second peak (~1.0 nm) of the intramolecular atomistic RDF. The bond, angle, and dihedral constants (K_b, K_θ, and K_φ) were determined by matching the height (g(r) value) of the first, second, and third peak of the same atomistic RDF. We found that the intramolecular atomistic RDFs are similar among all 10-mer homo-MC chains with different monomer substitution types, which is expected because the contour length of 10-mer chains (~5nm) is well below the persistence length of MC (~11nm)[69] and therefore these stiff chains do not show the effect of substitution. Based on the information from the intramolecular atomistic monomer RDF, we decided to average the RDFs obtained for all eight homo-MC chains and to use a single set of bonded parameters for all MCs. Note that when methylcellulose monomer is substituted at 3-position (i.e. 3-MC), the intra-chain hydrogen bonding network is disrupted and therefore the chain is more flexible[46]. Yet this effect is not predominant in a MC chain that is shorter than 40 monomers long. Therefore, we choose to capture this effect by tuning the non-bonded interaction, rather than bonded interaction. A summary of all bonded parameters is tabulated in Table 2.1.

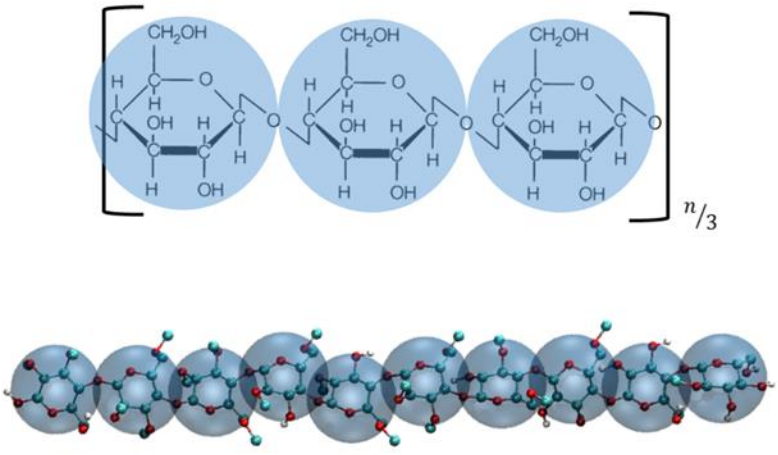


Figure 2.3: Schematics of the methylcellulose coarse grained model. Each methylcellulose monomer (DS ranging from 0 to 3) is represented by one bead centered at the monomer center-of-mass (COM). The beads are connected via hard harmonic springs.

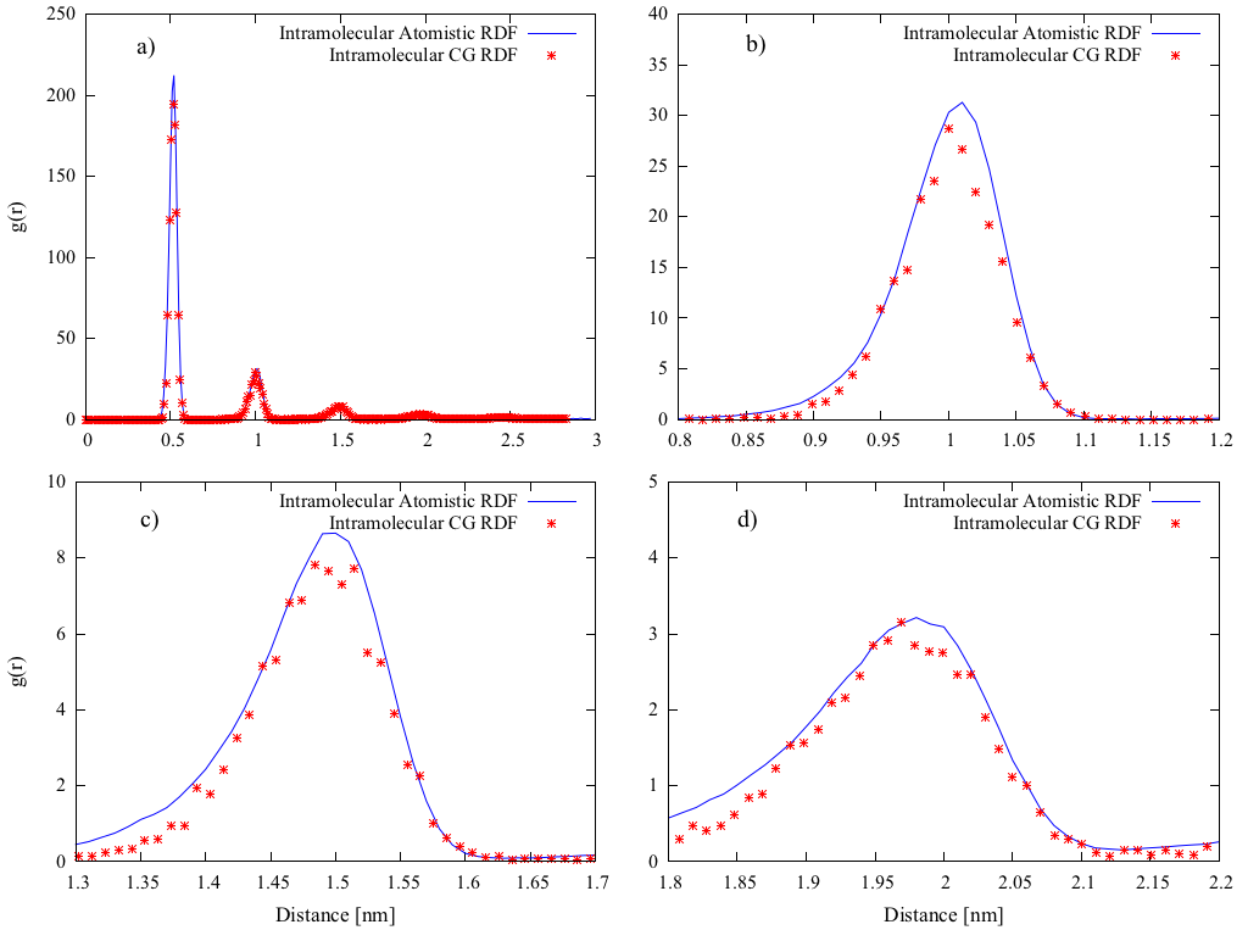


Figure 2.4: a) Intramolecular atomistic monomer RDFs obtained from atomistic simulations of 10-mer single chain homogenous methylcellulose. The RDF is an average of RDFs obtained from eight single homopolymer chain systems. Each homopolymer consists of a chain of 10 monomers of one of the eight methylcellulose monomer substitution types. b), c), and d) are zoomed-in view of the second, third, and fourth peak of those shown in a) at 1.0, 1.5, and 2.0nm respectively.

Bonded Parameters				
	Dimensionless Units		Dimensional Units	
l_0	1	σ	0.515	nm
k_{bond}	1000	ε/σ^2	2478.28	kJ/mol/nm^2
θ_0	165	Deg	2.88	Rad
k_{angle}	30	ε/Rad^2	74.35	kJ/mol/rad^2
n,d	1			
$k_{dihedral}$	2	ε	4.96	kJ/mol
Non-bonded Parameters				
	Dimensionless Units		Dimensional Units	
$\varepsilon_{ii\ long-range}$	0.2	ε	0.5	kJ/mol
$\sigma_{ii\ long-range}$	0.874	σ	0.450	nm
$r_{c\_ii\ long-range}$	4	σ	2.06	nm

Table 2.1: Summary of the intramolecular parameters in CG model. Here ε_{ii}, σ_{ii}, and r_{c_ii} are the input values to the Lennard-Jones 9-6 potential (Equation 2.6).

We selected the conversion factors between the three fundamental dimensionless units (mass m, distance σ, and energy ε) and their dimensional units counterparts. The unit mass was chosen to be 188 Da, which is the averaged molecular weight of all eight MC monomers. The unit length was set to be 0.515nm, which is the simulated average COM separation of monomers from the atomistic RDF. The unit energy was chosen to be 2.478 kJ mol^{-1}, corresponding to $1k_bT$ at 298K. The converted bonded parameters in real units are also tabulated in Table 2.1. All other dimensionless units can be expressed as combinations of these fundamental dimensionless units, as detailed in the LAMMPS manual, and therefore can be also converted to the demensional units.

The non-bonded interactions took the form of a truncated and shifted Lennard-Jones (LJ) 9-6 potential (Equation 2.6). The choice of the LJ 9-6 potential instead of the more commonly adopted LJ 12-6 potential reflects the flat geometry of the methylcellulose monomer. The potential was shifted upward so that the value of the potential function goes to zero at a cutoff distance (r_c). We introduced an additional weak intramolecular non-bonded interaction between any four and five consecutive beads on one chain to fine-tune the CG RDF so that the remaining peak positions (4^{th} and 5^{th} peaks) on the intramolecular atomistic monomer RDFs were matched with CG RDFs. Because these RDFs were similar among all MC monomer substitution types, these weak non-bonded 1-4 and 1-5 interactions were kept the same for all MCs, and are tabulated in Table 2.1.

$$U_{non-bonded}(r) = \begin{cases} \frac{27}{4}\varepsilon_{ii}\left[\left(\frac{\sigma_{ii}}{r}\right)^9 - \left(\frac{\sigma_{ii}}{r}\right)^6 - \left(\frac{\sigma_{ii}}{r_c}\right)^9 + \left(\frac{\sigma_{ii}}{r_c}\right)^6\right] & r < r_c \\ 0 & r \geq r_c \end{cases} \quad (2.6)$$

The intermolecular non-bonded interaction parameters were fitted from intermolecular monomer COM-monomer COM RDFs, which are referred as “intermolecular atomistic monomer RDFs” in the following. These were generated from atomistic simulations with 10 wt% polymer loading at various chain lengths. These intermolecular RDFs were generated in a similar way as the intramolecular RDFs, except that the contributions from the five neighboring monomers of any given monomer were omitted. A typical intermolecular atomistic monomer RDF had several closely-spaced peaks at short distance (see Figure 2.5). Achieving a good fit of the entire RDF of this kind would require a tabulated potential, as demonstrated by Srinivas et al.[42] We, however,

decided to use an analytical form, namely the LJ 9-6 potential, even though this choice allowed us to achieve good fits of only the first main peak, which occurs roughly at r=0.6nm in all intermolecular atomistic monomer RDFs. Our choice was justified by the following reasons:

1) The first peak of the intermolecular atomistic monomer RDF reflects the equilibrium COM distance between two monomers on different chains, which correlates with the intermolecular interaction strength between these two monomers, and is captured by the 9-6 LJ potential in our CG force field.

2) Our goal is to simulate MC chains with realistic chain lengths (>400 monomers), which are well beyond the longest chain length we can afford to simulate at an atomistic scale (30 monomers). As a result, obtaining tabulated potentials for such long chains would be extremely challenging due to the lack of reference atomistic simulations. Our strategy is to use an analytical expression to obtain fit parameters for short chains, and extrapolate these to longer chains. This can only be achieved through the use of an analytical potential function.

3) Commercial MC products contain all eight MC monomer substitution types. Using an analytical potential function for each monomer substitution type is particularly convenient for simulating heterogeneous MC chains, where a geometric mixing rule is used to calculate the ε_{ij} and σ_{ij} between different pairs (Equation 2.7). If, however, tabulated potentials were to be employed, the 24 cross terms between all eight MC monomer substitution combination types would need to be computed at both short and long chain lengths, which would require a tremendous modeling effort.

$$\sigma_{ij} = \sqrt{\sigma_{ii} * \sigma_{jj}}$$

$$\varepsilon_{ij} = \sqrt{\varepsilon_{ii} * \varepsilon_{jj}} \qquad (2.7)$$

4) Our use of analytical potentials in a CG model can be adapted to model other related and important experimental polymers, including hydroxypropyl methylcellulose (HPMC) and hydroxypropyl methylcellulose acetate succinate (HPMCAS), as we shown in section 2.7. These polymers are often used as drug carriers and the sizes of the polymer-drug complexes are usually around a few hundred nanometers.

We used a modified iterative Boltzmann inversion (IBI) scheme to obtain the intermolecular non-bonded interaction parameters. The standard IBI scheme (Equation 2.8) has been used in many studies to achieve very good fitting between CG and atomistic RDFs[42,75–78]:

$$V_{i+1}(r) = V_i(r) + \alpha k_B T \ln\left(\frac{g_i(r)}{g(r)}\right) \tag{2.8}$$

In this equation $V_i(r)$ and $V_{i+1}(r)$ are the tabulated potentials in the ith and $(i+1)$th iteration, respectively; $g_i(r)$ and $g(r)$ are the RDFs corresponding to the potential at the ith iteration and target RDF, respectively. α is the damping factor which is arbitrarily chosen and decreases after each iteration. We modified the standard IBI scheme and used an analytical potential (Equation 2.9-2.10) to substitute for the tabulated potential form of the ith iteration.

$$U_i(r) = \varepsilon_i f(r) \tag{2.9}$$

where

$$f(r) = \begin{cases} 4\left[\left(\frac{\sigma_{ii}}{r}\right)^9 - \left(\frac{\sigma_{ii}}{r}\right)^6 - \left(\frac{\sigma_{ii}}{r_c}\right)^9 + \left(\frac{\sigma_{ii}}{r_c}\right)^6\right] & r < r_c \\ 0 & r \geq r_c \end{cases} \tag{2.10}$$

If we used $U_i(r) = \varepsilon_i f(r)$ in the ith iteration, the objective was to find ε_{i+1} that provides the best fitting $g_{i+1}(r)$ for the $g(r)$, such that the difference between $U_{i+1}(r) = \varepsilon_{i+1} f(r)$ and $U_i(r) = \varepsilon_i f(r)$ approaches zero. In practice, we minimized the function

$$\begin{aligned} G &= \int w(r)\left[U_{i+1}(r) - U_i(r) - \alpha k_B T \ln\left(\frac{g_i(r)}{g(r)}\right)\right]^2 dr \\ &= \int w(r)\left[(\epsilon_{i+1} - \epsilon_i) f(r) - \alpha k_B T \ln\left(\frac{g_i(r)}{g(r)}\right)\right]^2 dr \end{aligned} \tag{2.11}$$

with respect to ϵ_{i+1}. Here, $w(r)$ is a weighting factor that determines the relative importance of the fit at different r values. We used a Gaussian function centered at the location of the first peak, with a standard deviation (s_d) of 0.05σ as the weight function, to give the most weight to the first peak of the RDF (r_p):

$$w(r) = \exp\left(-\frac{(r-r_p)^2}{2*{s_d}^2}\right) \tag{2.12}$$

Taking a derivative of G with respect to ε_{i+1} yields the following expression for updating the ε value at each iteration

$$\varepsilon_{i+1} = \varepsilon_i + \alpha k_B T \frac{\int w(r) f(r) \ln\left(\frac{g_i(r)}{g(r)}\right) dr}{\int w(r) f^2(r) dr} \tag{2.13}$$

Because the RDF is a discrete function, we used Simpson's rule to handle the integral in Equation 2.13. α is set to unity for the first iteration and it is decreased by a factor of 0.8 after each iteration until the ε_i value is converged (with tolerance of 10^{-6}).

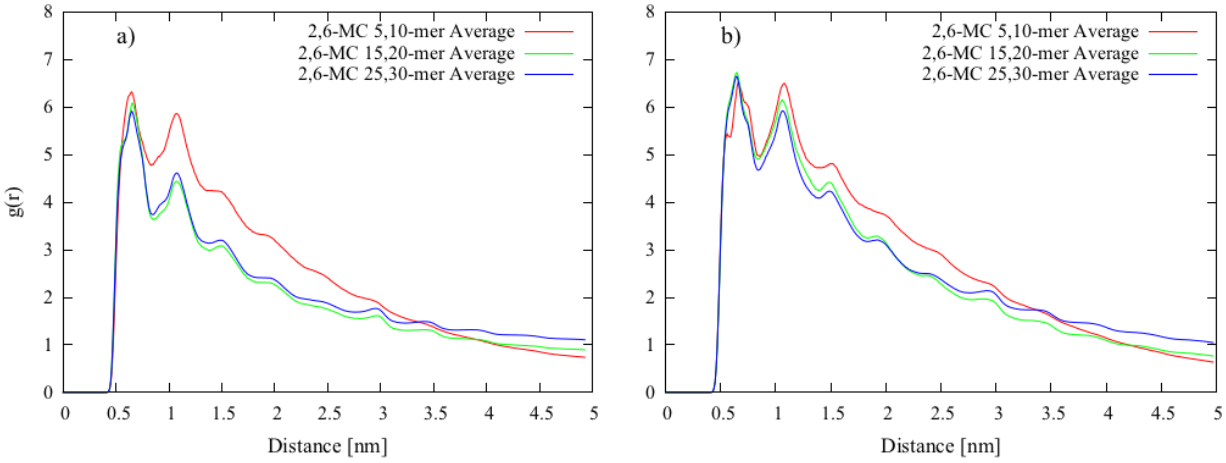


Figure 2.5: a) Intermolecular atomistic monomer RDFs for 10 wt% 2,6-MC with different chain lengths at 25℃. To reduce statistical noise, we averaged RDFs of 5 and 10-mers (red), 15 and 20-mers (green), and 25 and 30-mers (blue). b) Same as a) except these RDFs were generated at 50℃

We parameterized our CG force field at two different temperatures, 25℃ and 50℃. 50℃ is the typical gelation temperature for a dilute MC solution (<2 wt%)[24]. For each temperature, we set up reference atomistic simulations for all eight MC monomer substitution types at three different chain lengths, namely a 15-mer, a 20-mer, and a 25-mer. In addition, we set up reference simulations of 30-mers of Cellulose (C), 3-MC, 2,6-MC, and 2,3,6-MC, and reference simulations of 5-mers and 10-mers for 2,6-MC. The polymer concentration for all reference simulations are 10 wt%. To ensure proper convergence of the atomistic RDFs, each atomistic simulation was repeated twice with different initial configurations, and for up to 40 ns during production runs. The two RDFs from two individual runs converge well with each other. As noted in section 2.3, atomistic RDFs do not change significantly beyond 25ns after starting the production runs. Although one may argue the atomistic simulation cannot properly produce the long-range structure of meta-stable aggregates that form under some conditions, we only match the first peak of the atomistic RDF, which reflects the short-range structure which can be more readily equilibrated in our atomistic simulations. Therefore, we believe that we can derive the intermolecular interaction potentials needed for our CG model from these intermolecular

atomistic monomer RDFs. We found that the intermolecular atomistic monomer RDFs become relatively insensitive to the molecular weight beyond a 15-mer (Figure 2.5) at both temperatures.

We plot the fitted intermolecular CG RDFs against their atomistic counterparts for 20-mer C, 3-MC, 2,6-MC, and 2,3,6-MC at both temperatures in Figure 2.6. We assigned three fitting parameters for each system, namely the LJ parameters ε_{ii} and σ_{ii}, and the cutoff distance r_c. Specifically, the σ_{ii} value was determined based on the relative position between the zero cross position σ_{ii} (r at which U(r) = 0 and r < r_c in Equation 2.6) and the position of the minimum potential r_p (corresponding to the first peak of the RDF) of the LJ 9-6 potential, $r_p = 1.1447\sigma_{ii}$. The cutoff distance r_c was chosen arbitrarily based on the following relationship, $r_{c\_ii} = r_p/0.7$. The potential was truncated and shifted at a distance shortly beyond the position of first main peak in the intermolecular atomistic monomer RDF to ensure that the first peak of the RDF is fitted properly. We have tested this by choosing different cutoff distances, including $r_p(\sigma)/0.7$, 2σ, and 3σ, lowering the magnitude of ε_{ii} to maintain the fit to the first atomistic RDF peak as the cutoff distance increases. We observed that the first CG RDF peak was wider than the atomistic RDF peak when a long cutoff distance (e.g. 3σ) was used, and therefore decided to use $r_p(\sigma)/0.7$ as our cutoff distance in the CG model (Figure 2.7). The ε_{ii} value was fitted iteratively based on the scheme described in the previous section. For 15-mer systems, the intermolecular atomistic monomer RDFs had higher peak heights at elevated temperature than at room temperature, which agrees with the experimental observation that MC chains tend to aggregate and gel at elevated temperature (Figure 2.6). Cellulose, however, had very similar intermolecular RDFs for both temperatures. This was possibly due to the combined effect of weakened hydrogen bonding network and stronger hydrophobic interactions at elevated temperature. We were able to reproduce the position and height of the first peak in the intermolecular atomistic monomer RDFs using the three CG parameters described above. The second peak position of all CG RDFs occurred at double the distance of those of the first peaks, which is typical for a LJ fluid system[79]. We demonstrated, through snapshots (Figure 2.8), that the solvation behavior of different MC monomer substitution types in atomistic simulations can be reproduced fairly well in CG simulations by just fitting the position and the height of the first peak of intermolecular atomistic RDF. Di- and tri-substituted MC formed dense aggregates, while un- and mono-substituted MC formed loosely aggregated structures. Note that the short rigid CG 2,3,6-MC

chains bound side-to-side to form rod-like fibril structures; we do not expect this to occur for chains longer than their persistence length, for which we expect more complex fibril structures.

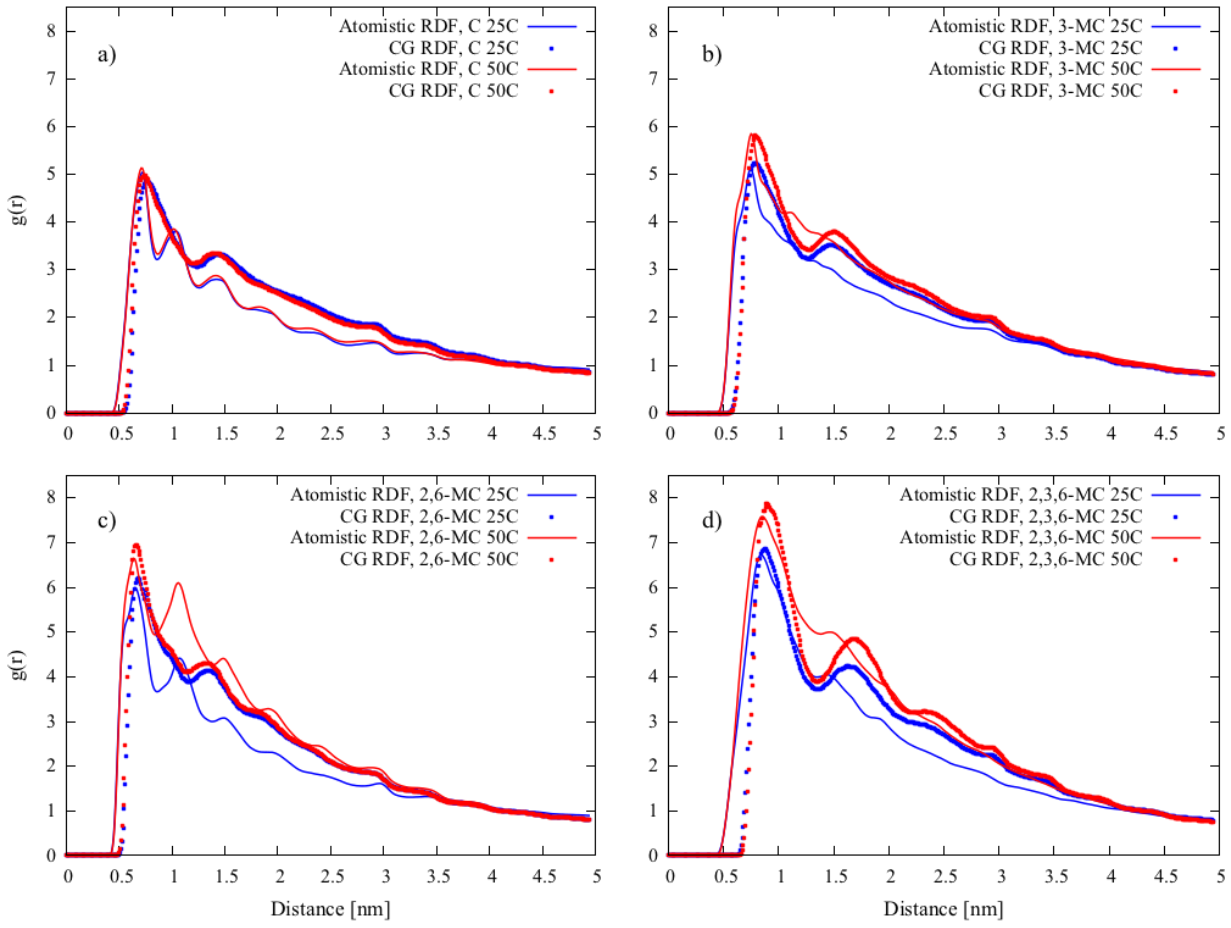


Figure 2.6: Fits of the intermolecular CG RDFs to intermolecular atomistic monomer RDFs for four homo-methylcellulose systems. The atomistic systems are cellulose, 3-MC, 2,6-MC, and 2,3,6-MC. The atomistic RDFs were obtained by computing the intermolecular monomer COM-monomer COM RDF for each 10wt% 15-mer homo-oligomer system. The CG RDFs were computed by the same method were fit to match the position (r_p) and height ($g(r_p)$) of the first peak in atomistic RDFs.

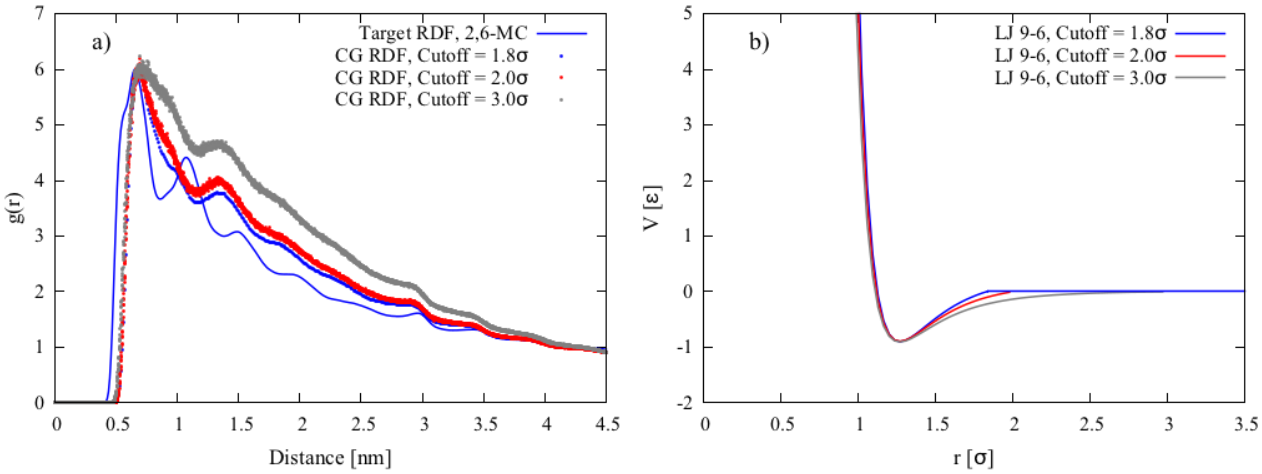


Figure 2.7: a) Intermolecular monomer RDF fitting with different cutoff settings applied to the b) truncated and shifted analytical LJ 9-6 potential. We choose Cutoff = 1.8 to be the cutoff distance throughout this study by setting $r_{c\_ii} = r_p/0.7$. The σ_{ii} value is kept constant in all cases, and ε_{ii} is adjusted to maintain the fitting of first peak and the well depth in the potential.

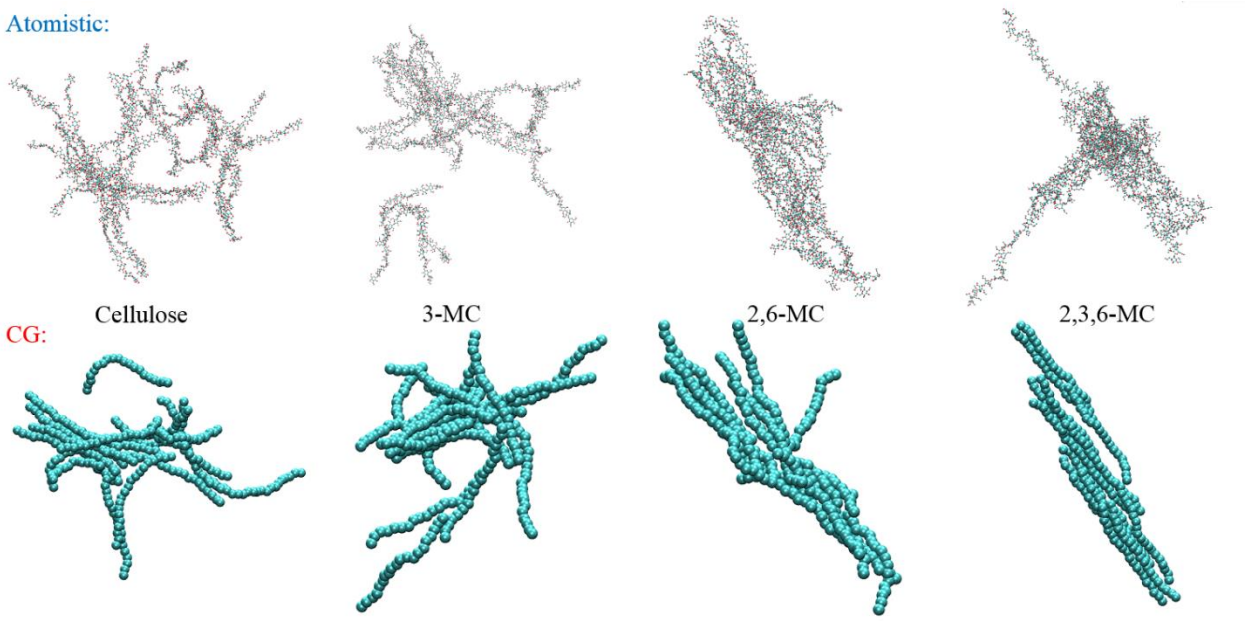


Figure 2.8: Snapshots of aggregation at 50 ℃ of CG chains and their atomistic counterparts. All systems contained 10wt% 20-mer homo-oligomers. The snapshots were obtained after simulation times of 40ns (atomistic) and $6x10^6$ steps (coarse grain).

We compared the fitted CG parameters for each system at different chain lengths. In 2,6-MC systems, the positions and heights of the first peak in the intermolecular atomistic RDFs for 5-mer and 10-mer were slightly different from those for 15-mers and longer chains. We conclude that the RDFs for 5-mers and 10-mers do not converge well and from here on we focus on 15-mer or longer chains. Among the RDFs for one monomer substitution type at one temperature, the positions of the first peak are similar for different chain lengths. This means that we can use a

single value of σ_{ii} and of r_{c_ii} for each monomer substitution of type i at that temperature, irrespective of chain length. We next focus on obtaining ε_{ii} values. As chain length increases, atomistic chains form more loosely packed structures. As a result, we observed a decrease in the fitted CG ε_{ii} values from 15-mers to 30-mers so that the longer CG chains also formed more loosely packed structures. We hypothesize that the ε_{ii} values will eventually plateau at large chain length, but we cannot afford long enough atomistic chains to reach the plateau. Therefore, we propose an analytical function to fit the chain-length dependence of the parameters obtained from atomistic simulations, and use this to extrapolate to obtain ε_{ii} values for long CG chains. We use a power-law fitting ($\epsilon_{ii} = AN^{-B}$) of the ε_{ii} values as a function of chain length (N), which allows the ε_{ii} values to decrease for long chain lengths. The functional form used for fitting would result in ε_{ii} decreasing asymptotically to zero at infinite chain length, but we limit use of this formula to chains no longer than 1000-mers so that the ε_{ii} values for long chains are no smaller than about half the values for 25-mers.

The extrapolation of the ε_{ii} values from chains no longer than 25 or 30 monomers to 1000 monomers is a long, and therefore dangerous, extrapolation. However, in what follows, we show that the predictions of the coarse-grained model using these extrapolated values yield remarkably good agreement with experimental results for both persistence lengths and chain collapse transitions that are associated with gelation. Thus, regardless of the uncertainty of our methods of obtaining them, it appears that we end up with coarse-grained parameters that are descriptive of the experimental system. We leave it to future work to justify these parameters or supply ones that are more convincingly derived than is possible here.

We tested the sensitivity of our fitting strategy on four representative MC homopolymers, namely C, 3-MC, 2,6-MC, and 2,3,6-MC. For each of these homopolymers, we obtained ε_{ii} values from 15-, 20-, 25-, and 30-mer chains. We then carried out 2-point, 3-point, and 4-point fittings (15,20-mer, 15,20,25-mer, and 15,20,25,30-mer respectively) on these ε_{ii} values using the power-law equation mentioned in the previous paragraph ($\epsilon_{ii} = AN^{-B}$). We obtained one pair of A and B fitting parameters for each of the 2-point, 3-point, and 4-point fittings. We then calculated the standard error (STE) for both A and B. We used A and B from 4-point fitting as our reference parameters and estimated an upper and lower bound of the extrapolated ε_{ii} values by plugging in A±STE and B±STE into the power law equation and re-compute the upper and

lower bounds for ε_{ii} values at different chain lengths. In Figure 2.9a, we show the ε_{ii} values obtained from fits to the atomistic simulations and the extrapolated curves with estimated error bounds for 2,6-MC system at two temperatures. The error bounds for all four systems are less than 10% for both temperatures, suggesting that the fittings obtained through 4-point and 3-point fittings would give extrapolated ε_{ii} values are accurate to within 10% at each chain length. Therefore, we used 4 point-fitting for the four representative MC systems we discuss in this section, and 3-point fitting for the rest of the MC systems in this study due to the expense of simulating 30-mer systems. For 2,6-MC, the extrapolated ε_{ii} values for a 1000-mer are 44% and 81% of those of a 20-mer for room and elevated temperature, respectively. For a 2,6-MC 20-mer, there is a 15% change in ε_{ii} between 25℃ and 50℃, but for a 1000-mer, this change grows to over 50%. This shows that the effect of the temperature on the interaction strength is more significant at longer chain lengths. Similar trends have been observed in the other three representative homopolymers as well. In Figure 2.9b, we magnify the long-chain region and plot the extrapolated ε_{ii} values for the four representative homopolymers at chain lengths between 250 and 1000 monomers. In general, σ_{ii} values and cutoff distances (r_{c_ii}) are very similar between the two temperatures for the same monomer substitution type (see Table 2.2). Values of ε_{ii}, therefore, roughly determine the strength of the intermolecular interaction for any specific monomer substitution type. We observe that for all homopolymers, ε_{ii} values were higher at elevated temperatures (red solid symbols) than at room temperature (blue hollow symbols). This is expected since a stronger interaction strength is mostly likely responsible for gel formation at elevated temperature. Note here that 2,3,6-MC has similar ε_{ii} values at both temperatures, and at elevated temperature ε_{ii} values for 2,3,6-MC is lower than those for 2,6-MC. This can be attributed to the fact that there is no hydroxyl group in 2,3,6-MC and it cannot form hydrogen bonds. Thus, the increment in ε_{ii} values with temperature for 2,3,6-MC is smaller than that for 2,6-MC, which can form hydrogen bonds. We tabulate the intermolecular interaction parameters we used for each monomer substitution type at both temperatures in Table 2.2.

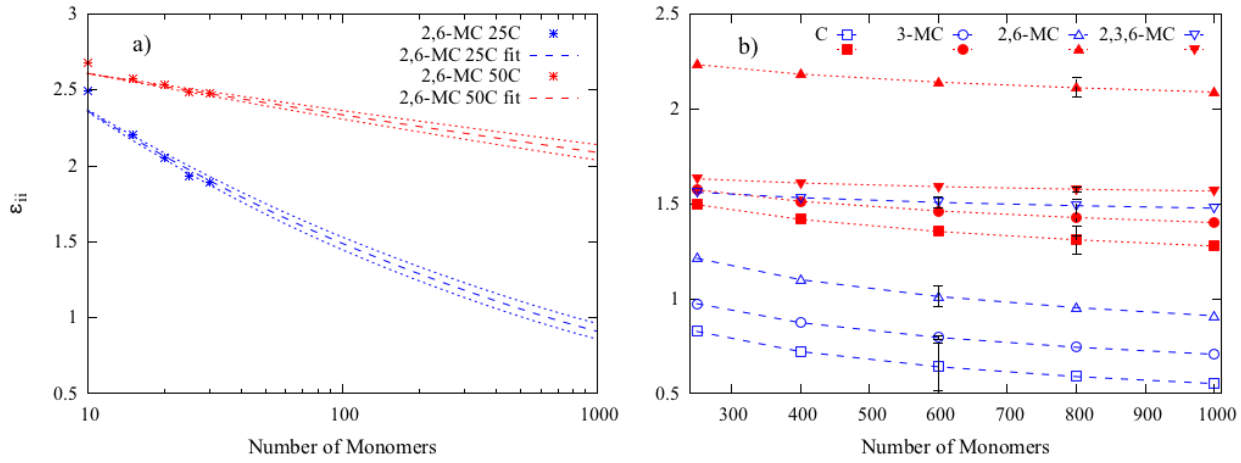


Figure 2.9: a) Extrapolation of the ε_{ii} values to long chain lengths. Blue and red stars are fitted values obtained from corresponding atomistic simulations. The dashed lines are fits of $\epsilon_{ii} = AN^{-B}$ to data from 15, 20, 25, and 30-mer results; while dotted lines are estimated error bounds from re-computing the ϵ_{ii} values using A±STE and B±STE (STE is short for "standard error"). b) Extrapolation of the ε_{ii} value to long chain lengths for C, 3-MC, 2,6-MC, and 2,3,6-MC at 25℃ (blue) and 50℃ (red)

Low Temperature (25℃)				
	A	B	σ_{ii}	r_{c_ii}
Cellulose (C)	4.1324±0.1880	0.2913±0.0198	1.2383	2.0250
2-MC	3.6765±0.0009	0.1801±0.0001	1.0687	1.7476
3-MC	3.4880±0.0103	0.2310±0.0014	1.2722	2.0804
6-MC	5.2975±0.8227	0.2969±0.0609	1.0517	1.7198
2,3-MC	3.9416±0.0887	0.2456±0.0099	1.3062	2.1359
2,6-MC	3.8093±0.0425	0.2072±0.0055	1.1196	1.8308
3,6-MC	2.9452±0.0878	0.2055±0.0143	1.3401	2.1914
2,3,6-MC	1.9479±0.0102	0.0400±0.0022	1.4079	2.3024
High Temperature (50℃)				
	A	B	σ_{ii}	r_c
Cellulose (C)	2.8070±0.0349	0.1139±0.0056	1.1988	1.9695
2-MC	2.4410±0.0207	0.0222±0.0036	1.0687	1.7476
3-MC	2.5033±0.0381	0.0840±0.0068	1.2892	2.1082
6-MC	2.4760±0.0181	0.0140±0.0031	1.1144	1.8308
2,3-MC	2.4819±0.0224	0.0362±0.0039	1.3001	2.1359
2,6-MC	2.9161±0.0167	0.0483±0.0025	1.0975	1.8031
3,6-MC	2.2424±0.0318	0.0489±0.0061	1.3170	2.1637
2,3,6-MC	1.9172±0.0039	0.0292±0.0008	1.4521	2.3856

Table 2.2: Summary of the intermolecular parameters of the LJ 9-6 potential used in the CG model, namely ε_{ii}, σ_{ii}, and R_c. The ε_{ii} value is determined by an analytical expression of chain length (N), $\varepsilon_{ii} = AN^{-B}$, where the constants A and B are tabulated.

2.6 Coarse-Grained Model Validation

Dissociation Free Energy of 10-mer Homo-Oligomers

To validate the CG force field we have developed, we set up "pulling" simulations using both atomistic and CG force fields to estimate the dissociation free energy between different pairs of side-to-side bound 10-mer homo-oligomers. In these simulations, two parallel homo-oligomers were pulled apart and the dissociation free energy calculated using umbrella sampling, as illustrated in Figure 2.10a. In the CG simulation, the two parallel chains were simulated for 50,000 steps and one configuration was selected every 15,000 steps (a total of three configurations) from the run as starting configurations for the pulling simulations to test the effect of initial configuration on the final free energy value. Performing a "pulling" simulation using atomistic MD, however, is computationally demanding. Therefore, we estimated the standard error of the simulated dissociation free energy via block averaging our data over three 10 ns intervals (10-20 ns, 20-30ns, 30-40ns respectively) for each window.

Figure 2.10b shows the potential of mean force (PMF) curves for 2,6-MC at room temperature, from both atomistic and CG simulations. The estimated ΔG values are $10.8 \pm 0.7\ k_bT$ and $9.7 \pm 0.3\ k_bT$ for atomistic and CG simulations respectively. Note that the shapes of the PMFs at short distances (<1.5nm) do not agree well, likely due to the lack of explicit solvent molecules in our CG simulation. Water molecules form cages around hydrophobic molecules such as 2,6-MC[80], which need to be disrupted, with consequent free energy penalty, to pull two hydrophobic chains apart. The atomistic PMF curve plateaus at ≈1.2 nm, indicating that the chains have already reached bulk solution conditions and no longer feel each other's presence at this separation. For the CG PMF, the plateau is reached at a longer distance (≈2.2 nm) possibly as a result of the absence of the explicit solvent molecules, but the final plateau is similar to that of the atomistic simulations, indicating that the solvent entropic contributions are captured implicitly through the fitting of the CG RDFs to the atomistic ones.

In addition to 2,6-MC, we also estimated the dissociation free energy for C, 3-MC, and 2,3,6-MC systems at both room and elevated temperatures using CG pulling simulations, tabulated in Table 2.3. These results show that the dissociation free energies increase with degree of substitution and with increased temperature, as expected.

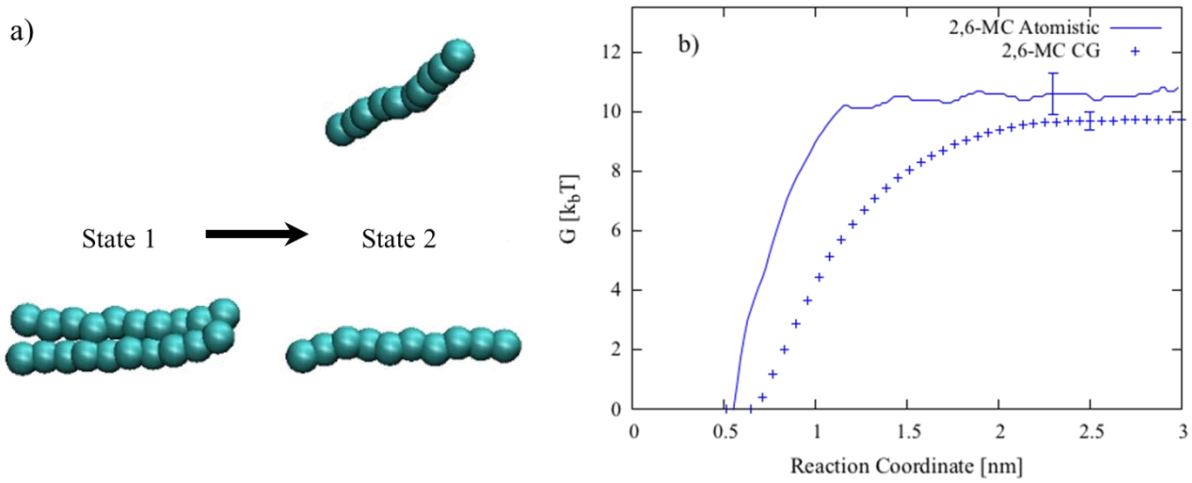


Figure 2.10: a) Schematic of the CG pulling simulation. Two 10-mer homo-oligomers are placed parallel to each other in the simulation box. A pulling force is applied vertically on the center-of-mass of the upper chain to pull the two chains apart. The atomistic pulling simulation is set up in a similar fashion b) Potential of mean force (PMF) diagrams with representative error bars of two 2,6-MC chains from atomistic (solid line) and CG simulations (points).

	ΔG [k_bT], 25℃		ΔG [k_bT], 50℃
C	7.4±1.0		8.6±0.6
3-MC	8.0±0.9		9.1±0.4
2,6-MC	9.7±0.3		10.2±0.2
2,3,6-MC	12.3±1.1		14.1±0.3

Table 2.3: Dissociation free energy of four representative methylcellulose homopolymers (C, 3-MC, 2,6-MC, and 2,3,6-MC) at both room temperature (25℃) and elevated temperature (50℃)

Scaling of Time in Coarse-Grained Simulations

We now relate the CG simulation time scale to the experimental time scale by matching the self-diffusivity of CG chains with the experimental diffusivity of MC chains. From experiments, the diffusion coefficient of the polymer chains can be estimated from their hydrodynamic radius (R_h) using the Stokes law (Equation 2.14) where η_s is the solvent viscosity (1cP for water). We estimate the ratio of hydrodynamic radius (R_h) over radius of gyration (R_g) for MC to be 1.4 from the data published by Keary[22]. Courtesy of Dr. Li from the Dow Chemical Company, we obtained the R_g values for various METHOCEL™ A samples and estimated the diffusivity of these samples from the above ratio.

$$D_t = \frac{k_b T}{6\pi\eta_s R_h} \tag{2.14}$$

From simulations, the self-diffusion coefficient of the CG chain can be computed using the polymer center-of-mass (COM) mean-square displacement (MSD) (Equation 2.15) where r is the COM position of the polymer chain.

$$D_t = \frac{\langle [r_{COM}(t) - r_{COM}(0)]^2 \rangle}{6t} \tag{2.15}$$

We set up three runs for each single-chain system with chain length ranging from 10 to 400 using the intermolecular interaction parameter derived for cellulose. The simulated self-diffusion coefficient has units of σ^2/τ, while in experiments the units are m^2/s. The ratio of these two units gives the conversion factor between CG time and real time. We pick the conversion factor, $\tau = 0.028ns$, so that the simulated diffusion coefficient for a 250-mer chain agrees with the experimentally measured diffusion coefficient for a chain of similar length (see Figure 2.11). The typical step size in our CG simulation was 0.001τ, which converts to 28fs per CG time step. This step size is comparable to a typical step size used in simulations with the MARTINI force field, which ranges from 20-40 fs[74]. We plot the diffusion coefficients from both simulation and experiment in Figure 2.11. The diffusion coefficients (D) measured in the experiments follow a $D \propto N^{-0.59}$ scaling law, N being the number of monomers, as predicted by Zimm theory for the chain-length dependent diffusivity for dilute polymer chains in a good solvent[81]. The diffusion coefficients calculated from simulations, however, follow a $D \propto N^{-1}$ scaling law, as predicted by the Rouse theory[82]. This discrepancy, which has also been observed by Chen et al.,[77] results from the lack of hydrodynamic interaction in our BD simulations.

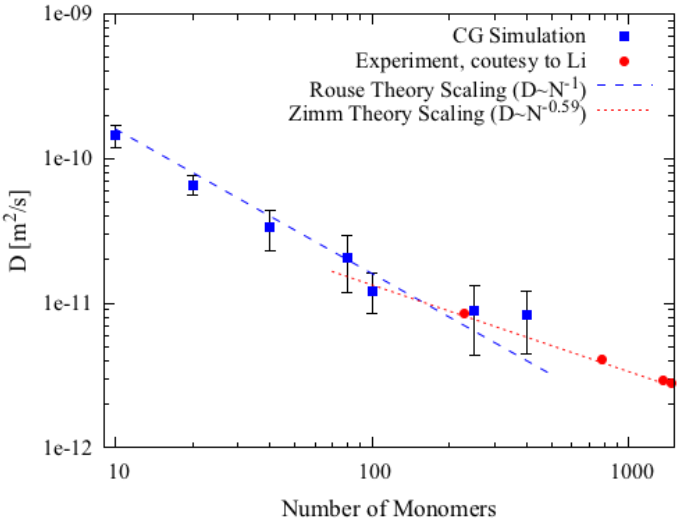


Figure 2.11: Polymer self-diffusivity estimated from experiments (red) and simulations (blue) plotted against number of monomers in the chain.

Chain-Length Dependence of Homogenous MC Radius of Gyration

We applied the CG parameters to simulate the four representative MC systems chosen earlier (Cellulose (C), 3-MC, 2,6-MC, and 2,3,6-MC), at various chain lengths at both room and elevated temperatures. We built homopolymer chains of each of these monomers with lengths ranging from 10 to 1000 monomers, and correspondingly varying contour lengths L. We simulated single chains of these and obtained the equilibrium radius of gyration (R_g),versus L, which we fitted with the Kratky-Porod Model[70] for semi-flexible chains (Equation 2.1).

The fittings of persistence length were again conducted using the curve-fitting toolbox in MATLAB (MathWorks, R2014b). The R_g values for different homopolymers are plotted in Figure 2.12. In Figure 2.12a, we show the dependence of R_g on chain length for the four homopolymers at room temperature. We observe that at low chain lengths (less than 100 monomers), R_g values are independent of monomer type. The R_g values in this regime follow rod-like scaling, as expected for semi-flexible polymers. Between 100 to 400 monomers, a transition occurs to the long-chain flexible regime (with more than 400 monomers), where a new power-law is obtained for three of the four monomer substitution types, namely C, 3-MC, and 2,6-MC. In the flexible regime, R_g becomes dependent on monomer substitution type. C and 3-MC are relatively stiff and have persistence lengths above 12nm, while 2,6-MC, due to its stronger self-interaction, has a persistence length of around 7.5nm. On the other hand, 2,3,6-MC shows a collapse transition between 400 and 600 monomers, which agrees with the observation that homogenous 2,3,6-MC is insoluble in water at room temperature.

In Figure 2.12b, we show R_g versus chain length for the same homopolymers at elevated temperature. C and 3-MC polymers have lower persistence lengths (8.5nm) at $50\,^{\circ}$C than at room temperature, reflecting the stronger intermolecular interactions at elevated temperature. Notice that as chain length increases, 2,6-MC also shows a collapse transition between 400 and 600 monomers. In addition, 2,3,6-MC has a collapse transition at shorter chain length (around 250 monomers) than at room temperature (400 monomers). These strong interactions for high DS substitutions are very likely the driving force for aggregation and gelation of MC chain at elevated temperature, as previous simulation and experimental works suggested. We summarize the persistence lengths estimated for these monomer substitution types in Table 2.4.

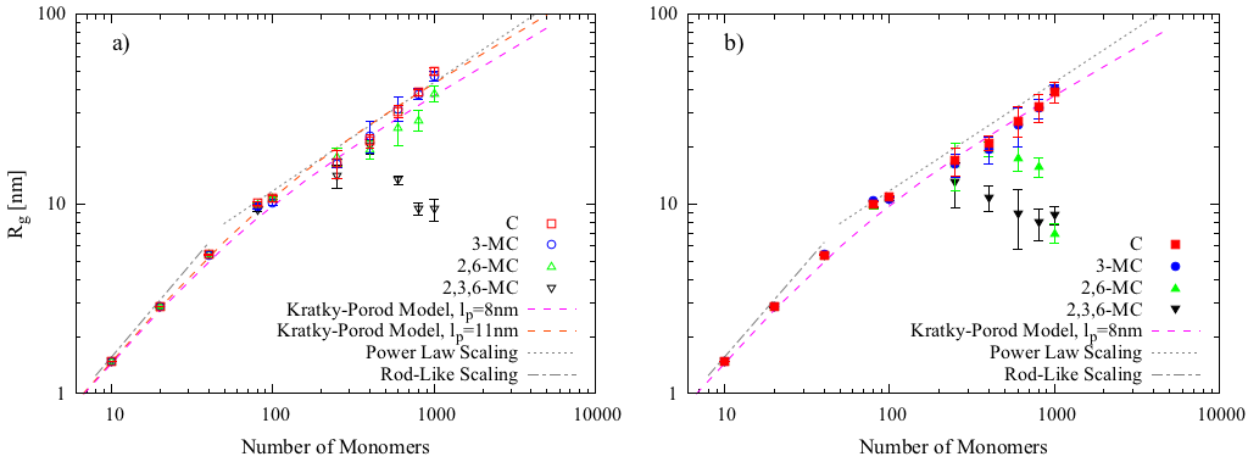


Figure 2.12: a) R_g versus chain length of four representative methylcellulose homopolymers (C, 3-MC, 2,6-MC, and 2,3,6-MC) at room temperature, with Kratky-Porod fits giving persistence lengths of 8nm and 11nm. b) same as a) except R_g values were obtained at elevated temperature.

	Persistence Length [nm], 25℃		Persistence Length [nm], 50℃
C	12.9 ±0.8		8.3 ±1.2
3-MC	12.1 ±1.0		8.5 ±0.9
2,6-MC	7.3 ±0.9		Collapse
2,3,6-MC	Collapse		Collapse

Table 2.4: Fitted persistence lengths of four representative methylcellulose monomer substitution types (C, 3-MC, 2,6-MC, and 2,3,6-MC) at both room temperature and elevated temperature. The persistence length is fitted using the Kratky-Porod model.

2.7 Coarse-Grained Hydroxypropyl Methylcellulose Acetate Succinate (HPMCAS) Model

We now turn our attention to the polymer HPMCAS. The CG HPMCAS chains were modeled using beads and stiff springs, similar to that described in the previous section, shown in Figure 2.13. Here we will outline a few key points and differences. The CG force field is implemented in GROMACS. This is chosen for the following two reasons 1) The force field is compatible with the phenytoin force field recently developed by Mandal et al[83], which was also implemented in GROMACS; and 2) GROMACS has more efficient algorithm when computing intermolecular interaction comparing to LAMMPS, thus offer higher computational performance[84]. A total of 10 HPMCAS monomer substitution types were modeled in this work, listed in Table 2.5. We used one bead to represent each backbone ring including the methyl groups attached, due to methyl group's relatively small sizes compared to those of other functional groups. The bead was centered at the backbone atoms center-of-mass (COM). We used a separate bead to represent the atoms in each HPMCAS functional group, namely in hydroxypropyl acetyl (HPAc), acetyl (Ac), protonated succinyl (Su), and deprotonated succinyl (SuDP), where the bead is centered at the

COM of the atoms in each respective functional group. This resulted in 1-3 beads per monomer for all 10 monomer substitution types that we have modeled. Note we did not include any monomer substitution type that has all three positions substituted with HPMCAS functional groups (i.e. 4 beads/monomer). We included both bonded and non-bonded interactions in our bead/stiff-spring model are the same as those in methylcellulose model (Equation 2.2).

Similar to the CG force field for MC, the bonded interaction parameters were determined from a single 10-mer chain atomistic simulation by mapping the intramolecular atomistic monomer RDF. We show the fitting for five monomer substitution types in Figure 2.13. Note that for each monomer shown here, with the exception of 2,6-Me, there are three intramolecular atomistic RDFs. This is because each monomer contains two CG beads, namely the backbone bead and functional group bead, and therefore there are two RDFs between the same bead type and one RDF between the two different bead types. The bonded parameters were determined by matching each of the three CG RDFs to their atomistic counterparts.

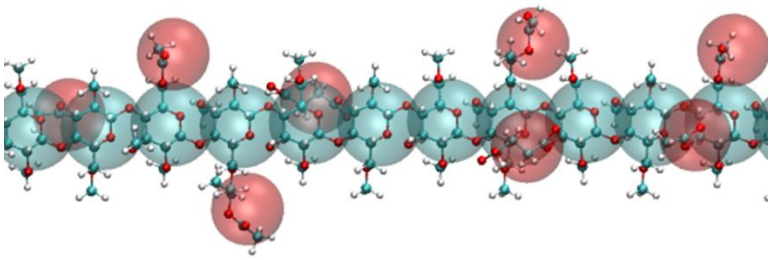


Figure 2.13: Molecular structure of atomistic and coarse-grained HPMCAS. Each HPMACS monomer can be either unsubstituted (H) or substituted with methyl (Me), hydroxypropyl acetyl (HPAc), acetyl (Ac), or succinyl (Su). Each HPMCAS monomer backbone, including any attached methyl groups, is represented by one bead (cyan) centered at the backbone atoms center-of-mass (COM). The HPMACS substitution groups, namely HPAc, Ac, and Su are each modeled by one bead (red) centered at the COM of the functional group atoms.

Monomer Substitution Type	# of CG Beads Types	Mol%
2,3-Me-6-HPAc	2	10
2,6-Me-3-Ac	2	8.5
2,6-Me-3-Su	2	8.5
2,6-Me-3-SuDP	2	(8.5)
2,6-Me	1	15.5
2-HPAc-3,6-Me	2	10
2-Me-6-Ac	2	10
2-Me-3,6-Ac	3	5.5
2,3-Me-6-Ac	2	7
2,3,6-Me	1	25

Table 2.5: HPMCAS monomers and co-polymer composition modeled in this work. In the first column, the numbers refer to the substitution positions, shown in Figure 1. The second column lists the number of CG bead types used to model each monomer substitution type. The total mole percentages of each monomer substitution type are shown in the third column, which represent an example of commercial HPMCAS polymer. The protonated and deprotonated 2,6-Me-3-Su substitution types are one single monomer substitution type under different pH values in the commercial polymer product, and therefore share the same mole percentages (shown in parenthesis). We assume random substitution of these monomer substitution types when constructing heterogeneous chains in this work.

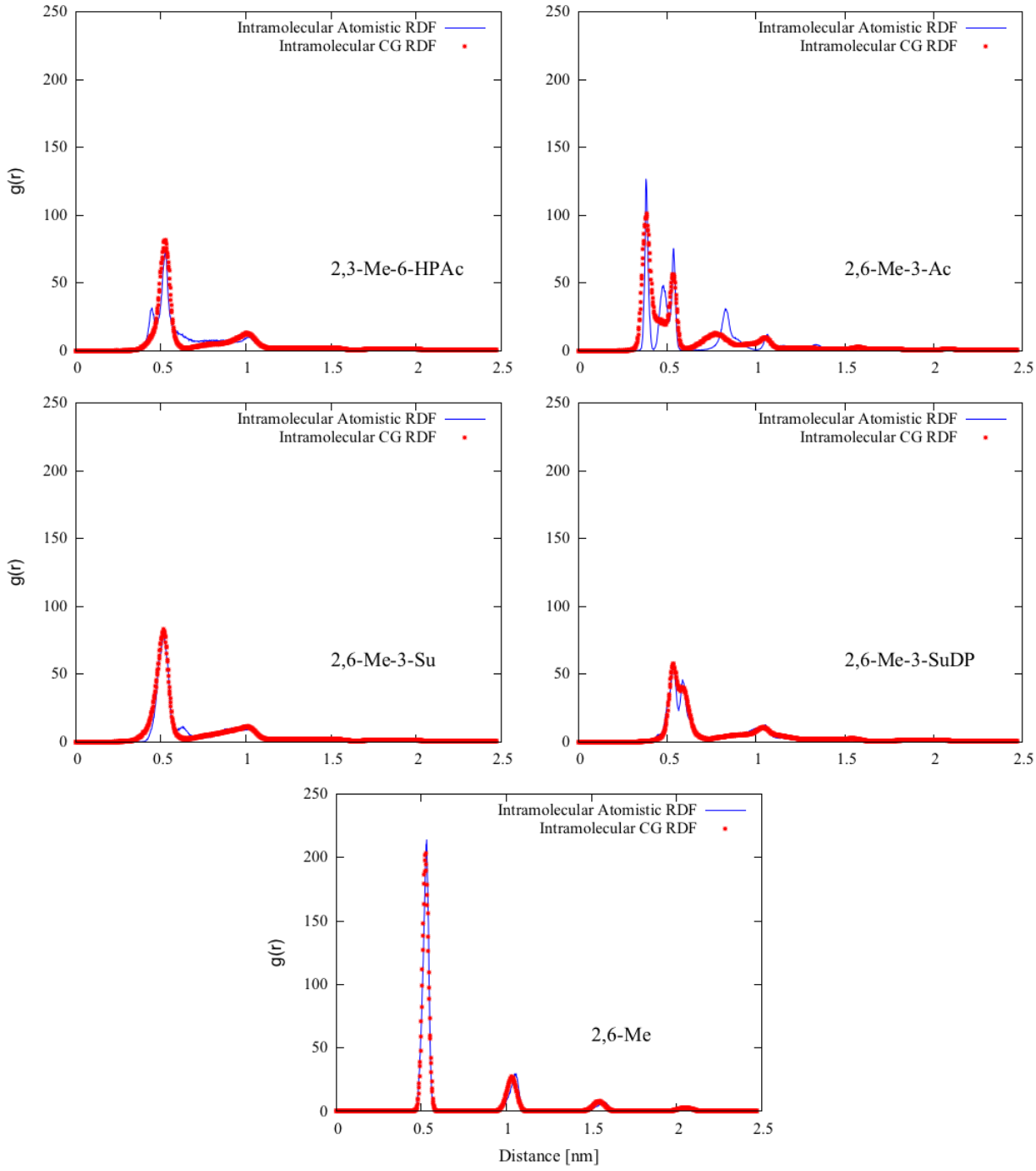


Figure 2.14: Intramolecular atomistic monomer radial distribution functions (RDFs) obtained from atomistic simulations of a 10-mer single chain homogenous HPMCAS model oligomer (blue). The bonded parameters were determined so that the intramolecular CG RDFs (red) reach good agreement with their atomistic counterparts.

The intermolecular non-bonded interactions were modeled with a shifted and truncated Lennard-Jones (LJ) 12-6 potential (Equation 2.21 with cutoff distance (r_c) set to be 2.0nm. The LJ 12-6 potential is chosen instead of the LJ 9-6 is because this is the only analytical Lennard-Jones potential that is supported by GROMCAS natively. The parameters were determined by fitting the intermolecular atomistic monomer RDF. These intermolecular atomistic monomer RDFs were generated from atomistic simulations of 15 homogenous chains, each 20 monomers long, in a cubic simulation box with 12nm on the side.

$$U_{non-bonded}(r) = \begin{cases} 4\varepsilon_{ii}\left[\left(\frac{\sigma_{ii}}{r}\right)^{12} - \left(\frac{\sigma_{ii}}{r}\right)^{6} - \left(\frac{\sigma_{ii}}{r_c}\right)^{12} + \left(\frac{\sigma_{ii}}{r_c}\right)^{6}\right] & r < r_c \\ 0 & r \geq r_c \end{cases} \quad (2.16)$$

We show the fittings of the five substitution types in Figure 2.15. As with the intramolecular RDFs, with the exception of 2,6-Me, for each monomer, there are three RDFs. An intermolecular atomistic monomer RDF has multiple discrete peaks at short distances, and instead of fitting the entire RDF, which would require a tabulated potential, we choose to only fit the first peak with an analytical LJ potential. We have discussed in previous section, compared to the simulations using tabulated potentials, using an analytical LJ potential not only allows the simulation to run faster and allows the cross interactions among different bead types to be handled easily, but also captures sufficient intermolecular interaction information to reproduce the polymer aggregation behaviors seen in the experimental systems. We adopted the modified iterative Boltzmann inversion (IBI) scheme described in section 2.5 to achieve fittings between the CG and atomistic intermolecular RDFs. We note that the backbone beads, even though they share the same methyl-group linkage (e.g., the same backbone atoms are present in 2,6-Me and in 2,6-Me-3-Ac), require different non-bonded interaction parameters, because of the influence of the side groups on the RDFs of the backbone beads. The same behavior holds for the functional group beads (e.g., the functional-group beads in 2-Me-3,6-Ac and in 2,3-Me-6-Ac). Therefore, we assigned unique bead types to all backbone beads and functional beads for all 10 monomer substitution types that we modeled, resulting in a total of 19 CG bead types in our polymer force field (Table 2.5).

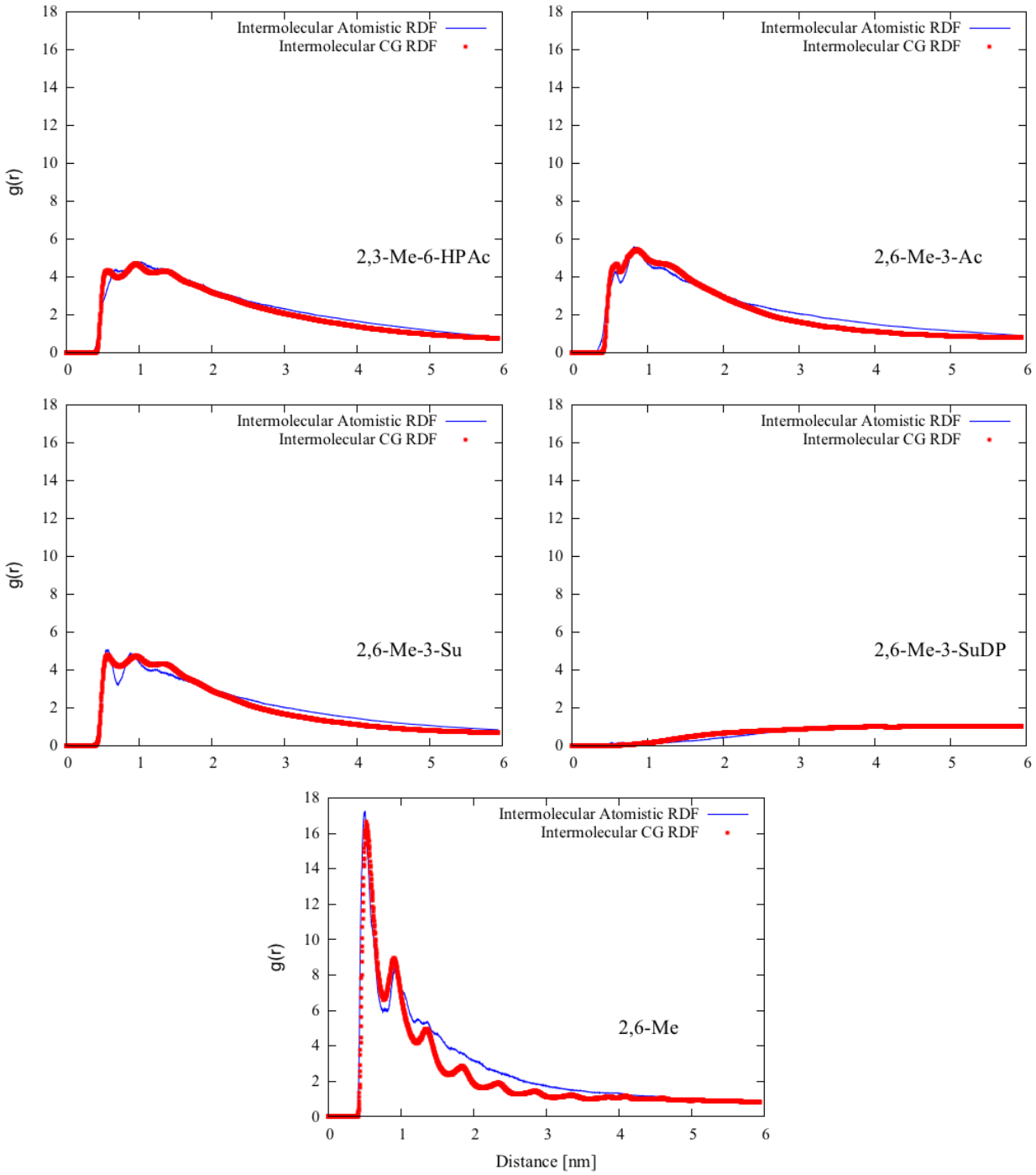


Figure 2.15: Intermolecular atomistic monomer RDFs obtained from atomistic simulations of 15 chains of 20-mer homogenous HPMCAS model oligomers (blue). The non-bonded parameters were determined by fitting the intermolecular CG RDFs (red) to their atomistic counterparts using a modified Iterative Boltzmann Inversion (IBI) technique.

2.8 Coarse-Grained Phenytoin Model

In this work, we use phenytoin molecule as the model drug. Phenytoin has been extensively studied in both experimental and simulation work.[18,19,40] In particular, Mandal *et al.*[83] developed a coarse-grained phenytoin force field that is capable of simulating phenytoin crystal growth. The non-bonded interactions in their force field were modeled using tabulated potentials. However, the use of a tabulated potential is computationally too costly for simulations of phenytoin molecules mixed with heterogeneous HPMCAS polymer chains due to the large number of tables need to be constructed and looked up. Here, we present a CG phenytoin model with the same bead mapping scheme but a different intermolecular interaction parameterization approach. Specifically, we modeled the intermolecular interactions with analytical LJ potentials, in which each phenytoin molecule was represented by three CG beads and three stiff springs. The two identical phenyl group rings were labeled as bead type A, while the middle ring containing the amide groups was labeled as bead type B. Similar to the polymer model, the phenytoin model includes both bonded and non-bonded interactions (Equation 2.17). Because three beads are used to model each phenytoin molecule, it is sufficient to include only a harmonic bond potential (Equation 2.3) in the bonded interaction (i.e. A-A and A-B, shown in Figure 2.16). The bonding potential parameters were again determined by mapping the intramolecular phenytoin CG bead-bead RDF onto the corresponding atomistic intramolecular phenytoin ring COM-ring COM RDF, referred to hereafter as the "phenytoin intramolecular atomistic ring RDF" (Figure 2.17). The non-bonded interactions were modeled using truncated and shifted LJ 12-6 potentials, with a cutoff distance of 2.0nm. Atomistic simulations of 120 phenytoin aggregates in a cubic simulation box of 10nm on the side were conducted, and the non-bonded interaction parameters were obtained by mapping the intermolecular phenytoin CG bead-bead RDF onto the corresponding atomistic intermolecular phenytoin ring COM-ring COM RDF, using the same modified IBI scheme described in the polymer CG force field section. The bonded and non-bonded parameters for the phenytoin CG force field are tabulated in the supplemental information.

$$U_{CG,phenytoin} = U_{bonded} + U_{non-bonded} \tag{2.17}$$

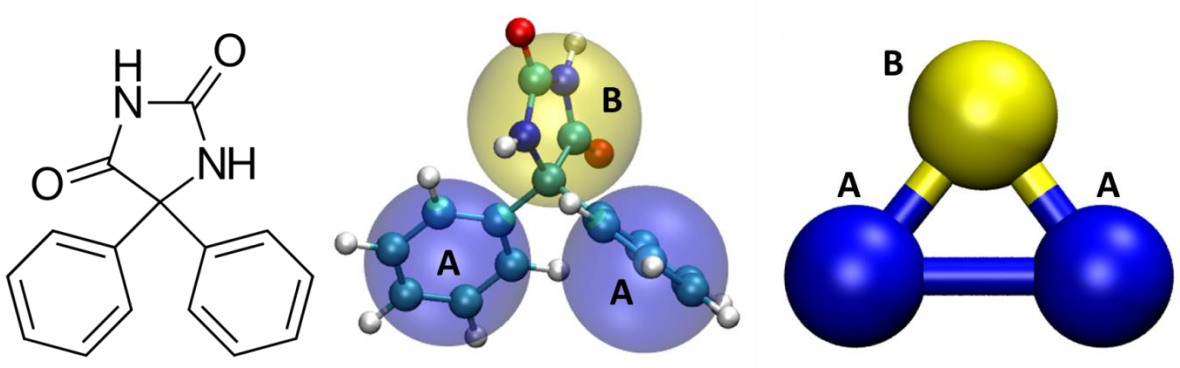


Figure 2.16: Chemical structure of phenytoin (left), map of phenytoin CG beads (middle), and the schematics of the phenytoin CG model (right). Each phenytoin molecule was represented by three CG beads connected with stiff springs. The two phenyl groups were labeled as bead type A (blue), while the middle group containing the amide groups end was labeled as bead type B (yellow). Three bonds are used to connect the three phenytoin beads, namely A-A bond and two A-B bonds.

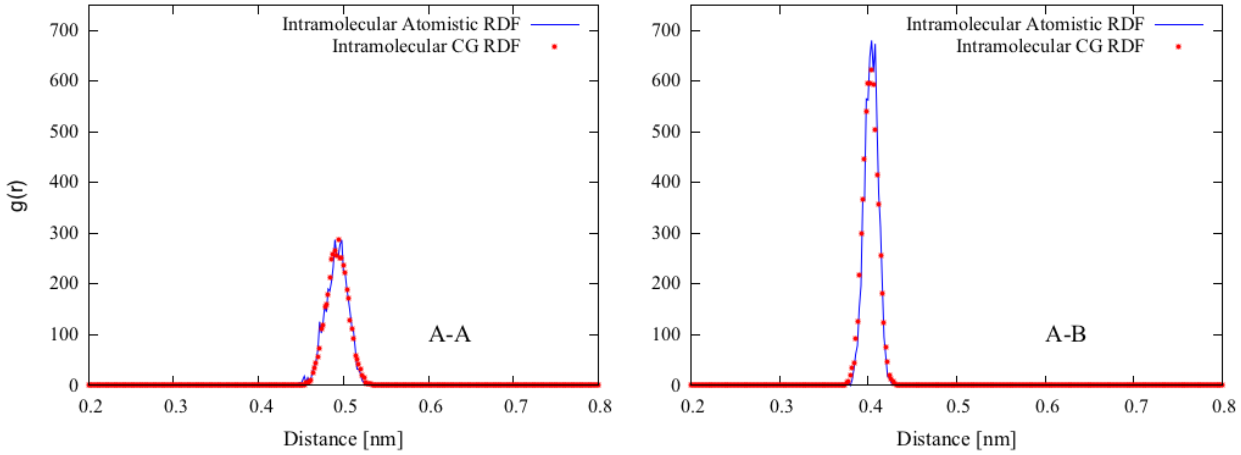


Figure 2.17: Intramolecular atomistic monomer RDFs obtained from atomistic simulation of a single phenytoin molecule (blue). The bonded parameters were determined by matching intramolecular CG RDFs (red) to their atomistic counterparts.

We adopted two approaches to parameterize the cross interaction terms in our force field. We used geometric mixing rules (Equation 2.7) to approximate the cross interactions between polymer bead types from different monomer substitution types. This approximation has been shown to be reasonable for modeling the solvation behavior of heterogeneous methylcellulose chains in the previous section. To model the cross interactions between polymer beads and phenytoin beads, we obtained the interaction parameters explicitly by conducting atomistic simulations with 15 homogenous 20-mer oligomer chains and 150 drug molecules in a cubic simulation box of 12nm on the side. We mapped the CG intermolecular polymer bead-drug bead RDFs to their atomistic counterparts, shown in Figure 2.18. For each such interaction there are multiple RDFs, because each monomer contains up to three CG bead types and each phenytoin

molecule contains two CG bead types. The explicit cross terms are listed in the supplemental information.

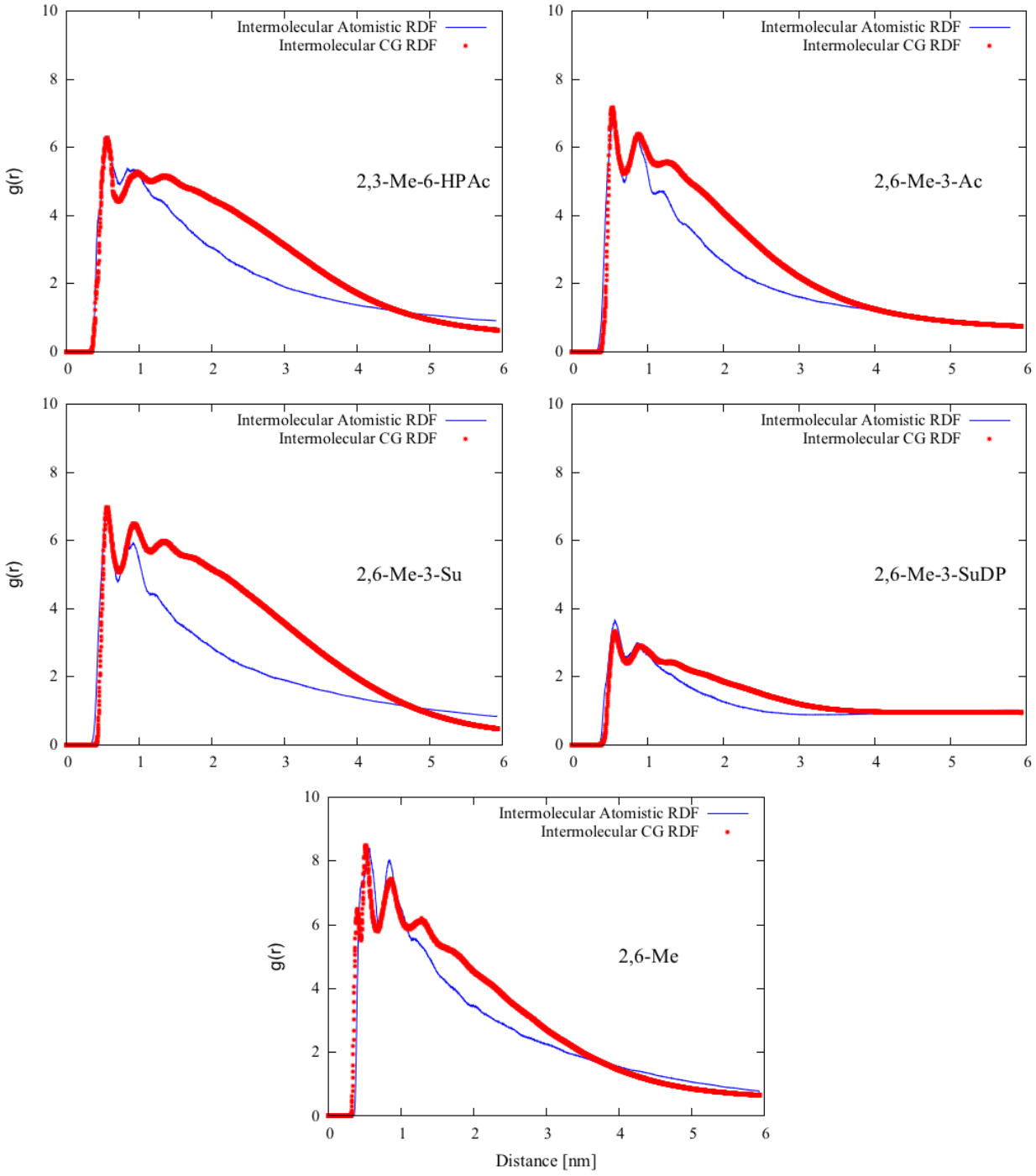


Figure 2.18: Intermolecular atomistic oligomer-drug monomer RDFs obtained from atomistic simulations of 15 chains of 20-mer homogenous HPMCAS model oligomers and 150 phenytoin molecules (blue). The explicit cross interaction parameters were determined by fitting the intermolecular CG oligomer-drug RDFs (red) to their atomistic counterparts using a modified IBI technique.

2.9 Coarse-Grained Phenytoin Model Validation

The CG phenytoin model presented in this work, referred to here as an "aggregation-based" CG force field, has two key differences from the CG phenytoin model developed by Mandal *et al.*,[83] referred to here as a "crystal-based" CG force field. First, the aggregation-based non-bonded parameters are derived from simulations of amorphous phenytoin aggregates in this work instead of crystalline phenytoin. Second, an analytical LJ potential is used to model the non-bonded interaction instead of a tabulated potential. Even with detailed tabulated potential capturing all intermolecular interaction details, the CG simulation using the crystal-based CG force field cannot capture the nucleation event due to the limitation of simulation time, although it can capture crystal growth from a starting crystal seed. Because our main purpose is to model the interactions between polymer and drug in a solid dispersion, and phenytoin nucleation does not occur over the time scale of microseconds accessible with our CG simulations, we believe that an analytical LJ potential captures enough of the details of the intermolecular interaction for our purposes. Nevertheless, it is still worth comparing simulation results generated by these two phenytoin force fields, to see the sensitivity of results to the details of the phenytoin force field.

In Figure 2.19 we show snapshots from atomistic and CG simulations of phenytoin. Starting from a randomly dissolved state, phenytoin molecules aggregate in a short period of time. The two CG phenytoin force fields yield very similar phenytoin cluster structures, and they both resemble the atomistic phenytoin cluster. The intermolecular molecule COM RDFs of the three structures are shown in Figure 2.20. The intermolecular RDF of the atomistic structure has a number of discrete peaks, due to the mapping of explicit atomistic level details into the bead representation. The aggregation-based CG force field yields an RDF that captures primarily the first peak. This is because a bias is placed on fitting the first peak when adopting the modified IBI fitting scheme. The crystal-based CG force field yields a RDF that captures primarily the second peak. We have computed the intermolecular molecule COM RDF generated from a phenytoin crystalline slab (Figure 2.20 inset) and find that, unlike the RDF generated from atomistic phenytoin aggregate, the second peak in the RDF is the most predominant one.

Therefore, the second peak is captured during the parameterization and manifested in the RDF computed for a phenytoin cluster using a crystal-based CG force field.

We have also tested the effect of box size on the concentration required for phenytoin molecules to aggregate, referred to as the aggregation concentration. In a cubic simulation box of 12nm on each side, 55 phenytoin molecules were required to form a persistent phenytoin aggregate. This translates to a phenytoin aggregation concentration of over 10,000 μg/mL, which is far greater than the 32 μg/mL reported in the solubility database[85]. However, as the size of the box increases, the aggregation concentration decreases substantially to less than 6,000 μg/mL for a simulation box of 105nm on each side. It is reasonable to conclude that as the system size increases, the aggregation concentration will continue to decrease and might eventually approach the phenytoin solubility value for a macroscopic system. In addition, given that phenytoin has a high crystallization tendency, we believe the phenytoin aggregates observed in our simulation will eventually turn into crystals, if the force field is capable of simulating and capturing nucleation event.

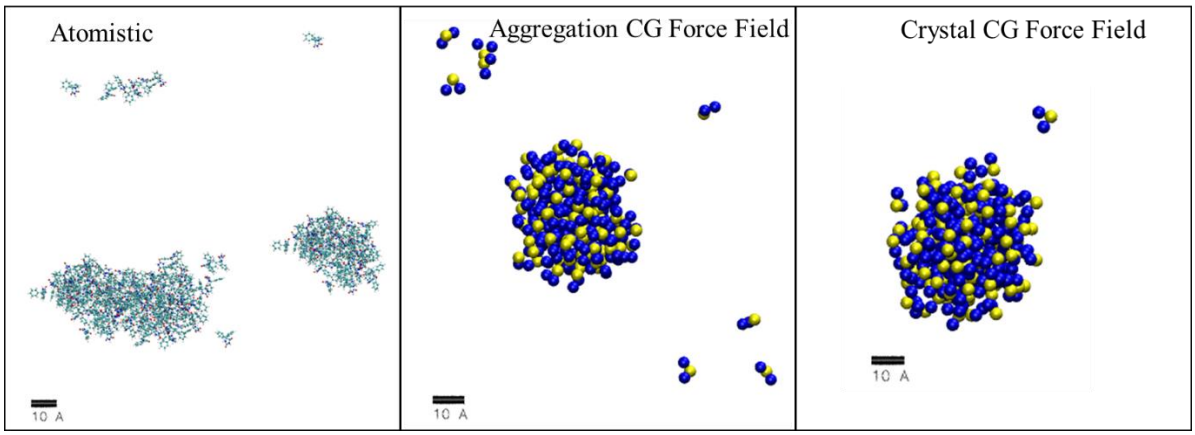


Figure 2.19: Snapshots from atomistic and CG simulations of phenytoin-only systems containing 150 phenytoin molecules in cubic simulation boxes of 12nm on the side (~3.5 wt%), after a simulation time of 30ns and 500ns for atomistic and CG simulations respectively. The CG simulations use the aggregation-based CG force field (middle) and the crystal-based CG force field (right), respectively.

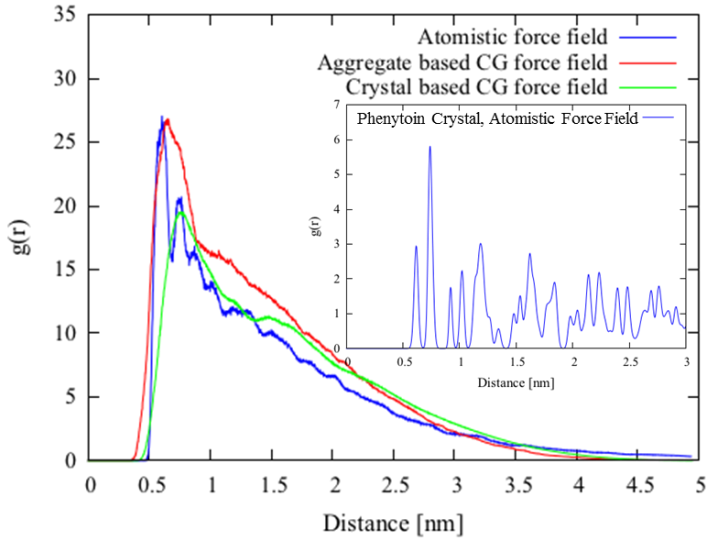


Figure 2.20: Intermolecular bead-bead RDFs obtained from atomistic and CG simulations of phenytoin aggregates. The inset shows the intermolecular bead-bead RDF obtained from atomistic simulation of a bulk phenytoin crystal.

2.10 Concluding Remark

In this chapter, we described the force fields used to model two cellulosic polymers, namely methylcellulose (MC) and hydroxypropyl methylcellulose acetate succinate (HPMCAS), and drug molecule phenytoin. We first compared the two atomistic GROMOS force fields and the AMBER force field. Despite some differences in conformation distributions at the dimer level, both GROMOS force fields gave similar predictions for the stiffness of longer chains. We then presented a systematic approach to develop coarse-grained force fields for methylcellulose and HPMCAS using a single bead per monomer backbone scheme. These coarse-grained force fields are obtained based on atomistic simulations of short methylcellulose oligomers. We have validated these CG force fields against available experimental data and existing simulation results in the literature, including radius of gyration, diffusivity, dissociation free energy, and radial distribution function. In the next two chapters, we will use these force fields to study the solvation behavior of methylcellulose oligomers and explore the gelation mechanism, and study the interaction mode between HPMCAS oligomer and phenytoin drug molecules.

Chapter 3: Modeling Commercial Methylcellulose

Some of the materials in this chapter are results of a collaborative work with Dr. Indranil S. Dalal, Dr. Prateek K. Jha, Rahul Ramesh, Dr. Valeriy V. Ginzbury, Robert L. Sammler, Ming Huang, and Qi Lei.

3.1 Introduction

Commercial methylcellulose (MC) products developed by the Dow Chemical Company, marketed under the brand name METHOCEL™ A are widely used in agricultural, ceramic processing, construction, and pharmaceutical industries. MC is categorized as safe to use as a food additive by the U.S. Food and Drug Administration[15]. Besides its commercial value, MC is of considerable scientific interest as a self-attractive semi-flexible water-soluble random copolymer that forms a gel at elevated temperatures, and has been studied both experimentally and theoretically[86,87]. To form the MC monomer, up to three reactive hydroxyl groups (-OH) on the natural cellulose monomer are substituted with hydrophobic methyl groups ($-CH_3$). Methylcellulose monomer, as a result of this substitution, can have a range of hydrophobicities and degrees of substitution (DS), defined as the moles of substituents (i.e. hydrophobic methyl groups) per mole of MC monomer[22]. For convenience, we will here take the term "methylcellulose (MC) monomer" to include the cellulose monomer as a limiting case with DS=0. We will also use the term "cellulosic" to encompass both unmethylated and methylated monomers and polymers throughout the paper.

Commercial METHOCEL™ A product has two important properties – it is soluble in water and can form a thermoreversible gel at elevated temperatures. Fully substituted MC (i.e. DS=3) is insoluble in water because of its strong hydrophobic interactions. Cellulose (DS=0) is also insoluble in water as the presence of multiple hydroxyl groups on a monomer gives rise to intramolecular hydrogen bonds results in cellulose crystallization. However, MC with DS around 2 is water soluble, as a result of the balanced effect of breaking the tendency of cellulose to crystallize via incorporation of hydrophobic methyl groups while retaining enough hydrophilic hydroxyl groups to sustain enough hydrogen bonding with water[50]. The MC gel morphology has

been revealed recently by Lott *et al.*[27] using cryo-TEM. They found the gel to be a network of fibril structures with a uniform diameter of around 14±2nm above 55℃. The fibril formation mechanism however is still unclear. Here we show through atomistic MD simulations that the hydrophobic interaction is the major driving force for oligomers to form clusters, in line with the earlier theory proposed by Kato et al.[28] However, the earlier theoretical works could not provide a satisfactory answer regarding why the diameter of the fibril stops increasing beyond 14nm.

Collapse transitions of various self-attractive semi-flexible polymers have been well-documented in experimental studies[88,89]. Examples of self-attractive semi-flexible polymers include biopolymers such as F-actin and DNA, and synthetic polymers such as Kevlar and Zylon used in high-performance fibers. Many simulation studies[90–93] have been carried out over the past decade to characterize the collapse transitions of these self-attractive semi-flexible polymers under different solvent conditions. A recent systematic simulation study by Kong and Larson[92] for example, detailed various collapsed states and collapse paths exhibited by semi-flexible polymers with different persistence lengths and attractive interaction strengths. Other important simulation studies have shown similar collapsed states[94–97]. These transitions often involve long-lived intermediate states such as "hairpin structures" and final compact states including toruses, condensed globules, and folded bundles. These previous simulation studies of semi-flexible polymers utilized a top-down approach, involving generic bead-spring models that were not targeted to any specific polymer chemistry. Here, we describe a bottom-up approach, with the force field designed to represent the specific physico-chemical properties of the semi-flexible methylcellulose polymer. While the CG force field is specialized to the polymer of interest, our method of developing it is general and the structures we observe in our simulations, especially rings or toruses, have shown up in simulations of generic models, indicating their likely occurrence in collapse transitions of other semi-flexible polymers.

3.2 Atomistic Simulation Results

We apply GROMOS 56Acarbo force field to simulate model cellulosic hetero-oligomers whose monomers occur with probability matching that of the METHOCEL™ A chemistry. To do so, we set up four model systems with oligomer lengths of 10, 20, and 40 monomers. To increase the randomness of our model oligomers, we first generate one 90-monomer random sequence, and break it into nine 10-mers, each with different sequence. Similarly, we generate two 400-

monomer sequences and break one of them into 20 random 20-mers and the other into 10 random 40-mers. Each initial sequence of 90 or 400 monomers contains monomer probabilities based on the monomer probabilities of METHCOEL™ A chemistry, defined in Table 3.1, with a degree of substitution (DS) of around 1.9. We choose 20nm cubic boxes for the 20-mer systems and 22nm cubic boxes for the 40-mer system so that the side of each box is larger than both the length of the oligomer chain and the diameter of a methylcellulose fibril gel reported in the literature, namely 15nm[27]. (While we initially had some hopes of seeing evidence of such fibrils in our simulations, we were not able to detect them, probably because of inadequate length and time scales.) The oligomer concentrations range from 3wt% to 6.1wt%. Specifically, the simulations labeled "10-mer 4wt%", "20-mer 3wt%", "20-mer 6wt%", and "40-mer 4wt%" contain the oligomers at concentrations of 4.5wt%, 3.1wt%, 6.1wt%, and 4.4wt% respectively. Although these concentrations are slightly higher than in the solutions used for many experiments (2wt%), these higher concentrations are chosen to increase the number of oligomers in boxes that are small enough to be simulated with available computational power. Nevertheless, the experimentally determined gelation temperature is known for methylcellulose solutions in the range of concentrations simulated here, and is between 40℃ and 45℃. We therefore simulate at two temperatures, 25℃ and 50℃, which are respectively below and above the experimental gelation temperature.

	Average Mole Fraction
Cellulose	0.05
2-MC	0.13
3-MC	0.02
6-MC	0.10
2,3-MC	0.10
2,6-MC	0.26
3,6-MC	0.05
2,3,6-MC	0.29
Total DS (Me)	1.94

Table 3.1: Average mole fraction of each methylated cellulose monomer in METHOCEL™ A chemistry

The snap-shots of the four systems at 25℃ and 50℃ are shown in Figure 3.1 and Figure 3.2. We observe that the "10-mer 4wt%" system, which has a DS of 1.87, behaves similarly to the 10-mer multiple homo-oligomer di-substituted methylcellulose systems (DS=2) we discussed in section 2.3, with chains forming a globular aggregate at both the low and high temperatures. The "20-mer 3wt%" system, however, forms a ramified aggregate at low temperature and a more rod-like

structure that spans the box at high temperature. The "20-mer 6wt%" and the "40-mer 4wt%" systems form three-dimensional "gel-like" networks at both low and high temperatures. At the elevated temperature in Figure 3.2, the snap-shots for these two systems do not appear significantly different from those at low temperature, suggesting that these two systems might be trapped in a local energy minimum. We show close-in views of the "40-mer 4wt%" and "20-mer 6wt%" structures at 50℃, in Figure 3.3, which reveal that the oligomers condense into bundles with parallel chain alignment, seemingly to minimize the contact between hydrophobic methyl groups and water. Although this parallel configuration between polymer associating chains agrees with the proposed structure-formation mechanism by Bodvik *et al.*[98], who observed a methylcellulose fibril structure in cryo-TEM images, we suspect this is not the actual configuration the flexible methylcellulose chains adopt in the cross-linked gel structure due to the following. Firstly, the parallel configuration observed in our simulations does not explain the observation of fibrils with a uniform diameter of 15nm in methylcellulose, as reported by Lott *et al.*[27], because there is no apparent reason that the bundle formed by chains condensing parallel to each other would thicken to a specific diameter of 15nm, rather than continuing to thicken indefinitely. Secondly, the flexible methylcellulose chains in METHOCEL™ A products have chain lengths ranging from hundreds to thousands of monomer units, which will allow formation of helices with radius in nanometer range (i.e. 15nm) due to self-attracting forces. The model oligomers simulated here (10-mer, 20-mer, and 40-mer) are too short, given their persistence length, to bend enough to form helices. Thus, in our simulations, the minimization of methyl-group hydrophobic interactions and formation of intermolecular hydrogen bonds occurs most readily through parallel configurations of different short chains, rather than formation of a helical structure by a single, or a few, much longer chains. We suggest that although the parallel chain configurations observed in our short-chain simulations may not be representative of the final gel structures that much longer chains would form, they might be precursor to the final self-assembly of methylcellulose chains at larger length scale and longer time scale, which are only accessible through coarse-grained simulations. Srinivas *et al.* have reported such self-reorganization from clusters of parallel chains to helically twisted columns in their coarse-grained simulations of amphiphilic molecules[99], and something similar occur with methylcellulose chains, which we will later demonstrate using CG simulations. However, an interesting experimental work is to probe the gelation mechanism of much shorter methyl

cellulose chains than have been studied heretofore, namely chains of only 10-40 monomers similar to those studied in our simulations. If short methylcellulose chains such as these do not form 15-nm fibrils, but form much less regular gel structures, this might support the suggestion here that a minimum chain length is required to form the regular fibrils seen experimentally.

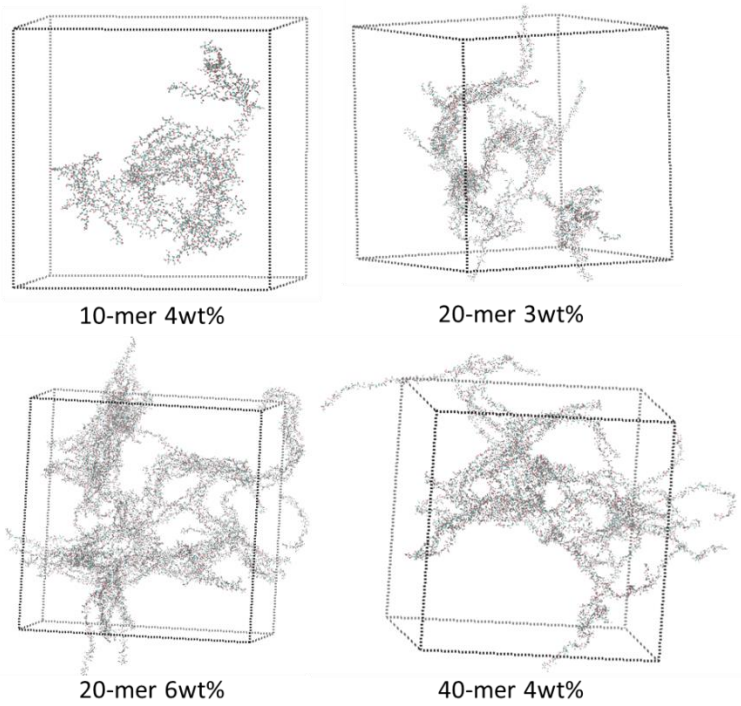


Figure 3.1: Snap-shots after 35ns for simulations of METHOCEL™ A in water at 25℃

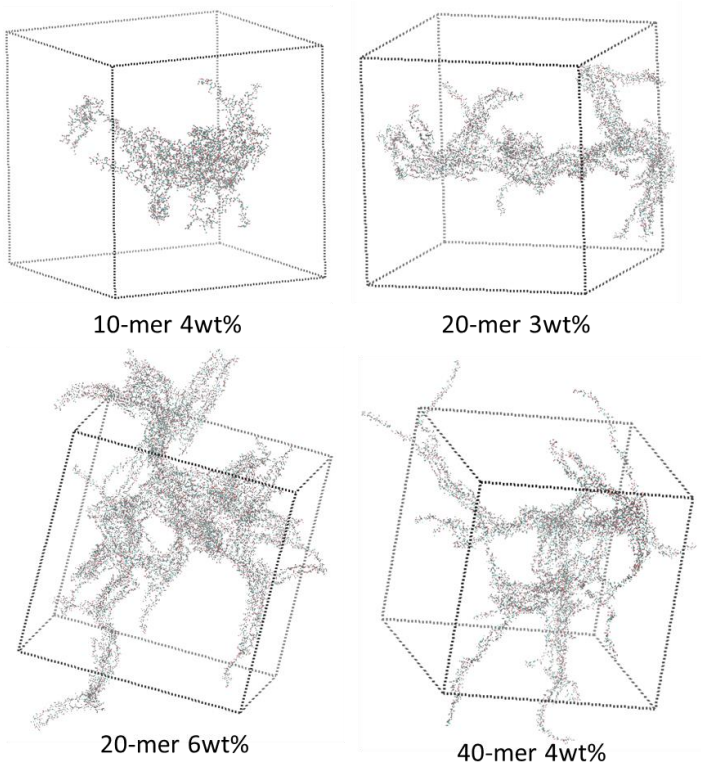


Figure 3.2: The same as Figure 3.1, except at 50℃

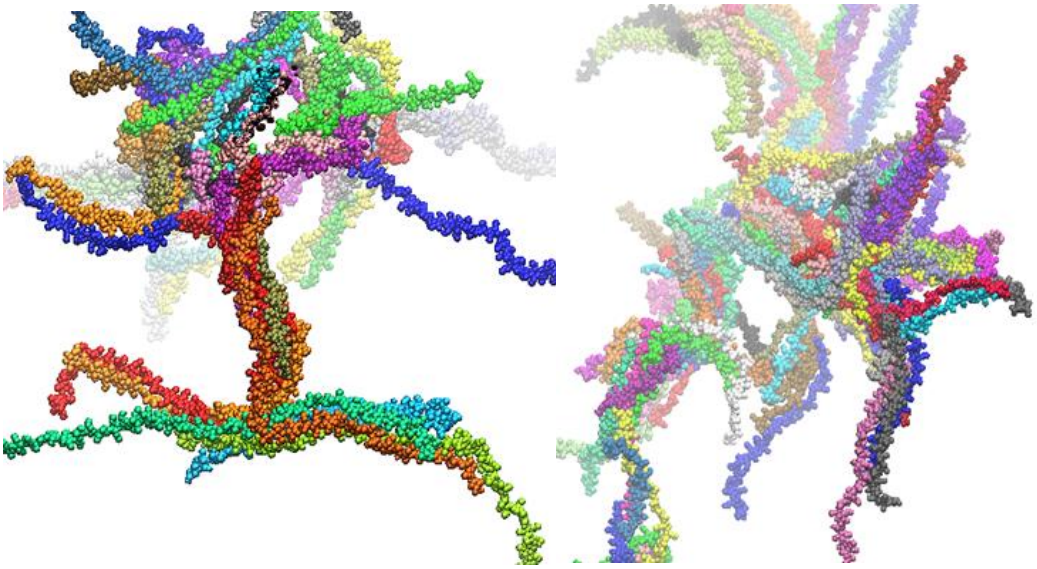


Figure 3.3: Close-in views of the METHOCEL™ A "40-mer 4wt%" (left) and "20-mer 6wt%" (right) systems at 50 ℃ show that model METHOCEL™ A oligomers condense parallel to each other. Each chain is colored differently.

We compute the oligomer-oligomer RDFs of all four systems at both room and elevated temperatures and plot the heights of the RDF peaks that occur roughly at r=1.2nm against oligomer weight percentage in Figure 3.4. The peaks of the RDFs at elevated temperature for all four systems are consistently higher, corresponding to closer-packed aggregates, than at room temperature, which is consistent with our earlier observations for homo-oligomers. In addition, it is clear that the heights of the RDF peaks decrease as the oligomer weight percentages increase. The 20-mer 6wt% system, for example, has the lowest RDF peak and shows barely any peak height change in response to the temperature change. Interestingly, even though the "10-mer 4wt%" and "40-mer 4wt%" simulations are in boxes of different sizes, namely 12nm and 22nm, they have very similar RDF peak positions (r≈1.2nm) and similar peak heights at both low and elevated temperatures. This suggests that for these relatively short oligomers (10 to 40 monomers), peak height is more sensitive to concentration than to oligomer size, and that we can obtain representative results using cheap simulations of "10-mer" oligomers in a small simulation box.

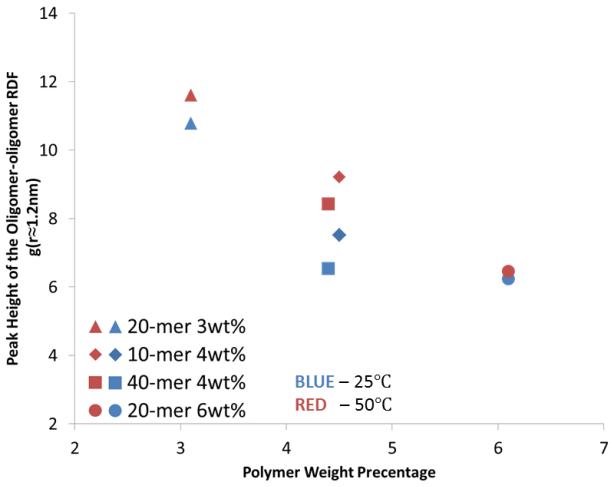


Figure 3.4: Peak heights of oligomer-oligomer RDFs (at r≈1.2nm) in 4 METHOCEL™ A model oligomer solutions at room temperature (25℃), shown in blue points, and elevated temperature (50℃), shown in red.

We next compute the numbers of hydrogen bonds formed, both between oligomers and water and between oligomers and oligomers (Table 3.2). We normalize the total hydrogen bond counts by the numbers of 10-mer lengths in each system so that we can not only compare results among the four model random co-oligomers but also compare the results for the random co-oligomers

with those for simulations of multiple homo-oligomer chains reported in the previous section. All four co-oligomers (DS≈1.9) show very similar oligomer-water and oligomer-oligomer hydrogen bond counts. Around 90% of the oligomer-oligomer hydrogen bonds in the co-oligomers are intramolecular, and the rest are intermolecular (i.e., between different chains). For all four systems, as the temperature increases, the number of oligomer-water hydrogen bonds decreases and the number of oligomer-oligomer intermolecular hydrogen bonds increases slightly. Thus, similar to the RDFs, the shortest oligomer, namely a 10-mer, yields the same normalized hydrogen bond counts and same temperature-dependence of hydrogen bonding as in the simulations in a much bigger box with longer model oligomers, namely the "20-mer 3wt%", "20-mer 6wt%", and "40-mer 4wt%" systems. Nevertheless, the larger aggregates in the "20-mer 3wt%", "20-mer 6wt%", and "40-mer 4wt%" systems are more ramified than those in the "10-mer 4wt%" solutions, which may be crucial to the gel-formation mechanism, although the simulation length and time scales are nowhere close to the scale of experimental gels. Our results indicate that, while we are not able to simulate the large-scales structures formed by these gels, the numbers of hydrogen bonds, RDF peaks, and other metrics of local structure on the scale of a few nanometers, are insensitive to the length of the oligomers, and therefore may be correctly predicted by the simulations.

	hydrogen bond breakdown				oligomer-oligomer hydrogen bond breakdown			
	25C		50C		25C		50C	
	oligomer-water	oligomer-oligomer	oligomer-water	oligomer-oligomer	intra	inter	intra	inter
10-mer 4wt%	30.7(7)	5.1(2)	26.5(5)	5.1(2)	91%	9%	89%	11%
20-mer 3wt%	28.8(5)	5.0(1)	25.3(7)	5.1(1)	92%	8%	91%	9%
20-mer 6wt%	28.2(4)	5.2(1)	24.8(7)	5.2(1)	92%	8%	90%	10%
40-mer 4wt%	29.7(6)	5.2(1)	25.2(3)	5.4(1)	93%	7%	92%	8%

Table 3.2: Numbers of hydrogen bonds, per 10 monomers in the chain, with standard error, and percentages of different hydrogen bond types in the METHOCEL™A model systems. The normalization of the hydrogen bond counts per numbers of 10-mer sub-chains, means that numbers of hydrogen bonds in a 20-mer chain and in a 40-mer chain are obtained from the entries in the table by multiplying by 2 and 4 respectively.

To determine whether methylcellulose gelation is induced by the more hydrophobic monomers or by the more hydrophilic ones in heterogeneous methylcellulose, such as those used commercially, we compute contact maps of our model systems. We define two monomers to be "in contact" if the distance between any one atom on one of the monomers is less than 0.4nm

distance from any atom on the other. For the "10-mer 4wt%", "20-mer 3wt%", "20-mer 6wt%", and "40-mer 4wt%" systems studied here, three repeat simulations with random initial configurations are set up and simulated for 35ns. A 200ps production run is then conducted for each simulation and five frames are taken separated by 50ps intervals and used to compute the average number of contacts for each of the 64 types of monomer-monomer pairs, where the monomer "types" are distinguished by the eight methyl substitution patterns, ranging from no methyl substituents, to tri-substituted MC. Given the fractions of each type of monomer present, and the total number of contacts, we compute the numbers of each pair of monomer types that would be expected if the contacts between them were random, and we use this to normalize the raw counts of actual contact numbers. A value greater than unity then implies that the contact occurs more often than would be expected if monomers form contacts randomly with other monomers. We then average these normalized counts over five frames, divide by the number of monomers in the system, and average the results over three simulations for each oligomer system to produce the contact map for each of the four systems studied at two different temperatures. At each temperature, we observe that the majority (~90%) of the contact values are within the standard error among the four systems. We therefore further average the values in the four contact maps at the same temperature to produce contact maps averaged over the four oligomer model systems, at 25℃ and 50℃, shown in Figure 3.5a and b respectively. The corresponding standard errors, presented in Figure 3.5c and d, are generally below 0.1, indicating the contact maps we generated are largely insensitive to the randomly generated sequence based on the statistics of the commercial METHOCEL^{TM} A products. A few contact-map values involving cellulose, 3-MC, and 3,6-MC, however, have relatively large standard error, likely due to the low probability of presence in the METHOCEL^{TM} A product (<0.05 mol%). We use the "hot spots" (i.e., high normalized contact number, marked in red or purple) on the contact maps to suggest which monomer pairs are most responsible for inducing gelation at elevated temperature. The normalized count of the "2,3,6-MC–2,3,6-MC" pair is the highest among all pairs at both temperatures, and the count at elevated temperature is larger than that at room temperature. This suggests that hydrophobic interactions between the tri-substituted monomers are responsible for the formation of aggregates at temperatures above the experimentally determined gelation temperature. The intensities of the "Un–Un" (Cellulose–Cellulose) pair also increases at elevated temperature. This observation correlates well with the increased number of intermolecular

hydrogen bonds at elevated temperature shown in Table 3.2. However, the "Un-Un" pair intensities don't exceed the average pair contact intensity (e.g. intensity <1.0), suggesting that the hydrogen bonding is not the main driving force for the formation of aggregates in these methylcellulose oligomer systems. Our observation from these contact maps agrees with the methylcellulose gelation mechanism proposed by Kato *et al.*[28] and Li *et al.*[100] in which tri-substituted methylcellulose units act as hydrophobic junctions.

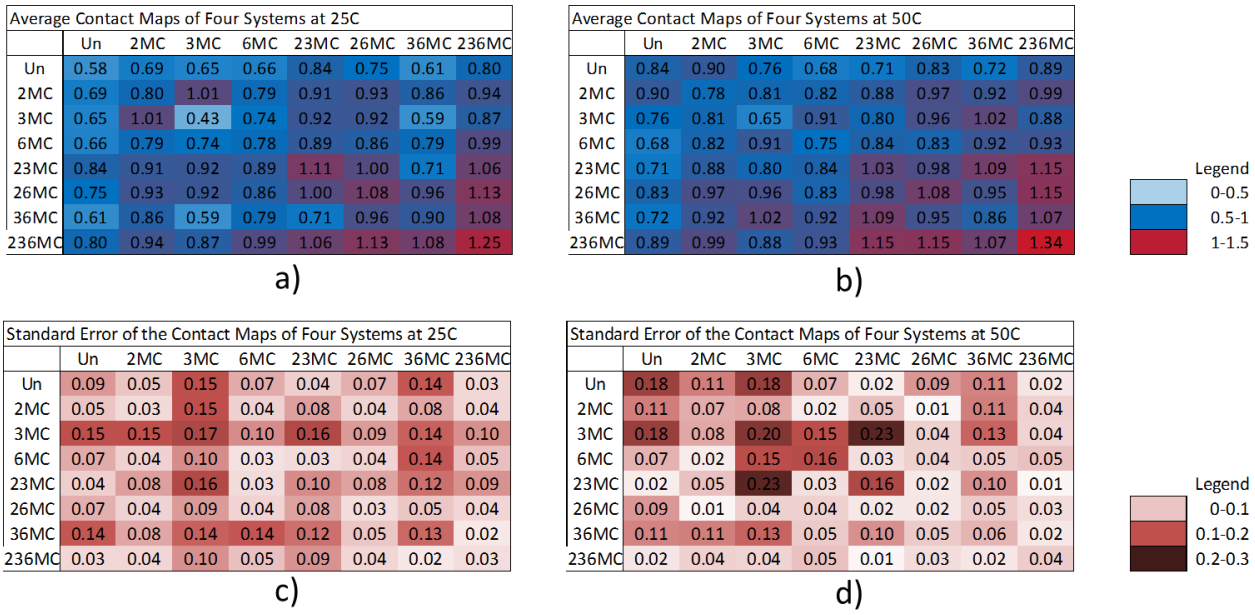


Figure 3.5: Contact maps averaged over four model METHOCEL™ A systems ("10-mer 4 wt%", "20-mer 3 wt%", "20-mer 6 wt%", and "40-mer 4 wt%") at 25℃(a) and at 50℃(b). Each tabulated value in contact maps is an averaged value over the actual number of contacts of that type obtained from the four model systems, divided by the number of that type that would exist if the same total number of contacts were assigned randomly, based on the fraction of each monomer type present. The corresponding standard errors of the two contact maps are shown in (c) and (d).

3.3 Coarse-Grained Simulation Results

Chain-Length Dependence of Heterogeneous MC Radius of Gyration

We now built model heterogeneous CG oligomers at various chain lengths based on the monomer substitution type mole fractions in the commercial polymer METHOCEL™ A. Based on the mole fractions shown in Table 3.1, we generated three random sequences at each chain length to minimize the sequence bias that can be introduced in the random substitution process. We simulated these chains at both room and elevated temperatures and obtained the R_g values. In Figure 3.6, we compare R_g from simulations to two sets of experimental data. The data from Li and from Patel et al. both fall nicely on the theoretical line for a persistence length of 10.9±0.6

nm. The l_p estimated from our simulation data at room temperature is 8.9±0.9nm, in good agreement with the experimental value. One possible reason for slightly lower persistence length from our CG model could be that the dihedral force constant (K_φ) we picked is rather weak. The dihedral angle potential function has a minimum value at 180 degrees, which corresponds to the preferred trans conformation among four consecutive monomers on a cellulosic chain. We parameterized the K_φ value based on the intramolecular RDF generated for a rod-like 10-mer chain, whose conformation is fairly insensitive to the choice of the K_φ value. In the long-chain regime though, the persistence length of the chain may be sensitive to the choice of K_φ value. We attempted to increase the K_φ value but we were forced to take very small time step to ensure stable simulations. Therefore we concluded that we should keep the K_φ value at 2ε, to aim for a balance between the simulation performance and model accuracy.

At elevated temperature, a clear collapse transition occurs at chain lengths between 400 and 600 monomers. This transition point aligns well with that for 2,6-MC and 2,3,6-MC at 50C, which implies that the highly substituted monomers in a heterogeneous chain drive the collapse transition in MC, agreeing with the atomistic simulation shown in previous section. The collapsed chain maintains a ring conformation with outer diameter about 14nm, which will be discussed in more detail in the next section. We also tested the sensitivity to the strength of the intermolecular interaction, by lowering the ε_{ii} values at 50℃ by 20% for all eight monomer substitution types and repeating the simulations. We have already shown for all monomer substitution types ε_{ii} is lower at 25℃ than at 50℃ (Figure 2.9). Thus, this reduction in ε_{ii} by 20% corresponds to cooling the MC solution from 50℃ to a 43℃ if we assume ε_{ii} value scales with temperature linearly. We plot the R_g values trend for chains at 43℃ in the inset to Figure 3.6 and obtain a persistence length l_p of 5.4±0.7nm. The chain is aggregating, but visual inspection of the structure shows that it is not forming a collapsed ring structure. This suggests that the chain starts to collapse only at a temperature that is close to or above the gelation temperature.

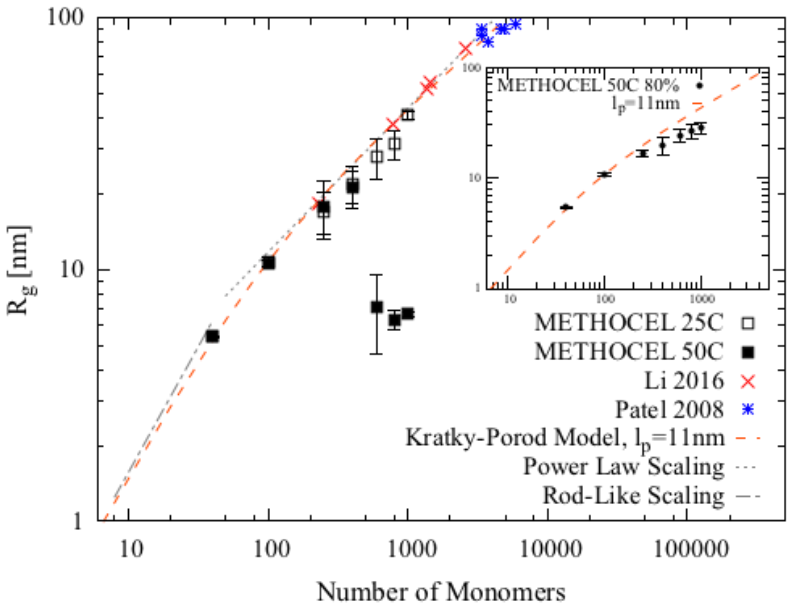


Figure 3.6: Simulated chain-length dependence of R_g for methylcellulose model oligomers at two temperatures (25℃ and 50℃), along with experimental data provided by Li (2016) and Patel et al. (2008). The inset shows R_g for methylcellulose model oligomer with intermolecular interaction value (ε_{ii}) for each monomer type set to 80% of the value at 50℃, corresponding to 43℃ if we assume ε_{ii} value scales with temperature linearly.

Formation of Ring Structure as a Precursor to Fibrilar Gel Formation

At a temperature that is slightly above the typical gelation temperature for a relatively dilute MC solution (<2 wt%, 50℃), we observe the formation of a ring structure in single-chain simulations. In Figure 3.7a, we plot R_g as a function of time in a simulation that started with a fully stretched chain. The stable ring is formed after 2μs and shows a consistent R_g value around 10nm. If we simulate a quench from 50℃ to 25℃ by changing the interaction parameters to the lower-temperature ones, the ring structure unravels and reaches the equilibrium random coil conformation in about 1μs (Figure 3.7b). We can use the radius of gyration tensor to further characterize the structure of the ring. The eigenvalues of the gyration tensor are expressed in Equation 3.1:

$$R_g^2 = \lambda_x^2 + \lambda_y^2 + \lambda_z^2 \tag{3.1}$$

The eigenvalues are ordered such that $\lambda_x^2 \geq \lambda_y^2 \geq \lambda_z^2$. We plot the eigenvalues of the radius of gyration tensor in Figure 3.8b and indeed we see λ_x and λ_y are almost identical. Some minor fluctuations are observed in the λ_z value. These correspond to self-reorganization and fluctuation of the structure. Snapshots of the ring structure reveal that it has an outer diameter of 13.9±0.4 nm, and an inner diameter of 7.1±0.2 nm. In next few sections, we show that the ring structure is

a precursor to the formation of a long MC fibril. The conditions where we observe the formation of a ring structure can be compared to the conditions predicted for ring formation in a dimensionless phase diagram describing the collapsed states of semi-flexible self-attracting polymers.

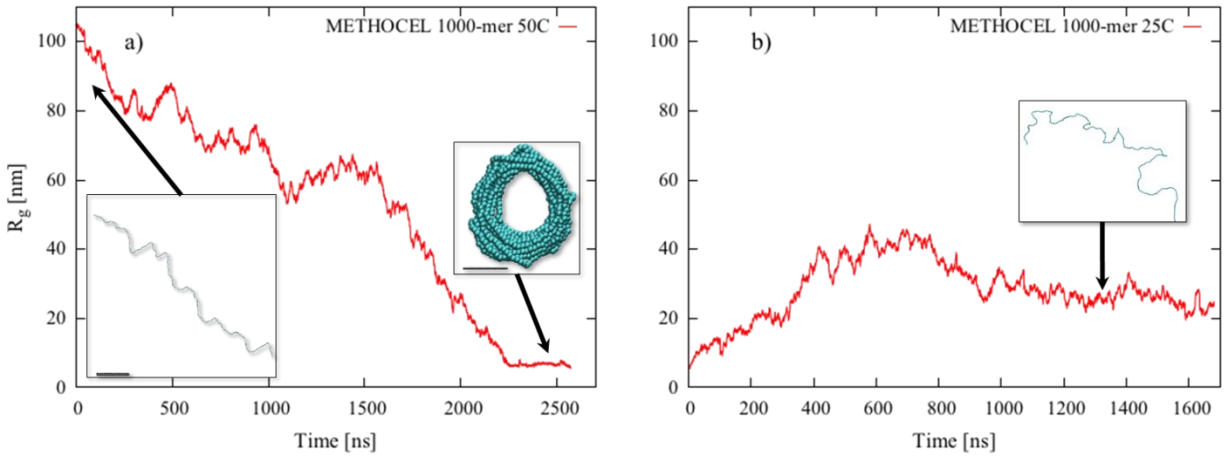


Figure 3.7: R_g versus time for model METHOCEL^{TM} A oligomer. a) At elevated temperature the chain transforms from a stretched state (left inset) to a collapsed ring state (right inset). The scale bar for the left inset is 50nm and 5nm for the right inset. b) R_g at room temperature starting from the ring structure formed at high temperature, showing re-expansion of the polymer.

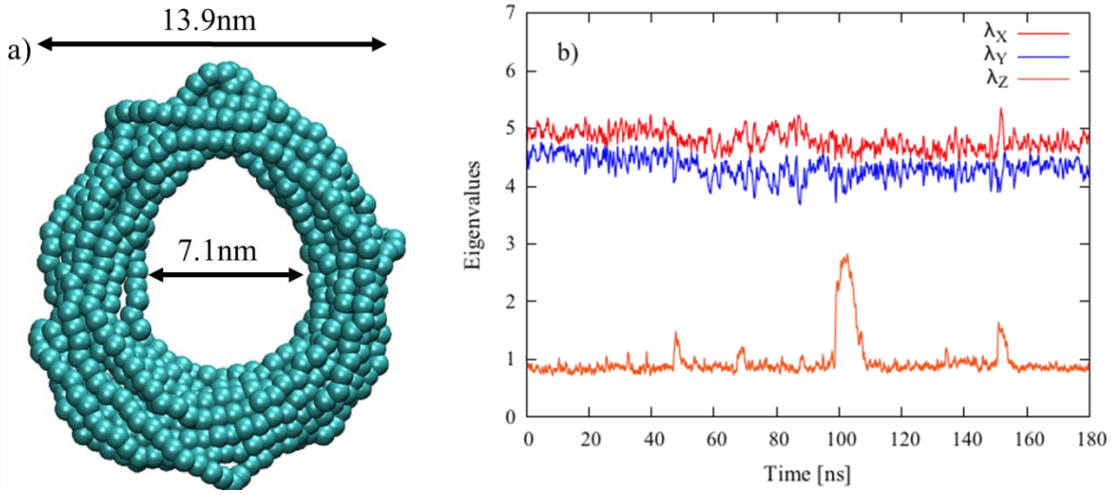


Figure 3.8: a) Snapshot of the stable ring structure formed by heterogamous methylcellulose chain at elevated temperature. b) Three eigenvalues of the radius of gyration tensor.

Concentration-Dependent Assembly of Multiple Methylcellulose Chains into Rings

We now focus on the ring formation of multiple methylcellulose chains. Figure 3.9 depicts simulation starting from a random configuration of three MC chains of 1000 monomers long at

different concentrations. The cubic simulation box size was set to 400, 180, 80, 60, and 45 nm on the side, corresponding to concentrations of 0.002, 0.017, 0.19, 0.46, and 1.09 weight percent. The model polymer chains were randomly placed in the box (Figure 3.9a) and allowed time for adequate equilibration at 25 ℃. The temperature was then increased to 50 ℃ and the final snapshots from each simulation at different concentrations are shown in Figure 3.9b-f. At a very dilute concentration, namely 0.002 weight percent, individual chains formed isolated rings. These rings have been observed in a single-chain study, and the diameters of these rings are independent of the specific repeat unit sequence in the randomly substituted chains; the outer diameter is estimated to be 13.9±0.4nm and the inner diameter is 7.1±0.2 nm. The outer diameter compares well to that (14±2 nm) measured experimentally for long MC fibrils. Similarly, the inside diameter compares reasonably well to that (10.8±3 nm) estimated with the measured fibril diameter and water content assuming an ideal cylindrical hollow tube. Note that at a concentration of 0.002 weight percent, the isolated ring structures did not interact with each other during the course of simulation.

In a dilute solution, namely between 0.017 and 0.3 weight percent, at least one chain first formed an isolated single-chain ring structure, which we refer to as a "seed ring structure". The other chains, whether still in the random coil state or having also formed ring structures, then came into contact with the seed ring and fused with it, forming a single tubular structure. The "height" (extent in the direction parallel to the main axis) of the individual ring is 2.6±0.3 nm. The single stable "proto-tube" structure formed by the three rings has an outer diameter of 16.4±1.3 nm, an inner diameter of 7.9±1.1 nm, and a total height of 4.7±0.3 nm. The ring is packed with many revolutions of methylcellulose CG beads. The thickness of the proto-tube wall increases by 1.25 nm on average after individual rings fuse, corresponding to two revolutions of CG beads, since each CG bead has a diameter of 0.515 nm. The height of the three-molecule proto-tube is 2.1 nm more than the height of a single ring, corresponding to an addition of 5 revolutions in axial direction.

As the concentration of the polymer in solution increases to roughly between 0.5 and 1.1 weight percent, initially dispersed chains formed a bundle structure and evolved into a three dimensional network of bundles. Note that at 0.5 weight percent, bundles formed and bent into imperfect ring **(**Figure 3.9e), suggesting another possible pathway to a ring structure. At 1.1 weight percent however (Figure 3.9f), the bundles formed a three dimensional network of bundles with no rings.

This three dimensional network is similar to the conventional gelation network for methylcellulose proposed by Kato *et al*,[3] though not necessarily easy to reconcile with the fibrillar network observed experimentally. It is also possible, as hypothesized by Lodge and co-workers,[21] that both fibrils and bundles could be metastable structures, but formation of fibrils is preferred for kinetic reasons, for example by rapid growth from seed ring structures that are presented in the solution due to compositional inhomogeneity. At this time, we do not have a clear explanation for why MC solutions of concentration greater than 1.1 weight percent would form hollow fibers rather than fibrillar gels with fibers composed of bundles of parallel aligned MC molecules. If the latter were to occur, however, the fibril diameters would have no strongly preferred value and a distinct fibril diameter of 14 nm would not be expected, nor would the rather low (40%) density of polymer in the fibers be explicable. Hence, for now we hypothesize that some thermodynamic or kinetic factor favors formation of rings even at concentrations above that for which our simulations predict them to form. We proceed, then, to simulate the assembly of such rings into tubular fibrils.

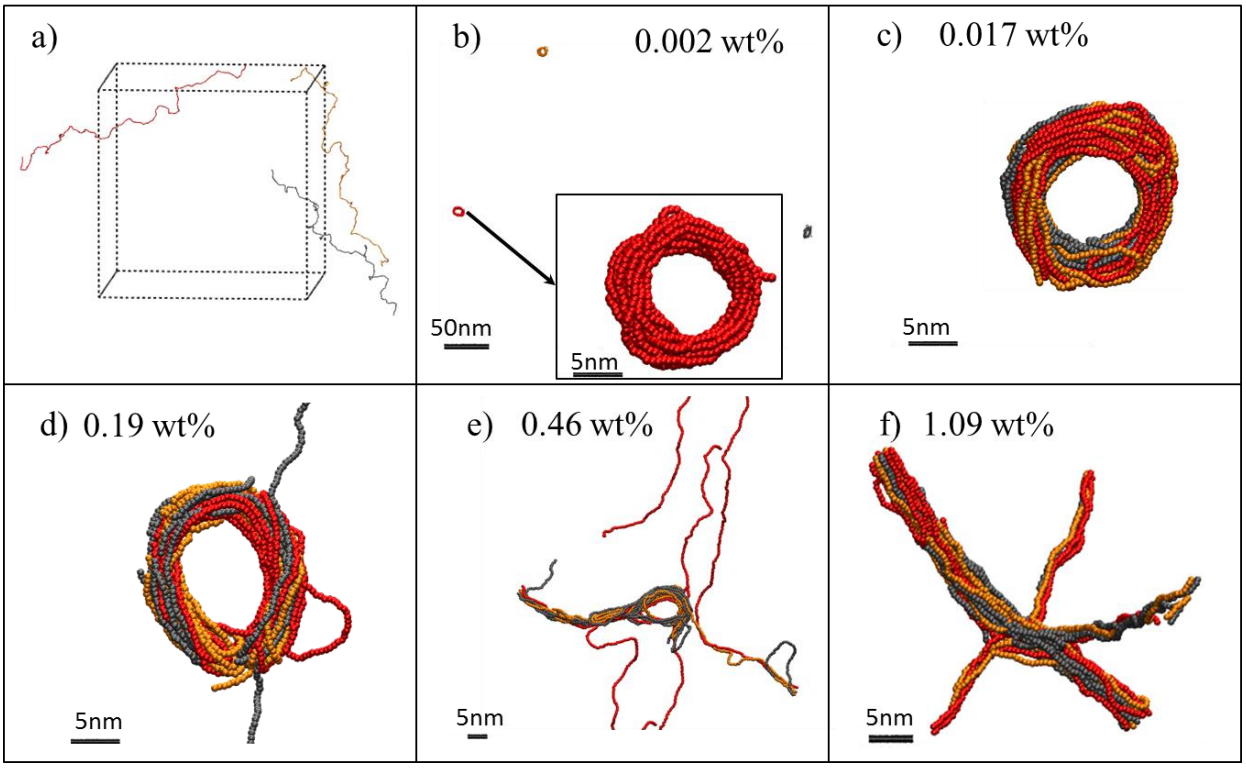


Figure 3.9**.** Snapshots from five simulations of chains with DP = 1000 with varying polymer concentrations. a) At the beginning, three chains are randomly placed in a cubic simulation box with periodic boundary conditions. To vary the concentration, we change the box size L = 400, 180, 80, 60, and 45 nm, corresponding to polymer concentrations of 0.002, 0.017, 0.19, 0.46, and 1.09 wt %. b-f) Final snapshots of the simulations corresponding to various MC concentrations. All simulations were performed at T = 50 ℃.

Self-Assembly of Multiple Rings into a Proto-Tube

Next, we examined how the rings could aggregate into a tubular structure. An initial ring structure was generated by simulating a 1000-repeat unit chain at elevated (50 ℃) temperature, and then replicated five times. We then performed several simulations in which the initial positions of the five rings were varied. In the first simulation (Figure 3.10a), the replicates were initially placed on top of each other with 2nm center-to-center distances and eventually fused into a single proto-tube structure. In the second simulation (Figure 3.10b), the rings were initially placed randomly and ended up forming a single ring, however with some defects. By increasing the simulation box size, we reduced the effective concentration to 1.3 weight percent (Figure 3.10c), and observed that the rings self-assembled and fused into a single tube. This suggests that in heated methylcellulose solution, at lower polymer concentrations, single-chain rings self-assemble into tubular structures. At higher concentrations, the ring self-assembly is more likely to include branch points and other defects.

The proto-tubes, of course, can "polymerize" further, given the right conditions. For example, taking two five-ring structures and putting them one on top of another leads to the formation of a ten-chain tube (Figure 3.11a). The tube has an outer diameter of 17.4±0.6 nm and an inner diameter of 6.3±0.7 nm (Figure 3.11b). We calculate the void fraction in the center of this structure as the square of the ratio of inner diameter over the outer diameter. The void fraction is thus estimated to be approximately 13%, which is substantially smaller than the water volume fraction observed in experiments.[18-20] However, we note the effective density of coarse-grained repeat units is much lower than that of the polymer itself because of the coarse-graining which replaces flat repeat unit by spherical beads. We estimate each cellulosic repeat unit to be a thin cylindrical disk shape that has a height of 0.15nm and diameter of 0.515nm. With this assumption, polymeric material occupies about 45% of volume, and the remaining 55% are effectively interatomic voids. Assuming the wall of the ring is densely packed with CG beads, we can re-map the CG beads to atomistic MC repeat units, and thereby find that 52% of the total tubular structure volume will be voids and can be accessed by water molecule. Note however, that to simulate such a back-mapped atomistic system with 5000 repeat units and explicit water

molecules present is not currently possible. Therefore, while we do not know how the atomistic chains would pack if simulated *de novo*, this mapping calculation yields a reasonably good agreement with the 60% water content percentage reported in the experiments.

To further test if the proto-tube structures can self-assemble without a prearranged or ordered initial structure (e.g. the stacking configuration with a small separation), we set up a simulation with one five-chain proto-tube structure space 5 nm to the side of a ten-chain proto-tube structure (Figure 3.12a). The five-chain structure came into contact with the "cap" of the ten-chain structure and formed a metastable "proto-junction" (Figure 3.12b). This metastable structure persisted for around 100ns before re-arranging to form a single, higher aspect-ratio tube (Figure 3.12c). We hypothesize that if another proto-tube were nearby when the metastable structure is formed, the three proto-tubes could form a stable three-way junction.

To quantify the dimensions of the tubular structures more accurately, we used the radius of gyration tensor approach (Equation 3.1). We plot in Figure 3.13**,** the averaged eigenvalues for all final ring and tube structures obtained from simulations. Two eigenvalues, namely λ_x and λ_y, have very similar values in all structures. These two values correspond to the outer diameter of the axisymmetric ring and tube. Note that these two values increase upon transitioning from a one-chain to a five-chain proto-tubular structure, confirming the growth of the tubular structure's outer diameter observed through visual inspection. Among five-, ten-, and fifteen-chain tubular structures, these two eigenvalues are very consistent, indicating the tube structure's diameter does not change any further. More importantly, the λ_z value, which corresponds to the height of the ring and tube structures, increases almost linearly from a one-chain to a fifteen-chain tube, clearly indicating the tubular structure grows axially once its stable inner and outer diameters are established. Our simulation results thus support the proposed gelation mechanism in the theoretical model outlined below. Individual chains in a random coil state undergo a conformational transition when temperature rises and form isolated ring structures. These ring structures, similar to nucleation sites in crystal growth model, attract other rings and self-assemble into short proto-tubes, and eventually grow into long hollow (water-filled) fibrils with a uniform inner and outer diameter. In our multiple chains simulations, the outer diameter of all tubular structures ranges from 14 to 17 nm, in excellent agreement with 15±2 nm as reported by Lodge and co-workers.[18-21]

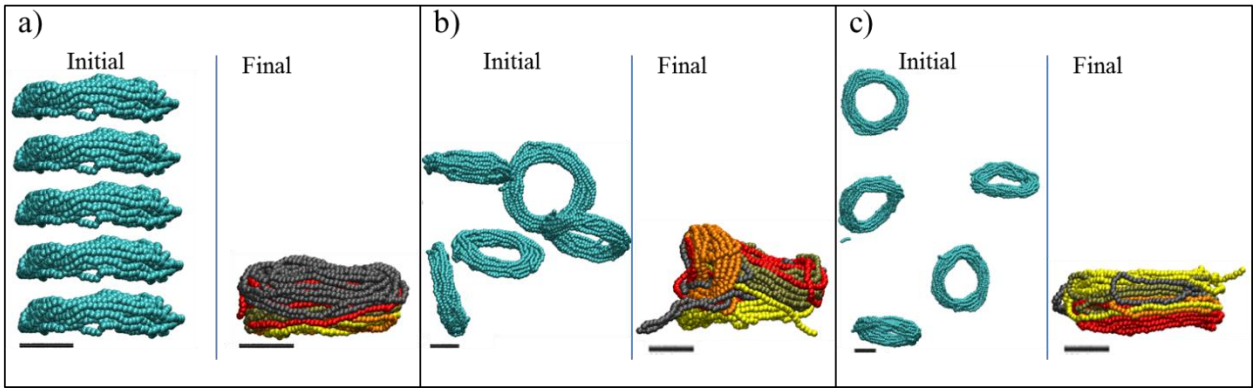


Figure 3.10**.** Initial and final snapshots from simulations of five replicated rings self-assembling into tubular structure. The ring is formed from a single heterogeneous MC chain with DP = 1000. All scale bars are 5nm. Three initial configurations were constructed. a) Replicates are placed on top of each other. b-c) Replicates are placed randomly in the simulation box at concentration of 3.9 and 1.3 weight percent respectively.

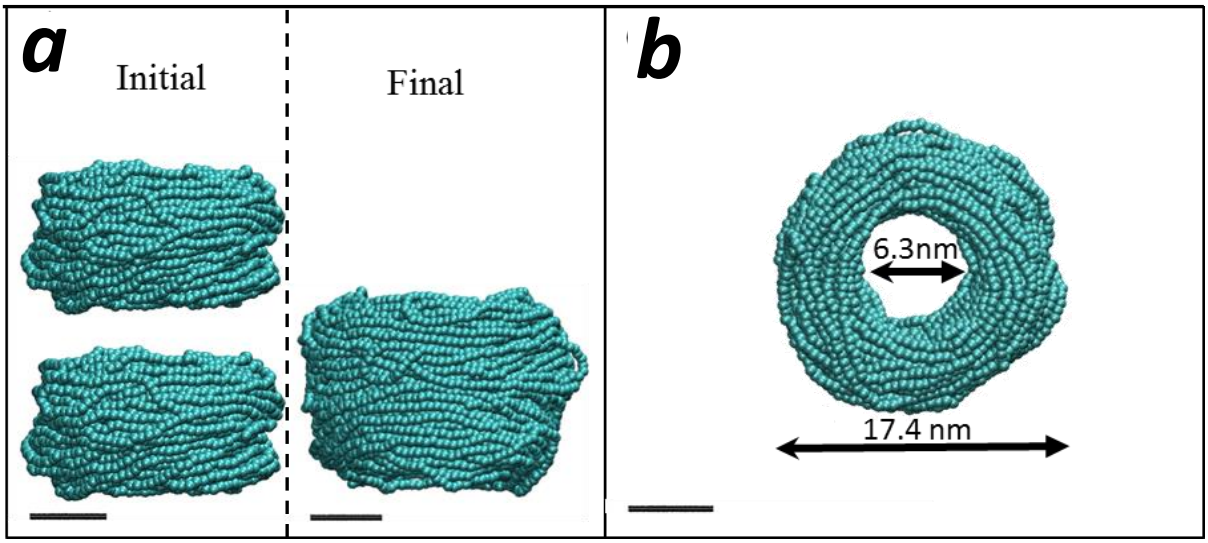


Figure 3.11. a) Two five-chain proto-tubes are brought together and equilibrated, forming a ten-chain tube (b) Cross-section and dimensions of a three-chain proto-tube. All scale bars in the figure are 5nm.

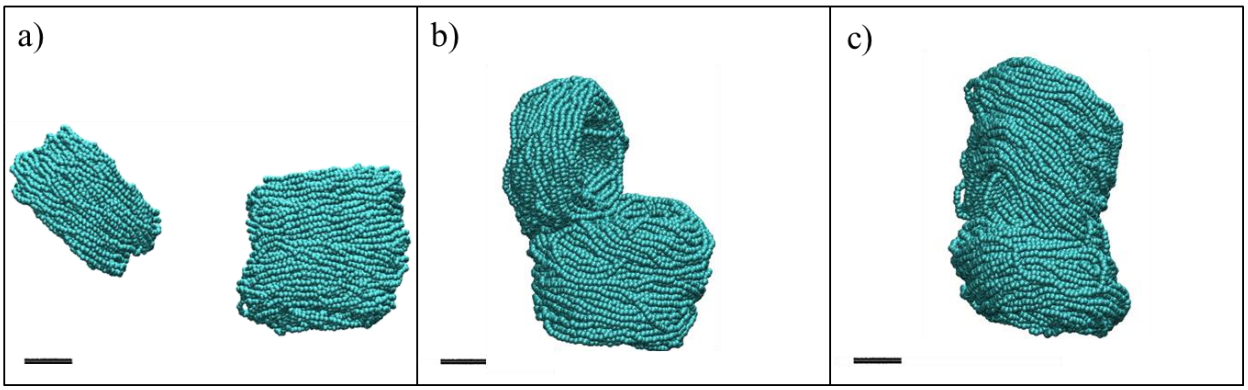


Figure 3.12. Snapshots of (a) initial, (b) metastable, and (c) final structures in the fifteen-chain tube structure growth simulations. All scale bars in the figure are 5nm.

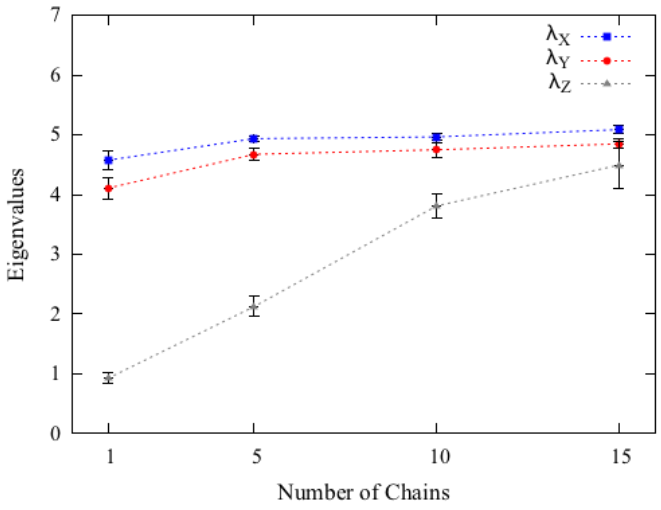


Figure 3.13**.** Three averaged eigenvalues (λ_x, λ_y, λ_z, in *nm*) of the radius of gyration tensor as a function of number of chains in the ring and tube structures.

Stability of the Tubular Structure under Cooling/Heating Cycle

Finally, we demonstrate that the formation of the tubular structure at elevated temperature can be reversed by cooling the ten-chain tubular structure obtained from simulations described in the previous section (Figure 3.14a). To do so, we lowered the system temperature to an arbitrary medium temperature, which is slightly below the gelation temperature. This is achieved by lower the ε_{ii} values derived at 50 ℃ by 20%, assuming ε_{ii} values scale linearly with the temperature. In response, the single ten-chain tubular structure broke into two smaller tubular structures with similar outer and inner diameters. Due to the effect of periodic boundary condition, the two proto-tubes remained in contact with the periodic images and therefore cannot move further apart (Figure 3.14b). This suggests that when temperature is lowered below the gelation temperature, segments of tubular structures will remain in solution and recombine when temperature is raised to above the gelation temperature again. Indeed, we recovered the original single ten-ring tube structure when we increased the system temperature back to 50℃, mimicking the cooling and heating cycle in a typical MC gelation experiment.

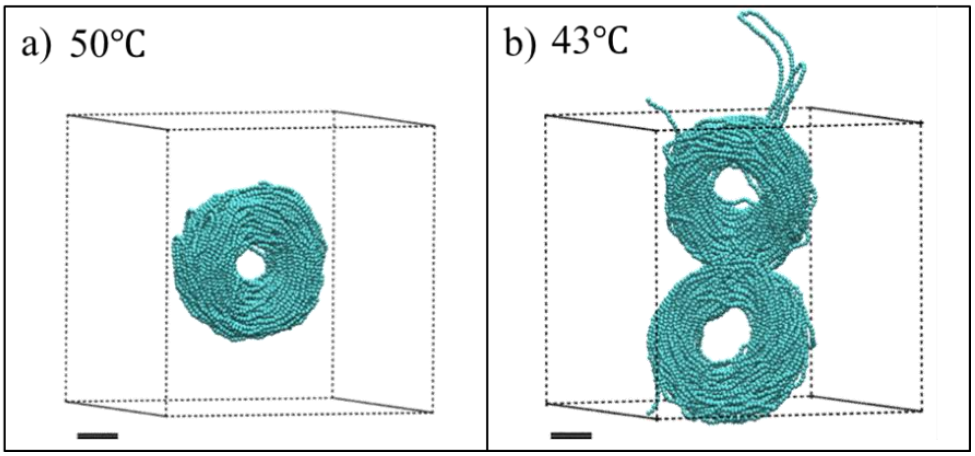


Figure 3.14. Snapshots of ten-chain tube structure at a) 50℃ and b) 43℃. All scale bars in the figure are 5nm.

3.4 Analytical Semi-Flexible Polymer Model

Broadly speaking, the ability to predict the collapsed state of a given semiflexible polymer chain based on its structural properties (e.g., monomer diameter, chain stiffness, polymer interaction with solvent, etc.) can be a powerful tool to design specialized polymer to form specific collapsed structures. In early work, Schnurr et al.[90] simulated single short stiff chains (2–3 Kuhn steps, N_k) using bead-spring "pearl necklace" chains and observed the chain evolve from an extended state to various collapsed states. More recently, Seaton et al.[97] reported the phase behavior of simulated 30-mer semiflexible bead-spring "pearl necklace" chains ($N_k < 10$) with various bending stiffnesses over a wide range dimensionless temperatures. With increasing chain stiffness, they observed globules at very small bending stiffness, to bundles of varying aspect ratios at higher bending stiffness, to tori, and to expanded coils at the highest bending stiffness. However, they considered only a single chain length and suggested that the phase diagram could change for much longer chains. In a recent simulation study, Kong et al.[92] detailed various collapsed states and collapse paths exhibited by semiflexible bead-spring polymers at different chain resolutions (i.e., number of beads per Kuhn length), self-attractive strengths, and chain diameters. A conformational phase diagram for polymer chains with a length of five Kuhn steps ($N_k = 5$) as a function of dimensionless self-attraction strength and ratio of chain diameter to Kuhn length was produced. In sum, simulation studies currently available in the literature are limited to bead-spring chains with rather short lengths. Because we can now simulate a single chain of length several hundreds of Kuhn steps, we would like to re-examine the collapsed phase behavior of semiflexible polymer chains over a wider range of chain lengths.

One important consideration in the polymer chain model is the choice of the bending potential. There are two typical bending potentials, namely the harmonic bending potential ($U_{\text{angle,h}}$) and the cosine bending potential ($U_{\text{angle,c}}$) (Equation 3.2-3.3). While Schnurr et al. and Kong et al. chose harmonic potential for their model and Seaton et al. chose cosine potential in their work, Stukan et al. have compared the effect of these two bending potentials on the stability of the torus and bundle. They have found that the choice of the bending potential can lead to different stability of the collapsed structures. For example, for large bending angles, such as the ones in the end fold, the cosine potential is "softer" than the harmonic potential, and therefore decreases the overall energy of the end fold. Because Stukan et al. only modeled the chains at low dimensionless temperature with ideal hexagonal packing, we would also like to re-evaluate the effect of the bending potentials on the conformational behavior at higher dimensionless temperature and over a wider range of dimensionless chain length range in this work.

$$U_{\text{angle,h}} = \frac{1}{2} K_{\theta}(\theta - \theta_0)^2 \tag{3.2}$$

$$U_{\text{angle,c}} = \frac{1}{2} K_{\theta}[1 - \cos(\theta - \theta_0)] \tag{3.3}$$

$$U_{\text{angle,s}} = 8U_{\text{angle,h}} - 14U_{\text{angle,c}} \tag{3.4}$$

Here, we develop a much simpler analytical model to predict the formation of torus or bundle states by representing these collapsed states as ideal geometries with specified dimensions, and corresponding surface areas, such as the areas of the folded bundle ends and sides. We then subsume the properties of the chain, namely the number of monomers, monomer diameter, and chain stiffness into surface free energies of both ends and sides, as well as the bending energy per unit length for a chain with a radius of curvature set by the radii of the torus. We then find the geometry of minimum free energy of both torus and bundle for a given set of parameters, and find whether the torus or the bundle has the lower minimum free energy. A phase diagram is produced by recording these minimum free energy conformations. To validate the theoretical derived phase diagram, we also simulate bead-spring "pearl-necklace" chains using Brownian Dynamics (BD) simulations with three different bending potentials. Specifically, in addition to the harmonic and cosine bending potential, we adopt a "stiff" potential ($U_{\text{angle,s}}$) using a linear combination of the two aforementioned potentials (Equation 3.4) to study the effect of bending

potential systematically. We show that the predicted phase diagram agrees with the BD simulation results qualitatively.

Model Detail

We consider three possible collapsed conformations for a single polymer chain, namely torus (T), bundle (B), and globule (G). The "pearl-necklace" polymer chain is modeled using beads connected by short stiff springs; each chain contains *N* beads of diameter σ, and the contour length of the chain is *L*. We introduce energy penalty parameters for exposed monomer surface (γ_s), end folds in the bundle (γ_e), and chain bending in the torus (γ_b), respectively. The former two parameters have units of free energy per unit area, while the latter has units of free energy per unit curvature squared per unit chain length. This yields three dimensionless ratios, namely $\gamma_b/\gamma_e\sigma^3$, γ_e/γ_s, and $L^* \equiv L/\sigma$. The other quantities with units of length are also scaled by σ to make them dimensionless. We approximate the stretched chain as a long and thin tube, with the volume of the chain V_c is made dimensionless as $V_c/\sigma^3 \equiv \pi L^*/4$. We focus on the two collapsed conformations, namely the torus and the bundle. For each of these, we estimate the free energy by summing up the contributions from the lateral surface (for the bundle and the torus), and from the areas of the two end caps (for the bundle), and from chain bending (for the torus). We then differentiate with respect to radius r^* ($\equiv r/\sigma$) for each structure (Figure 3.15) to obtain the lowest free energy state.

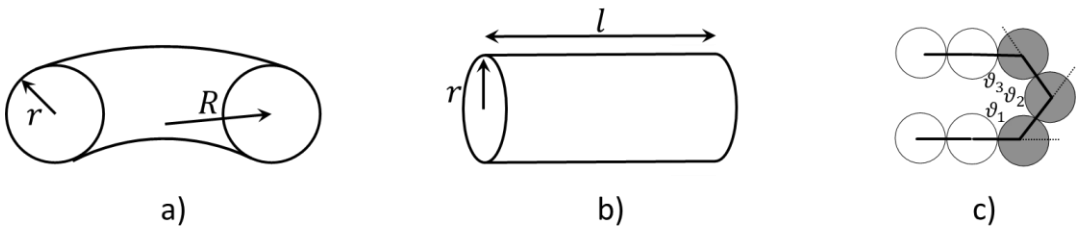


Figure 3.15. Schematics of torus (a) and bundle (b); (c) Schematic of a generic end fold we consider in this work, which has three exposed monomers (colored in gray). The angles in the end fold, ϑ_{1-3}, are formed among three consecutive monomer beads. For simplicity, these angles are taken to be 120°.

In the torus model, there are two dimensionless lengths, namely r_t^* ($\equiv r/\sigma$) and R^* ($\equiv R/\sigma$). As shown in Figure 3.15a, *r* is the radius of the cross section of the torus, and *R* is the distance from the center of the torus to the center of the cross section. We start by equating the volume of the free chain (V_c) to the volume of the torus (V_t) (Equation 3.5). We then express the free energy of the torus (G_t) as the sum of surface energy and bending energy, and derive the expression for

dimensionless free energy G_t^*. This free energy is made dimensionless with $\gamma_s\sigma^2$, which is comparable to the free energy of exposure of a single bead to solvent. We perform an analytical differentiation and obtain the value for r_t^* at which the free energy is minimized (note: the second derivative $d^2G_t^*/dr_t^{*2}$ is always positive) (Equation 3.6). We do not allow R^* to be smaller than $2r_t^*$, or the torus will self-intersect and the "donut hole" of the torus will disappear.

$$\frac{1}{4}\pi L^* = \pi r_t^{*2} 2\pi R^* \tag{3.5}$$

$$G_t = \gamma_s A_s + \gamma_b \frac{L}{R^2} \Longrightarrow r_t^* = \left(\frac{L^{*2}}{512\pi\frac{\gamma_b}{\gamma_s\sigma^3}}\right)^{\frac{1}{5}} \tag{3.6}$$

In the bundle model, we define the dimensionless radius r_b^* ($\equiv r/\sigma$) and bundle length l^* ($\equiv l/\sigma$), as shown in Figure 3.15b, where r is the radius of the bundle end cap. Again, we equate the volume of the free chain (V_c) to the volume of the bundle (V_b) (Equation 3.7). We then express the free energy of the bundle (G_b) as the sum of lateral surface energy and the end cap energy, and derive the expression for dimensionless G_b^*. Note that the surface energy due to the exposed monomers in the end caps is included in the first term. We differentiate the expression and obtain the value of r^* at which the free energy is minimized (note: the second derivative $d^2G_b^*/dr_b^{*2}$ is always positive) (Equation 3.8). We do not allow l^* to be smaller than $2r_b^*$ to maintain the ratio of the bundle diameter to its length below unity ($2r_b^*/l^* \leq 1$), so that it is truly a "bundle" and not a condensed globule or disk. If, on the other hand, the ratio $2r_b^*/l^*$ is greater or equal to 0.5, we consider the structure to be a globule.

$$\frac{1}{4}\pi L^* = \pi r_b^{*2} l^* \tag{3.7}$$

$$G_b = \gamma_s A_s + \gamma_e A_e \Longrightarrow r_b^* = \frac{1}{2}\left(\frac{L^*}{1+\frac{\gamma_e}{\gamma_s}}\right)^{\frac{1}{3}} \tag{3.8}$$

Defining $k_1 \equiv \gamma_e/\gamma_s$, and $k_2 \equiv \gamma_b/\gamma_e\sigma^3$, we derive the exact solutions for the boundaries between globule and bundle, and between bundle and torus in the asymptotic limit of large L^*. A bundle can be considered to be a globule when the ratio $2r_b^*/l^*$ is greater or equal to 0.5 (Equation 3.9). Because the shape of the globule, like that of a bundle, is solely determined by the ratio γ_e/γ_s, the exact solution to the boundary between globule and bundle is simply the horizontal line at which γ_e/γ_s equals 1 (Equation 3.11).

To derive the boundary between the bundle and the globule, we consider the constraint for globule geometry

$$0.5 = \frac{2r_b^*}{l^*} \tag{3.9}$$

Inserting l^* from Equation 3.8, r_b^* from Equation 3.9, and $k_1 \equiv \gamma_e/\gamma_s$ gives:

$$0.5 = \frac{\left(\frac{L^*}{1+k_1}\right)^{\frac{1}{3}}}{\frac{L^*}{4{r_b^*}^2}} \tag{3.10}$$

Inserting r_b^* from Equation 3.9 again and Equation 3.10 reduces to Equation 3.11, which leads to a trivial solution for k_1:

$$0.5 = \frac{1}{1+k_1}; \text{ thus } k_1 = 1 \tag{3.11}$$

Next, we derive the exact solution to the boundary between the torus and bundle phase, obtained by equating their free energies yields a relationship between k_2 and k_1 (Equation 3.12). At the limit of large k_1, Equation 3.12 reduces to a simple power law relationship $k_2 \sim {k_1}^{2/3}$ (Equation 3.14). We take the two constraints on the torus into consideration, namely a torus has to be non-self-intersecting, and the thickness of the cross section has to be more than one bead, which gives an upper and a lower bound on the k_1 value in the expression for phase boundary (Equation 3.15). As a result, k_1 must exceed 2.63, given by Equation 3.18. The upper bound on k_1, in Equation 3.18, is a function of the dimensionless chain length (L^*).

To derive the boundary between the torus and the bundle, we equate their optimized free energies

$$G_t^* = G_b^* \tag{3.12}$$

Using the expression for G_t^* from Equation 3.6 and G_b^* from Equation 3.8 gives:

$$\frac{1}{2}\pi L^* {r_t^*}^{-1} + k_1 k_2 \frac{64\pi^2}{L^*} {r_t^*}^4 = \frac{1}{2}\pi L^* {r_b^*}^{-1} + 2\pi {r_b^*}^2 + k_1 2\pi {r_b^*}^2 \tag{3.13}$$

Inserting r_t^* and r_b^*, then gives:

$$k_2 = \alpha^5 {L^*}^{\frac{1}{3}} (8 + 8k_1)^{\frac{3}{5}} {k_1}^{-1}$$

where:

$$\alpha = \frac{\frac{6}{5}}{(512\pi)^{\frac{1}{5}}} \tag{3.14}$$

Equation 3.14 is thus the boundary between torus and bundle phase. For large k_1, Equation 3.14 reduces to:

$$k_2 = 32\alpha^5 L^{*\frac{1}{3}} k_1^{\frac{2}{3}} \tag{3.15}$$

Next, we consider the constraints on the torus geometry:

$$\frac{1}{2} \leq r_t^* \leq \frac{R^*}{2} \tag{3.16}$$

Inserting r_t^* and R^* gives:

$$\frac{1}{2} \leq \frac{L^{*\frac{2}{5}}}{(512\pi)^{\frac{1}{5}}(k_1 k_2)^{\frac{1}{5}}} \leq \frac{L^{*\frac{1}{3}}}{(16\pi)^{\frac{1}{3}}} \tag{3.17}$$

Inserting k_2, we obtain the upper and lower limit for the k_1 value, due to the constraints on the torus geometry:

$$2.63 \leq k_1 \leq 0.58L^* - 1 \tag{3.18}$$

Because our theory does not consider fluctuations on the scale of the structure (torus, bundle, or globule), and also neglects local bead ordering, the range of temperatures for which our model is valid must be such that the temperature is high relative to the energy of interactions of individual beads, which is of order ε, but must be very low relative to the total interaction energy of the entire molecule, which is of order N ε. For both of these conditions to hold, our model is restricted to long chains (i.e., L^* > 100). Because we set the dimensionless temperature ($T^* = k_B T/\varepsilon$) to be unity, our theory is expected to hold at dimensionless temperatures between 1 and << N.

Simulation Details

The Brownian Dynamics (BD) simulations were performed using the LAMMPS simulation package. Harmonic bond and interactions were used to model the bonded interactions, while the Lennard-Jones 12-6 potential was used to model the non-bonded interaction. We considered three bending potentials in the simulations, namely harmonic, cosine, and stiff bending potential (Equation 3.2-3.4). By design, the stiff bending potential produces a progressively higher energy penalty than the harmonic bending potential as the bending angle increases. Therefore, given the

same generic end fold, such as the one shown in Figure 3.15c, the cosine potential gives the smallest energy penalty, followed by harmonic and the stiff potential. The bead diameter (σ) in simulation is the same as the bead diameter (σ) in our theory. In the simulations, we hold the interaction strength (ε) and bead diameter (σ) constant at 1.0, and vary the bending potential coefficient (K_θ) from 1.5 to 150. The value of persistence length (l_P) scaled by length unit (σ), when using harmonic potential, is slightly larger than the value of K_θ scaled by energy unit (ε). (i.e., K_θ/ε of 1.5 gives l_P/σ roughly 2).

We can map the energy terms used in the analytical model, namely γ_s, γ_e, and γ_b, to the parameters used in the simulations. γ_s is the free energy penalty a bead pays for being exposed to the solvent. We approximate this free energy by the LJ potential energy of an exposed bead. Beads that interact through an LJ potential in the dense state will pack closely together with numbers of neighbors not too far from 12, that of a face-centered cubic (FCC) with interaction strength of 1ε per contact, which is the depth of the potential well. The contribution per bead in the pair is 0.5ε per contact, and the total potential energy for a non-exposed bead is thus 6ε. Here, we consider an exposed bead loses potential energy contribution from 8 neighboring beads, loosely based on the structure we observed in the simulations. Therefore, γ_s is approximated to be the potential energy difference (4ε) divided by the cross sectional area of the bead ($\pi\sigma^2/4$). γ_b is the product of bending coefficient (K_θ) and bead diameter (σ) (Equation (5a–b)). We consider a generic end fold with three exposed monomer beads (Figure 3.15c). For simplicity, we take all three angles in the end fold to be 120 degrees. Therefore, γ_e is the ratio between energy of the fold (E_{fold}), calculated using the three bending potentials respectively, and the approximated exposed end fold surface area of the three monomer beads ($3\sigma^2$) (Equation (5c)). The above parameters for mapping are very rough approximations; in general, the structure of the fold and beads packing will depend on the parameters of the model and the dimensionless temperature. However, to maintain the simplicity and generality of the model, we avoid these complexities. The more detailed model of Stukan et al. [101] considered in detail the structure of the fold and beads packing, as reviewed above. Particularly, their model requires re-analysis for each specific end fold conformation, while our simple model should be quite general, although qualitative. In typical simulations, including those discussed here, both end fold and bending energies arise from the same bending potential; the fold is simply an extreme bend, whose energy is set by the same potential, but with smaller radius of curvature. Therefore, the

dimensionless ratio $\gamma_b/\gamma_e\sigma^3$ is a constant for a given bending potential. The three bending potentials that we considered allow us to access various $\gamma_b/\gamma_e\sigma^3$ ratios and validate the trend predicted by the theory.

$$\gamma_s = \frac{16\varepsilon}{\pi\sigma^2} \tag{3.19}$$

$$\gamma_b = K_\theta d \tag{3.20}$$

$$\gamma_e = \frac{E_{fold}}{A_{fold}} = \frac{E_{fold}}{3\sigma^2}$$

$$E_{fold} = 3U_{angle}(120°) \tag{3.21}$$

3.5 Analytical Model Prediction and Comparison with Simulation Results

Within our model, the collapsed state of a single polymer chain is controlled by three dimensionless quantities, namely $\gamma_b/\gamma_e\sigma^3$, γ_e/γ_s, and $L^* = L/\sigma$. A fourth dimensionless parameter, the dimensionless temperature, k_BT/ε, influences when the chain will remain a random coil, rather than collapsing, but we consider long chains here at moderate values of k_BT/ε, and this parameter therefore does not enter our theory. On the "phase diagram" shown in Figure 3.16, the y-axis is the ratio of the end fold energy to the surface energy, while the x axis is the ratio of the bending energy to the product of end fold energy and bead diameter cubed. We vary the two dimensionless ratios over a range of values to determine a "phase diagram" of lowest energy conformation. Specifically, for each γ_e/γ_s and $\gamma_b/\gamma_e\sigma^3$ pair, we compute the resulting minimized free energies for all three conformations, and record the lowest energy conformation on the phase diagram. The phase diagram for L^* = 600 is shown in Figure 3.16, which includes regions for each of the three phases that we modeled for a polymer chain. The globule (G) conformation is observed when surface energy penalty dominates (small γ_e/γ_s values; see inset V in Figure 3.16). Chains in this region can be regarded as flexible. The chain adopts either torus or bundle configuration when it has moderate bending and end fold energy penalties. Specifically, when the end surface energy penalty is high but the bending energy penalty is low, the polymer chain collapses into a torus (T). The aspect ratio of this torus (R^*/r^*) increases as the bending energy increases; the cross section of the torus becomes thinner and its radius increases, which helps the chain to reduce the high bending energy penalty. This can be seen from the insets I and II in Figure 3.16. On the other hand, when the end fold energy penalty is low but bending energy penalty is high, a bundle (B) is formed. The aspect ratio of the bundle

(*l*/r**) increases as the end fold energy term increases; the bundle becomes thinner and longer, which helps the chain to reduce the end fold energy at an expense of having more beads exposed along its side. Schematics of bundles with different aspect ratios are in the insets III and IV in Figure 3.16. We mark the exact solutions to the boundaries between collapsed phases, derived in the theory section on Figure 3.16. The boundary between globule and bundle (Equation 3.12) is a horizontal line at $\gamma_e/\gamma_s = 1$. We note that the upper bound of the boundary between torus and bundle that we present in Equation 3.19, namely $\gamma_e/\gamma_s \leq 0.58L^* - 1$, is not visible, because the value of γ_e/γ_s at the upper bound exceeds the x-axis range. For a small $\gamma_b/\gamma_e\sigma^3$ (*x*-axis value), even though the minimum-free-energy torus is self-intersecting at the given γ_e/γ_s and $\gamma_b/\gamma_e\sigma^3$ pair, a torus with minimum possible *R*/r** value of 2 can still form and its free energy remains lower than that of a bundle. This results in the horizontal boundary between bundle and torus phase on the phase diagram at small $\gamma_b/\gamma_e\sigma^3$ values. From the simulation trajectories, we observed conformations that fluctuate between two collapsed states, namely torus and bundle. We therefore assume that a chain can fluctuate between two states (i.e., T&B) if the calculated free energy of one state is less than 5% different from that of the other state, and add a transition phase (T&B) on the theoretical phase diagram. The resulting phase diagram captures both the predictions based on theory and the observations from the simulations. Note we expect to see a chain adopting a random coil (RC) conformation when both end fold energy and bending energy penalties dominate (large γ_e/γ_s and $\gamma_b/\gamma_e\sigma^3$ values), despite its large solvent-exposed surface area. However, because it is challenging to model the entropic contribution in the free energy of the random coil state to a reasonable accuracy, we are not presenting an analytical model for this state in the current work.

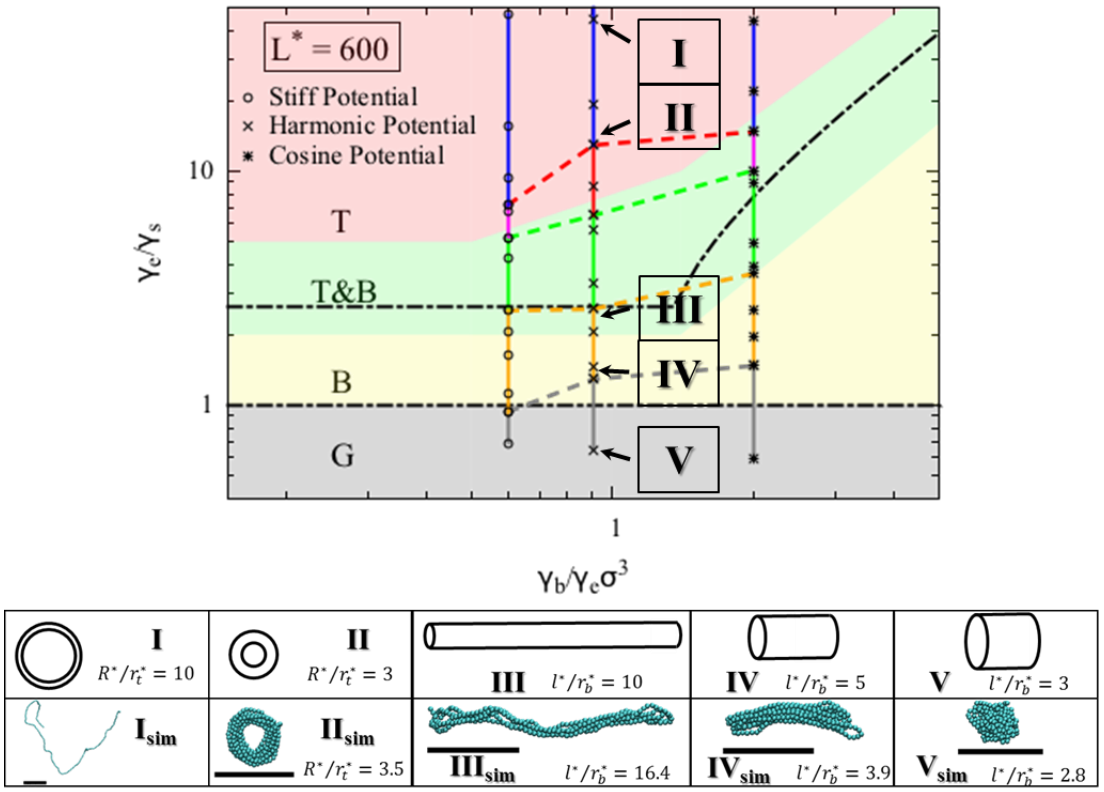


Figure 3.16**.** Phase diagrams for polymer chain with dimensionless chain length $L^* = L/\sigma = 600$. Each symbol on the phase diagram marks the parameter values at which a simulation was conducted. Specifically, open circles, crosses, and asterisk symbols are the simulation runs using stiff, harmonic, and cosine bending potentials respectively. A vertical bar connects simulations data producing the same type of structure. The color of the vertical bars, of the dashed lines, and of the regions of the diagram producing this structure coded as follows: blue—random coil (RC), pink—fluctuating between random coil and torus (RC&T), red—torus (T), green—fluctuating between torus and bundle (T&B), yellow—bundle (B), and black- globule (G). Note the pink and blue regions only occur in simulations because we do not model random coil phase. We regard the simulated T&RC phase simply as a torus phase (T) because we do not model the random coil (RC) phase analytically. The dashed lines connecting the vertical bars mark the upper boundaries of the phases obtained from the simulations (e.g., simulation results between yellow and grey dashed lines resulted in a bundle as the final collapsed structure). Exact solutions for the boundaries between the three collapsed phases are shown on the phase diagram (black dash-dotted lines). We have selected representative conformations in each phase, pointed by the arrows. We have calculated the theoretical aspect ratios for these conformations, depicted in the insets I–V. We have also conducted simulations using harmonic bending potential under the same conditions, and the final snapshots from the simulations are supplied in additional insets I–V with subscript "sim".

We compare the collapsed conformations predicted by the theory with the ones obtained from the simulations. As we noted earlier, because the bending and the end fold are modeled using the same energy potential, the ratio $\gamma_b/\gamma_e\sigma^3$ is constant for a given bending potential in our simulations. We can map the K_θ and ε values onto the phase diagram using Equation (9a–c), given a generic end fold (Figure 3.15c). From the simulations that use the harmonic bending potential, we observe various collapsed states as γ_e/γ_s value increases (Figure 3.16 inset I_{sim}

through V_{sim}). Specifically, for γ_e/γ_s values less than 1.3, we observe a globule, and for γ_e/γ_s between 1.3 and 2.7, we observe bundles with various aspect ratios. For γ_e/γ_s greater than 6.5, tori with various aspect ratios form. In between 2.7 and 6.5, we observe fluctuations between bundle and torus. As expected, simulations show random coils form at large γ_e/γ_s, here found to be a value greater than 13.

We compared the simulation results obtained using different bending potentials. The simulations show transitions from collapsed states to random coil state for all three bending potentials. The transitions from bundle (B) to fluctuation between bundle and torus (T&B) occur at very similar γ_e/γ_s values for simulations using stiff and harmonic potentials, and at a higher γ_e/γ_s value for simulations using the cosine potential. This trend is in agreement with theoretical prediction. The transition from fluctuation between bundle and torus (T&B) to torus (T) occurs at a higher γ_e/γ_s value as $\gamma_b/\gamma_e\sigma^3$ increases. This shows that a chain modeled with a "softer" bending potential that grows less steeply with increasing bending angle is more likely to form a bundle with large aspect ratio. We have marked the upper boundaries of the phases obtained from the simulation results in Figure 3.16 (i.e., chain in the simulation with mapped γ_e/γ_s value between yellow and grey dashed line will result in a bundle as the final collapsed structure). Because we do not model the random coil phase analytically, we simply regard the phase where chain fluctuates between torus and random coil (T&RC) as the torus phase (T). These boundaries obtained from simulations agree qualitatively with the ones predicted by the theory. We note that a globule is merely a very short bundle; therefore, it is often hard to distinguish between a short bundle and a globule from the simulation results, contributing to the deviation on the boundaries between globule and bundle obtained from simulation and theory.

3.6 Concluding Remark

In this chapter, we present simulation results of both homogenous and heterogeneous methylcellulose oligomers using both atomistic and coarse-grained (CG) force fields. The atomistic simulations reveal that the aggregation of the short oligomer chains is driven by the hydrophobic interaction. We demonstrated that the CG force field, which we derived based on the radial distribution functions generated from atomistic simulations, is capable of distinguishing the effect of monomer substitution type and temperature on polymer conformation. The heterogeneous methylcellulose chain, with methylation content chosen to match that of the

commercial METHOCEL^{TM} A product, has a persistence length of 9nm at room temperature and collapses into a ring structure for chain lengths of 600 monomers or more at elevated temperature. The ring structure formed has an outer diameter of 14nm and a void fraction of 26% and appears to be a precursor to the methylcellulose fibrils that compose the gel observed in experiments. We then also showed that individual MC rings subsequently self-assemble into stacks or "proto-fibrils" that sometimes gives rise to Y-junctions. These simulation results clearly support the experimental observation of the methylcellulose gel morphology and complement the theoretical work of methylcellulose gel formation mechanism.

In addition, we presented a simplified continuous analytical model to predict the collapse conformation of a single self-attractive semiflexible polymer chain in solution. We produced from this theory phase diagrams at various dimensionless chain lengths (L^*) by varying the ratios between three energy parameters, namely the solvent-water surface energy (γ_s), the energy of bundle end fold (γ_e), and the bending energy per unit length in a torus (γ_b). Three phases were modelled in this work – torus, bundle, and globule. We showed that a good qualitative agreement between theoretical and simulated results can be achieved at long chain length for transitions between collapsed states. Combined with the computational work presented in this chapter, our approach is useful for obtaining quick estimates of the collapsed state of a given polymer chain, and for designing polymer chemistries with controlled transitions between collapsed states.

Chapter 4: Modeling Hydroxypropyl Methylcellulose Acetate Succinate – Phenytoin Solid Dispersion Formulation

Some of the materials in this chapter are results of a collaborative work with Dr. Taraknath Mandal.

4.1 Introduction

Currently, majority of the new drugs candidates, or active pharmaceutical ingredients (APIs), in the research and development pipeline are estimated to have limited solubility in water, which in turn result in poor bioavailability when administrated orally.[3,4] Among many techniques that have been explored to promote the solubility or to maintain the supersaturation of the APIs, including complexation,[5,6] particle size reduction.[7] Amorphous solid dispersion[8–11], where API molecules in their amorphous form are mixed with polymer excipients, has stand out to be a very promising API solubility enhancement mechanism.

Hydroxypropyl methylcellulose acetate succinate (HPMCAS) has been identified as one of the most effective polymer excipients for solid dispersion formulation.[9,16,17] Each D-glucose monomer unit (Figure 1.1) in cellulose contains three substitution positions that allow functional groups with various sizes and hydrophobicity levels to be attached. The polyfunctionality thus allows HPMCAS-based solid dispersion formulations to maintain drug supersaturation for prolonged periods of time and to inhibit drug recrystallization. For example, the hydrophobic acetyl group stabilizes the hydrophobic drug molecules in the matrix, the unsubstituted groups allow hydration of the matrix upon solvation, and the succinyl group is ionized at pH level of 7 to provide colloidal stability.[9] However, because of the low precision in controlling the substitution pattern when manufacturing HPMCAS, it is very hard for researchers to understand precisely the effect of each functional group, and it is even more challenging to rationalize the design rules for optimizing the performance of HPMCAS for a given new drug candidate. To tackle this problem, Ting *et al*.[18] have synthesized analog polymers of the HPMCAS and studied their drug dissolution performance. They have concluded from this that tuning the acetyl-succinyl ratio in HPMCAS is useful to promote the solubility of drugs that are susceptible to

phase-separation, such as probucol, while the unsubstituted hydroxyl group plays an important role promoting the solubility of drugs that are strong crystallizers, such as danazol and phenytoin. Yin *et al.*[19] synthesized a new class of cellulosic polymer derivatives containing hydrophobic, hydrophilic, and pH-responsive functionalities. They demonstrated that a functional group that contains thioethers and weak electron-withdrawing groups can effectively inhibit nucleation of drug crystals, therefore outperforming the drug dissolution performance of HPMCAS. Ueda *et al.*[102] varied the percentage of succinyl substituent in HPMCAS and concluded that the lower the succinyl substituent percentage in HPMCAS, the more effective was the suppression of drug crystallization. An atomistic molecular dynamics study of Jha *et al.*[40] found that acetyl groups in HPMCAS have stronger interactions with phenytoin (as measured by radial distribution functions) than do protonated succinyl groups, while the deprotonated succinyl (at pH=7) only interacts with phenytoin molecule weakly. These simulation results support the previous understanding of the effect of each functional group in HPMCAS.[9]

Jha *et al.*[40] and Xiang *et al.*[41] have used molecular dynamics (MD) to model HPMCAS short oligomer melts. These studies are limited by the size of the system (~10nm) and the simulation time (up to 100ns), thus cannot be readily compared to any experimental dissolution study. We developed a coarse-grained force field based on the structural information from atomistic-level simulations, and allowed us to simulate multiple methylcellulose chains up to 1000 monomers long. Very recently, Mandal *et al.*[83] presented a coarse-grained force field with a similar level of resolution for phenytoin drug molecules that is capable of capturing phenytoin crystal growth. Combining these coarse-grained schemes and advanced high performance computing resources, it has now become feasible to simulate HPMCAS drug nanostructures with diameter between 20-100nm, which is beyond the range of atomistic simulations.

4.2 Simulation Results for HPMCAS

We have developed the CG force field for HPMCAS based on the atomistic simulations of homogenous 20-mer oligomers. We consider five ten different monomer substitution types (Table 2.5). Our goal is to apply this force field to simulate longer polymers chains that are present in the HPMCAS solid dispersion nanostructure. The typical molecular weight of HPMCAS used in solid dispersion formulations is around 20 kg/mol,[19,102] corresponding to a chain length around 50-100 monomers. We therefore conducted multiple homogenous CG chain

simulations with both 50-mer and 100-mer chains of each monomer substitution type. To keep the polymer mass concentration consistent at around 3 wt%, the 50-mer systems contained twice as many chains as the 100-mers systems. In Figure 4.1 we show snapshots from both atomistic and CG simulations for five different monomer substitution types. As hoped, the structures formed in the CG simulations of 20-mer oligomers resemble their atomistic counterparts well. Oligomers with HPAc, Ac, and Su substitution groups form loose aggregates, and the oligomers with the SuDP substitution group (i.e., deprotonated, or charged, succinate) do not aggregate. 2,6-Me oligomers, on the other hand, form aligned bundles due to the lack of a bulky side chain, with only one bead used to model each CG monomer.

Polymers with HPAc, Ac, and Su substitution groups form chain-length-dependent aggregates. In the 50-mer systems, these polymers form multiple aggregated bundles. However, these bundles do not form a network or merge into a single long bundle during the simulation time of 50μs. In the 100-mer systems, these polymers form a bundled network, presumably because long chains are more likely to interact with neighbor chains and form a connected network. The aggregation behaviors of polymers with SuDP and with only Me substitution groups are less sensitive to chain length. In particular, polymers with SuDP do not aggregate in either the 50-mer or 100-mer systems, while both 50-mer and 100-mer 2,6-Me polymers form a single bundle that bends into a ring structure. The ring structures observed in these simulations are not the typical ring structures that are believed to be the precursor to the methylcellulose fibrillar gel, as we described in the section 3.3. This is because 1) the length of the chains in the simulations presented here is too short for a single chain to form a self-associated ring structure and 2) we showed the ring formation can be only captured in the simulations at low concentration (< 1wt%), much lower than the concentration of the system simulated here. Comparing the bundle networks formed by 100-mer chains with different substitution groups, we find the packings of the bundles are affected by the size of the monomer substitution type. Functional groups HPAc, Ac, and Su are relatively bulky and therefore prevent the backbone beads from forming closely packed aligned bundles. Highly methylated homogenous methylcellulose chains, with two or more methyl groups attached to each monomer, have strong intermolecular interaction,[103] which, along with the small size of the methyl group, explains the observed tight packing of 2,6-Me chains into a bundle.

It is worth noting that in our CG model developed for commercial methylcellulose (section 2.5), we adopted chain length dependent intermolecular interaction parameter ε_{ii} for each monomer substitution type in order to correctly reflect the aggregation behavior for a long chain (up to 1000-mer). In particular, a decreasing exponential function ($\varepsilon_{ii} = AN^{-B}$ where N is the chain length and A,B are fitting parameters) was used to extrapolate the ε_{ii} values obtained from short chain (< 20-mer) atomistic simulations to the values for long chains. This resulted in a less than 20% drop between the ε_{ii} values obtained from 20-mer reference simulations and those extrapolated to 100-mer for all the methylcellulose monomer substitution types, including 2,6-Me and 2,3,6-Me. In this work, we are considering chains of length 50 to 100 monomers, and therefore expect the ε_{ii} values of all bead types to follow a similar chain length dependence. To test the sensitivity of the final structure on the ε_{ii} value, we conducted a simulation of 2,6-Me 100-mer system with the value of ε_{ii} decreased by 20%. The resulted aggregated structure is similar to the corresponding structure presented in Figure 4.1. Because we do not expect to observe significant difference in aggregation behavior due to the less than 20% deviation in ε_{ii} values, we are not considering the effect of chain length dependent ε_{ii} value in this study.

We quantify the compactness of the network structures formed by different monomer substitution types using the intermolecular CG RDFs, shown in Figure 4.2. The rank order for the five monomer substitution types, based on the height of the first peak, remains the same as the polymer chain length increases. Specifically, 2,6-Me has the highest first-peak height, shown on the 10-based log scale, followed by polymers with Ac, Su, HPAc, and SuDP functional groups. These results are consistent with the visual inspections of the tightness of packing of chains from the simulation snapshots. To test whether these CG polymer systems reach equilibrium after 50μs, we computed intermolecular CG RDF over every 10μs interval. We found that the RDFs converge well after 30μs and therefore we conclude our simulation time is sufficient for the system to reach equilibrium, or at least a relatively stable non-equilibrium structure.

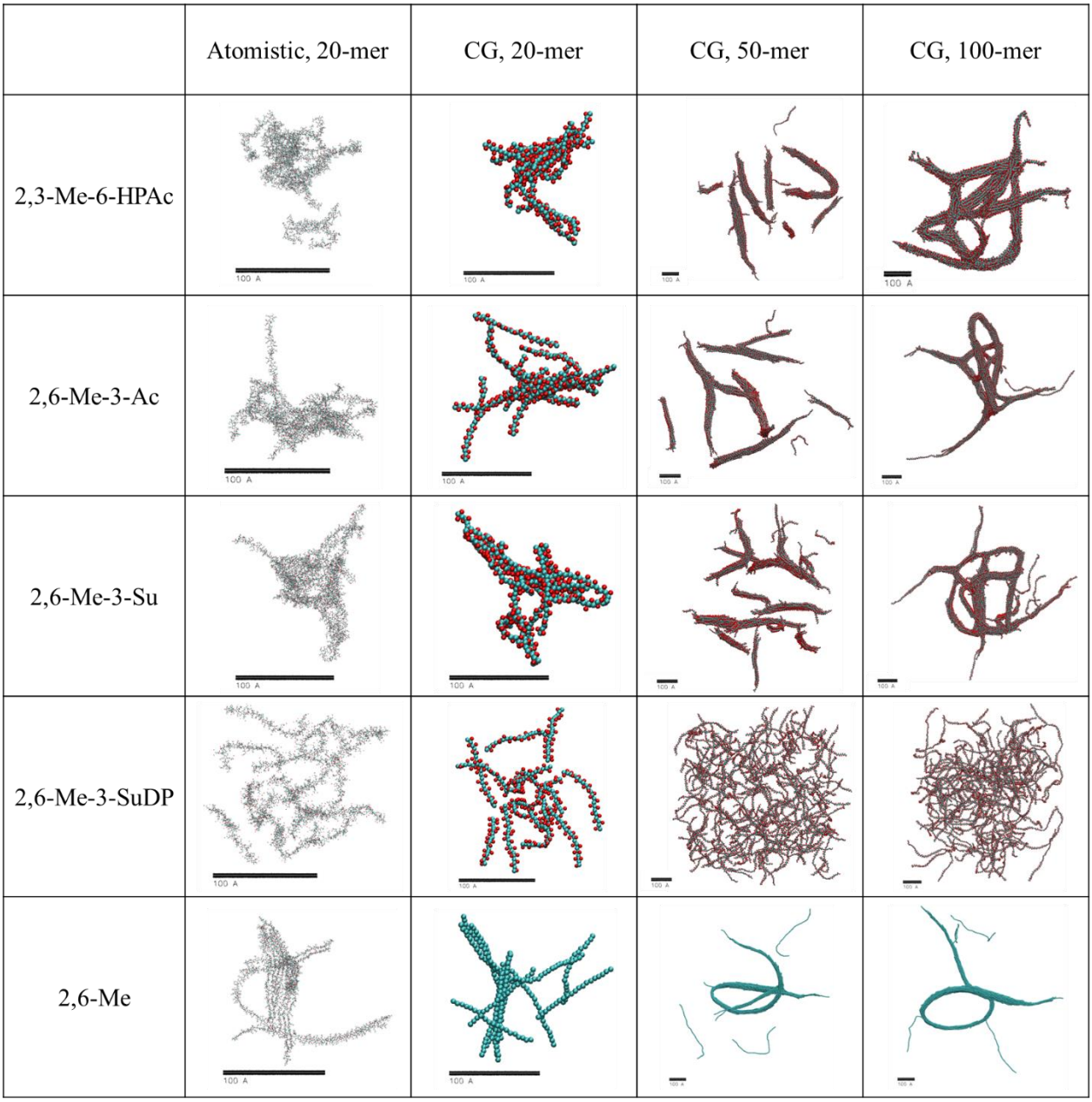


Figure 4.1: Snapshots from atomistic and CG simulations. The monomer substitution types of the homogenous oligomers and polymers simulated are listed as the row titles. Snapshots from the atomistic simulations are shown in the first column, and the snapshot from the CG simulations of 20-mer, 50-mer, and 100-mers are shown in the second, third, and fourth columns respectively. For the 20-mer systems, 15 chains were randomly placed in the cubic simulation box of 12nm on the side, leading to a concentration of 10wt%. For the 50-mer and 100-mer systems, 70 chains and 35 chains, respectively, were placed in the cubic simulation box of 35nm on the side, producing a concentration of 3wt%.

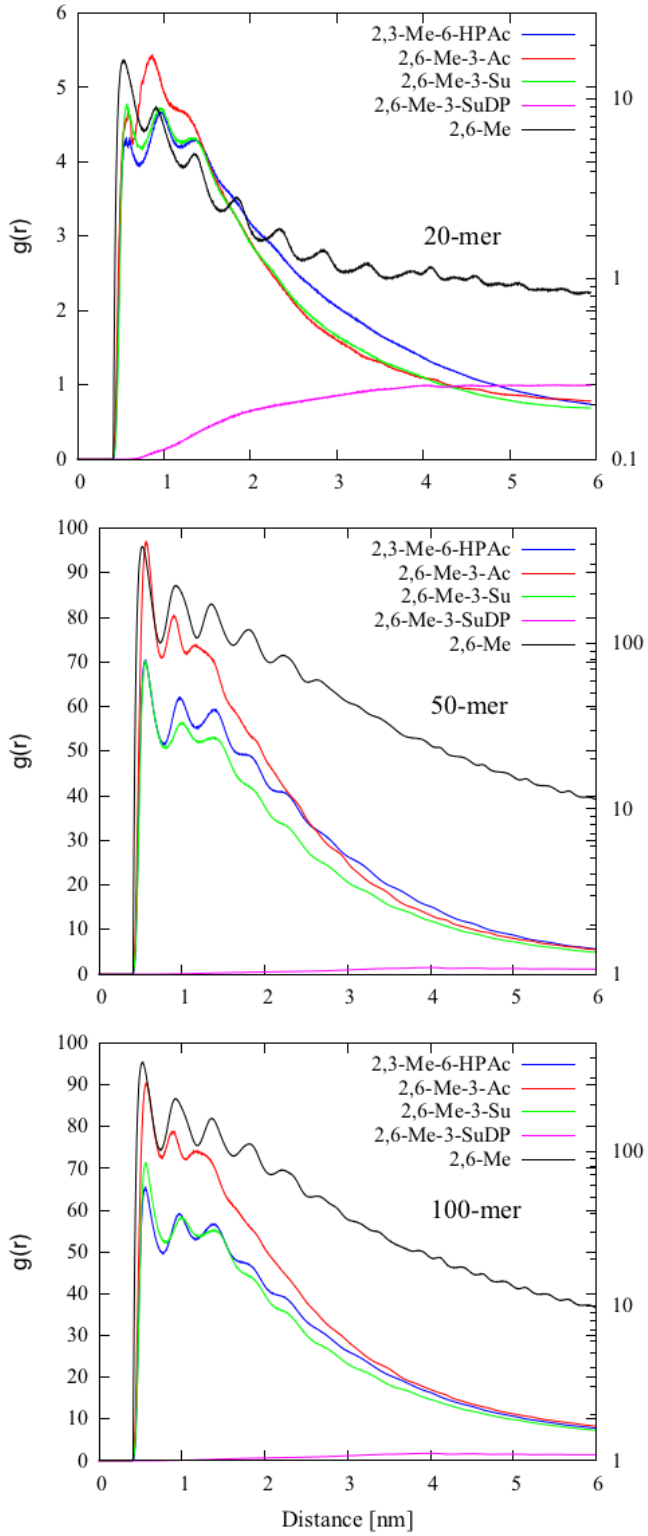


Figure 4.2: Intermolecular CG RDFs obtained from CG simulations of multiple 20-mer, 50-mer, and 100-mer homogenous HPMCAS model oligomers. The RDFs for 2,6-Me (black) are presented on the 10-based log scale on the right

4.3 Simulation of HPMCAS and Phenytoin

We now turn our attention to the interaction between HPMCAS and amorphous phenytoin. In particular, we want to compare the simulations results between the two phenytoin CG force fields, namely the aggregation-based and the crystal-based force fields. The crystal-based phenytoin CG force field[83] uses tabulated potentials to model the non-bonded interactions. Therefore, in simulations of HPMCAS homopolymers and amorphous phenytoin molecules, all interaction terms reported in this work, including the interactions between the polymer beads and cross interactions between the polymer beads and amorphous phenytoin beads, need to be converted to tabulated forms by explicitly computing the values of the potentials at distances spaced 0.002nm apart. However, despite using two different potential forms for phenytoin's interactions with itself, the interactions between polymer and amorphous phenytoin are taken to be identical for the two phenytoin-phenytoin interaction potentials. Readers are referred to ref [83] for more detailed implementation of the tabulated potentials. 10 tabulated potentials are required for each HPMCAS homopolymer/phenytoin system, as there are four different bead types; two for the drug and other two for the polymer, which is computationally fairly intensive. We have employed both force fields to model interactions between amorphous phenytoin and 2,6-Me-3-Su and 2,6-Me-3SuDP homopolymers, with drug and polymer molecules initially randomly dissolved in the box, and we show the final snapshots in Figure 4.3. The CG simulations of 20-mer oligomers and phenytoin molecules resemble their atomistic counterparts well. Phenytoin molecules have strong interactions with 2,6-Me-3-Su oligomers and as a result, an oligomer-drug complex is formed. However, phenytoin molecules interact only weakly with 2,6-Me-3-SuDP oligomers, which completely dissolve in solution. Therefore, no oligomer-drug complex is formed and phenytoin molecules form phenytoin-only aggregates, similar to those formed in phenytoin-only systems.

For 50-mer polymer chains, the snapshots are very similar between the simulations conducted using the two phenytoin force fields. Phenytoin molecules form a polymer-drug complex with 2,6-Me-3-Su polymer chains, wherein phenytoin molecules are mostly entrapped within the polymer capsule. Qualitatively, strong interactions between polymer and drug will block drug aggregation and thereby slow down the drug nucleation and/or crystal growth rate. On the other hand, phenytoin molecules interact weakly with 2,6-Me-3-SuDP polymer chains and form many small aggregates. The CG polymer-drug RDFs for the 50-mer system are shown in Figure 4.5.

The RDFs generated from the simulations using both CG force fields are very similar. The noticeable difference is that the 2,6-Me-3-SuDP system with the crystal-based CG force field gives a moderately higher RDF than does the RDF computed using the aggregate-based CG force field. This can be seen from the snapshots with free phenytoin molecules scattered in the polymer matrix in the system simulated using the aggregate-based CG force field, while all phenytoin molecules have aggregated into clusters in the simulations using the crystal-based CG force field. This discrepancy is likely due to the different atomistic reference structures used during the parameterization of the two force fields. Note that the qualitative aggregation behaviors of the drugs in the presence of polymers with two different drug force fields are similar as in both cases drugs are in the amorphous state and no nucleation of drugs occurred in our simulation. In fact, our run times are not long enough to see nucleation events, even in the absence of the polymer. Thus, the difference between the two force fields is minor for the purpose of our work, which is to study the polymer-drug interactions in the HPMCAS-amorphous phenytoin solid dispersion formulation at accessible simulation times. Since our simulations cannot reach time scales necessary to see nucleation events, any differences in behavior predicted by the two force fields is unlikely to emerge in our simulations. Therefore, in what follows, we will use the cheaper aggregate-based phenytoin CG force field. We note that, experimentally, if the excipient acts to block nucleation primarily by inhibiting drug aggregation, so that crystal nuclei have no chance to form, then either of our CG force fields might be adequate to predict the relevant excipient-drug interactions. If, on the other hand, the polymer interferes in some more detailed way in the mechanism of nucleation, then our aggregation-based force field will prove to be inadequate. At this stage, we are unable to resolve which of these scenarios is the more realistic one.

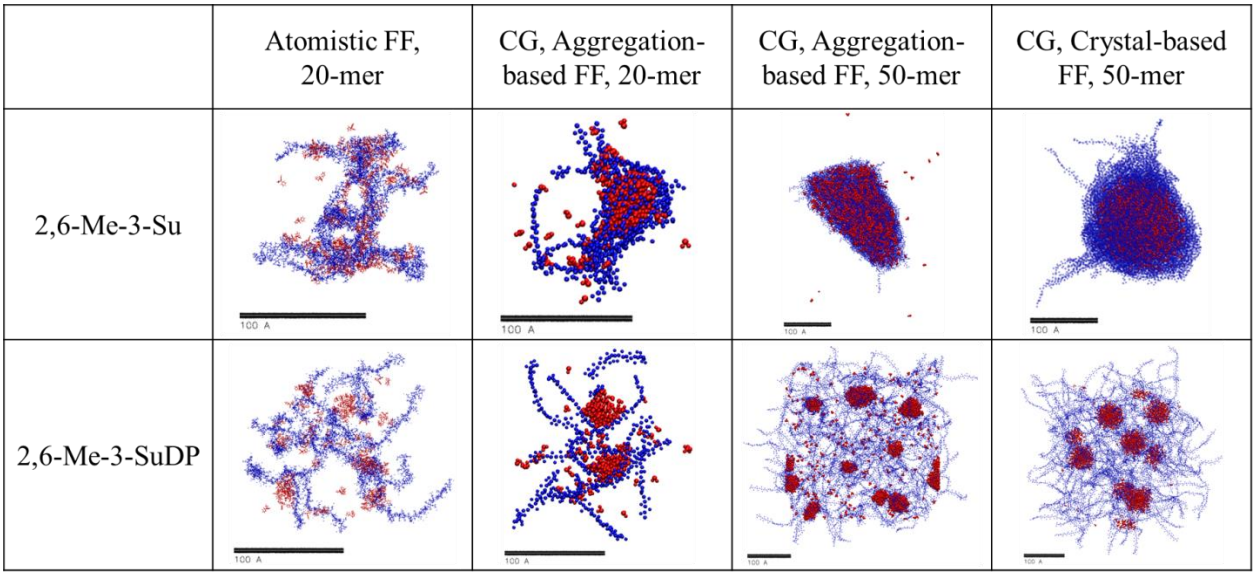


Figure 4.3: Snapshots from atomistic and CG simulations of model polymer (blue)-phenytoin (red) systems. The monomer substitution patterns of the homogenous oligomers and polymers simulated are listed as the row titles. Snapshots from the atomistic simulations are shown in the first column, and the snapshots from the CG simulations of 20-mers and 50-mers using the aggregation-based CG force field (FF), and 50-mers using the crystal-based CG FF, are shown in the second, third, and fourth columns, respectively. For the 20-mer systems, 15 oligomers (10wt%) and 150 drug molecules (3.5wt%) were randomly placed in the cubic simulation box of 12nm on the side. For the 50-mer systems, 70 polymers (3wt%) and 1500 drug molecules (1.5wt%) were placed in the cubic simulation box of 35nm on the side.

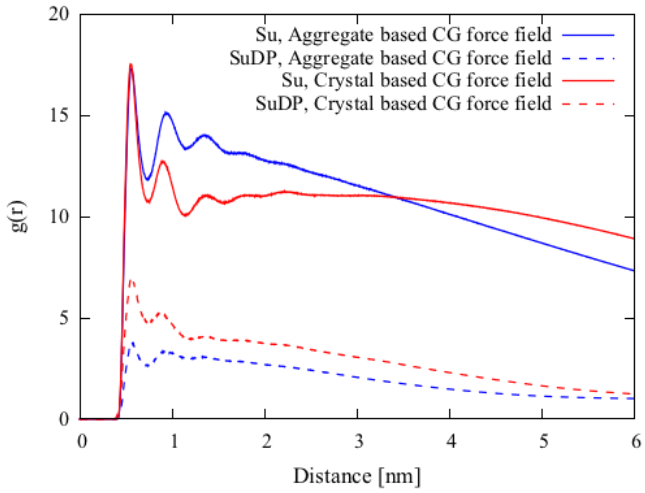


Figure 4.4: Intermolecular CG polymer-drug RDFs obtained from CG simulations of 70 50-mer homogenous model polymers and 1500 phenytoin molecules. RDFs of the simulations using aggregate-based CG force field and crystal-based CG force field are shown in blue and red respectively. RDFs of the polymer systems with protonated succniyl group (Su) are shown as solid lines and with deprotonated succniyl group (SuDP) as dashed lines.

To further test the robustness of the aggregate-based CG force field, we compute the polymer-drug RDFs generated for phenytoin-2,6-Me-3-Ac polymers as well as for phenytoin-2,6-Me-3-Su polymers. The RDFs show that phenytoin molecules have stronger interactions with 2,6-Me-

3-Ac polymers than with 2,6-Me-3-Su polymers, which agrees well with the relative interaction strengths reported in the atomistic simulations by Jha *et al.*[40] These RDFs also show in upper GI tract, where the phenytoin molecules are released, 2,6-Me-3-Ac polymer has much stronger interaction with phenytoin than 2,6-Me-3-SuDP polymer. This is in good qualitative agreement with experimental results, where HPMCAS with high succinyl substitution ratio was found to be less effective in maintaining phenytoin supersaturation than polymer with high acetyl substitution ratio[18,102]. We test the robustness of the structures observed in the CG simulations. To do so, we mimic a shift in the pH condition and switch between the protonated and deprotonated states by changing the non-bonded interaction parameters of 2,6-Me-3-Su and 2,6-Me-3-SuDP polymers. Upon changing from the protonated to the deprotonated state, the polymer-drug matrix quickly swells and dissolves in the solution, resulting in a few large phenytoin-only aggregates inside the polymer network. The opposite happens when deprotonated state is changed to the protonated state (Figure 4.5).

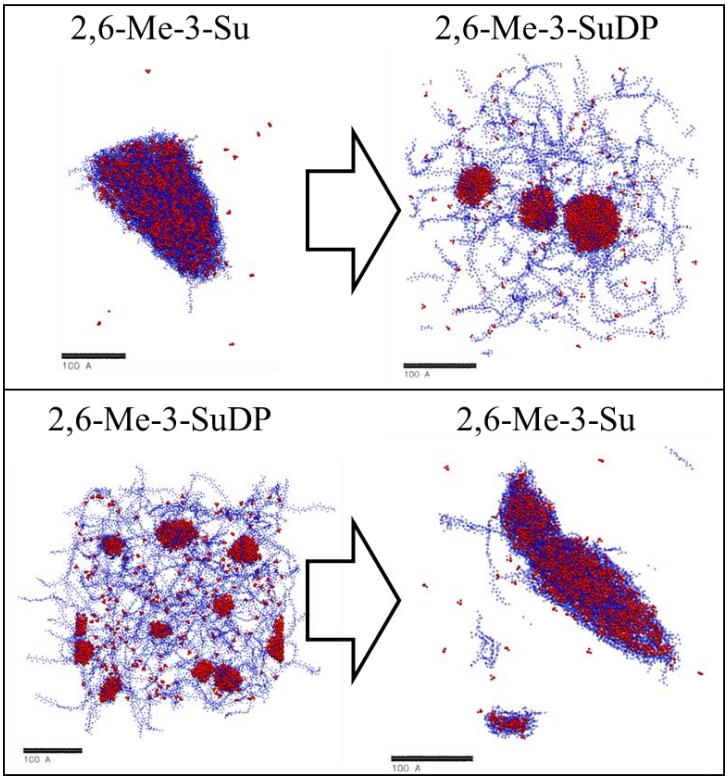


Figure 4.5: Snapshots from CG simulations of model polymer (blue)-phenytoin (red) systems. In the first row, we show the transformation from phenytoin-2,6-Me-3-Su system to phenytoin-2,6-Me-3-SuDP system, mimicking pH shift from 3 to 7. In the second row, we show the reversed transformation. The polymer and phenytoin concentrations are 3wt% and 1.5wt% respectively.

Next we build heterogeneous 50-mer HPMCAS model polymer chains based on the probability of occurrence of each monomer substitution type that represents an example of commercial HPMCAS polymer (Table 2.5). We assume random substitution patterns and construct three polymer sequences. Both protonated and deprotonated versions of these model polymer chains are simulated with phenytoin molecules. We solvate cubic simulation boxes of 35nm on each side containing 70 polymer chains (~3 wt%) and 1500 phenytoin molecules (~1.5 wt%). We also construct a system consisting of only 1500 phenytoin molecules, with no polymer. Snapshots from these systems after 20μs are shown in Figure 4.6. Without the polymer, phenytoin molecules quickly form small aggregates, and these aggregates then merge into bigger aggregates. Presumably these aggregates, given enough time, will nucleate and form crystals. When HPMCAS polymers are present, however, a polymer-drug complex is formed. Based on visual inspection, the complex formed between deprotonated HPMCAS and phenytoin is less compact than that between protonated HPMCAS and phenytoin.

We used cluster size analysis, implemented in GROMACS utility g_clustsize, to track the size of the largest drug cluster in the system. A drug molecule is considered to be a part of the cluster if COM distance between the molecule and any of the molecules that is already included in the cluster is less than 0.5nm. The sizes of largest drug clusters in all three systems as a function of simulation time are shown in Figure 4.7. In the drug-only system, the size of the cluster increases to around 200 molecules within the first 5μs. Thus, many small clusters are formed shortly after the initiation of the simulation. A merger between two small clusters occurs at around 15μs as shown by the doubling of the size of the largest cluster in Figure 4.7 (blue line), and we expect similar mergers to occur if we were to continue the simulation run beyond 20μs. In the polymer-drug systems, the size of the largest drug cluster quickly increases to around 500 due to the strong polymer-drug interactions and complexation. Protonated HPMCAS is thus more effective at inhibiting the growth of the drug cluster beyond this size, while the size of the largest drug cluster in the deprotonated HPMCAS-drug complex eventually plateaus at over 700 molecules. These simulations suggest that model HPMCAS polymer chains form a network in the HPMCAS-phenytoin complex, preventing the size of the phenytoin aggregate from growing beyond a certain limit. In addition, we hypothesize that the presence of the polymer in close proximity to the phenytoin cluster will greatly reduce the probability of nucleation and slow down the rate of crystallization, therefore prolonging the phenytoin supersaturation upon

solvation of the solid dispersion formulation. These simulations also shine light on the effect of pH on the solvation behavior of the HPMCAS. Upon shift of pH from 3 in the stomach to 7 in the small intestine, protonated succinyl groups becomes deprotonated, causing the polymer matrix to swell. As a result, drug molecules form slightly larger-sized aggregates. Although we are not considering the release of the drug molecules from the matrix in this set of simulations, it is reasonable to assume that during the release process, the diffusion of the drug molecules out of the complex will compete with the aggregation of drug molecules inside the complex.

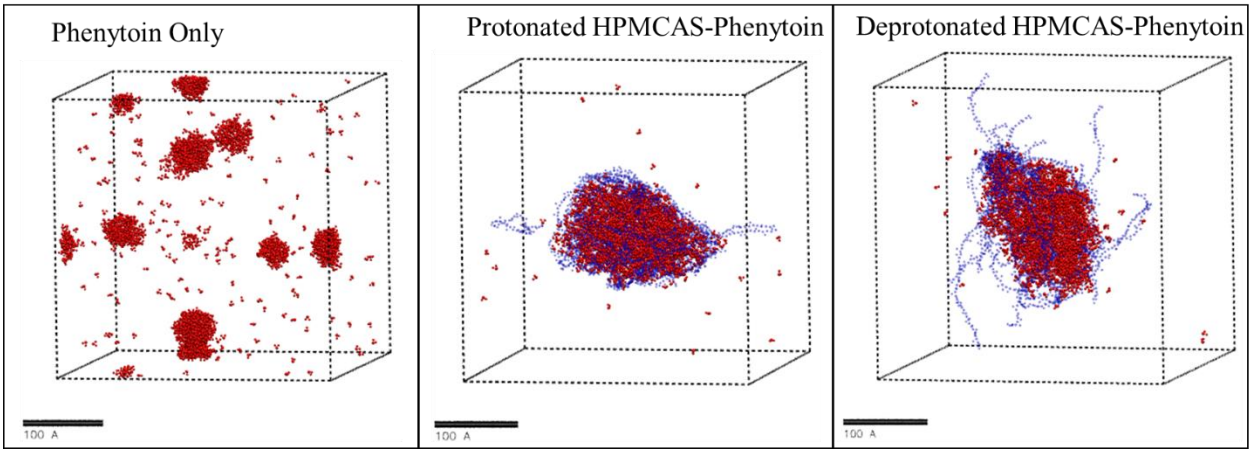


Figure 4.6: Snapshots from CG simulations of phenytoin (red) only and polymer (blue)-phenytoin (red) systems. The heterogeneous 50-mer model polymer chains are constructed using the percentages of each monomer substitution type shown in Table I. The polymer and phenytoin concentrations are 3wt% and 1.5wt% respectively.

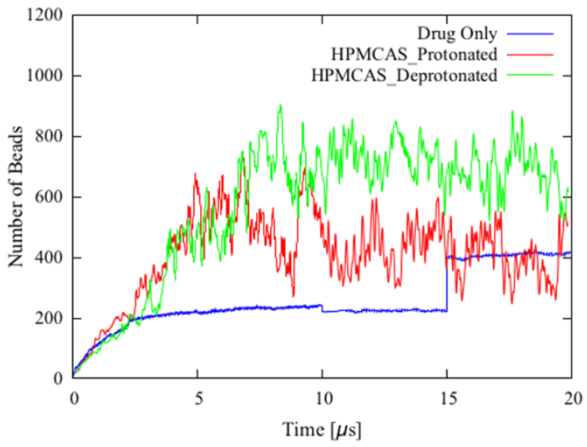


Figure 4.7: The sizes of the largest phenytoin clusters as a function of simulation time in three systems shown in Figure 4.6. The cluster sizes are tracked using the GROMACS utility g_clustersize.

4.4 Inhibition of Drug Aggregation in HPMCAS-Phenytoin Complex

We now turn our attention to understand how specific HPMCAS functional groups enhance the performance of the HPMCAS-Phenytoin solid dispersion formulation (SDD). Friesen et al.[9] suggested that the key structures in the HPMCAS based SDD that maintain the API supersaturation are the nanostructures with diameters ranging between 20-100nm. Our coarse-grained force fields developed for HPMCAS and phenytoin allow us to simulate these nanostructures directly and to understand the interaction modes between the polymer chains and drug molecules in these structures. An excellent polymeric excipient is expected to work in one or both of the following two ways 1) slowing down the drug diffusion rate, thus reducing the probability of a nucleation event and 2) interacting with drug molecule via specific intermolecular interaction, leading to suppression of the crystal nuclei formation. Because our phenytoin CG force field is not capable of simulating nucleation event or crystal growth, our main objective is therefore 1) to understand the intermolecular interaction mode between polymer chains and drug molecules, 2) to provide quantitative measurement of the drug diffusivity inside the polymer matrix, and 3) to study the effect of polymer properties, including concentration, chemistry, and chain length, on the diffusion and aggregation of the drug molecules in the cluster. In this section, we discuss our simulation results on how polymers inhibit the drug aggregation.

Effect of the Polymer Concentration

We study the effect of five functional groups, namely methyl (Me), hydroxypropyl acetyl (HPAc), acetyl (Ac), succinyl (Su), and deprotonated succinyl (SuDP). These functional groups are chosen due to their high probability of occurrence in HPMCAS (Table 2.5). We set up simulations with homogenous polymer chains and phenytoin molecules. Two concentrations, namely 20% and 33% drug loadings, and two polymer chain lengths, namely 50-mer and 100-mer, have been explored. For the 50-mer systems, the cubic simulation boxes of 35nm on each side are solvated with 70 polymer chains (~3 wt%) and 1500 phenytoin molecules (~1.5 wt%) for 33% drug loading, and 140 polymer chains (~6 wt%) and 1500 phenytoin molecules (~1.5 wt%) for 20% drug loading respectively. For the 100-mer systems, the numbers of polymers are halved in each simulation box comparing to the 50-mer systems, to maintain the same concentration. We plot the number of contacts between the drug molecules as a function of simulation time in these systems, shown in Figure 4.8. A contact is defined if the center of mass

of any one phenytoin molecule bead is less than 0.6nm distance away from the center of mass of another phenytoin molecule bead. Thus, a higher number of contacts correspond to more phenytoin aggregation. In Figure 4.8a, we show the number of contacts for the five polymer chemistries at two concentrations, in addition to the number of contacts without the presence of the polymer chains (i.e. free drug). The error bar for each system is taken to be the difference between two repeating simulations, which are small for all systems studied. The numbers of contacts plateau after 4μs in all systems, suggesting the size of the aggregates in the systems approach equilibrium. Comparing with the free drug only system, an effective polymer excipient should decrease the number of contacts between the drug molecules as a results of drug-drug interaction inhibition. As expected, all five polymer excipients form polymer-drug clusters, reducing the number of contacts between phenytoin molecules. Succinyl, hydroxypropyl acetyl, and acetyl are the top three performers, reducing the number of drug contacts by more than 40% comparing to those in the free drug only system. The effective aggregation inhibition of these functional groups can be attributed to both the strong intermolecular interaction strength and the bulkiness of the side group. Due to the size of the side groups, these three polymers form porous polymer-drug matrices, which accommodate the phenytoin drug molecules (see Figure 4.9a). In contrast, a polymeric excipient is less effective in inhibiting drug aggregation if the functional group has either weak intermolecular interaction strength or small size. For example, the size of the methyl functional group is small, thus only allows a well aligned bundled network to form due to its relatively strong interaction strength. As a result, drugs cannot be accommodated within polymer matrix (see Figure 4.9b). On the other hand, deprotonated succinyl group has weak intermolecular interaction with the drug molecules. Therefore, despite its large size, drug molecules diffuse fast inside the polymer matrix and form aggregates (see Figure 4.9c).

We also show the results from simulations of 33% drug loading in the inset of Figure 4.8a. The numbers of drug-drug contacts are higher in all simulations compare to jthose in simulations of 20% drug loading, suggesting the polymeric excipients are less effective at high drug loading. The order of polymeric excipient effectiveness, measured by the number of drug contacts in the simulation box, remains the same as the order obtained from the simulations of 20% drug loading. Both hydroxypropyl acetyl and succinyl functional groups are still highly effective. Acetyl group is less effective at high drug loading (i.e. 33%) than at low drug loading (i.e. 20%). We notice again that the deprotonated succinyl group is more effective than the methyl group.

This suggests that the bulkiness of the functional group is more important than strength of the intermolecular interaction for a polymer excipient to achieve effective drug aggregation inhibition. In Figure 4.8b, we show the same set of simulations conducted using polymers of 100 monomers long. We supply three snapshots from the simulations of 100-mer chains in Figure 4.10. Even though the rank of the polymeric excipient effectiveness remains the same, polymers are less effective at inhibiting drug aggregation. This is likely due to the fewer numbers of polymer-polymer contacts in the structure formed by long rigid chains (see Figure 4.10a).

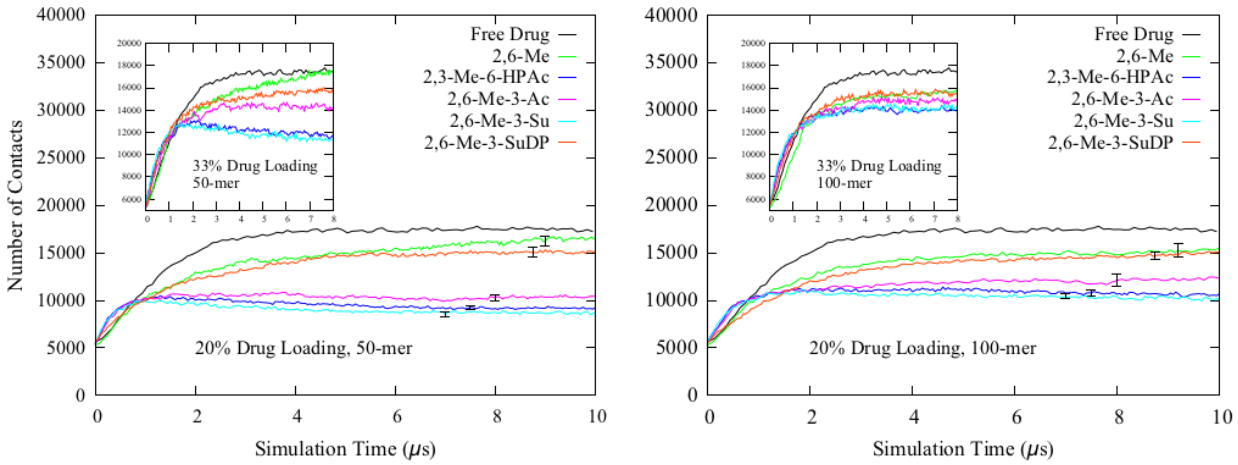


Figure 4.8: Number of contacts between phenytoin molecules as a function of simulation time. The 50-mer systems (left) contain 70 polymer chains (~3 wt%) and 1500 phenytoin molecules (~1.5 wt%) for the 33% drug loading, or 140 polymer chains (~6 wt%) and 1500 phenytoin molecules (~1.5 wt%) for 20% drug loading respectively. The 100-mer systems (right) contain half number of polymer chains comparing to their 50-mer counterparts.

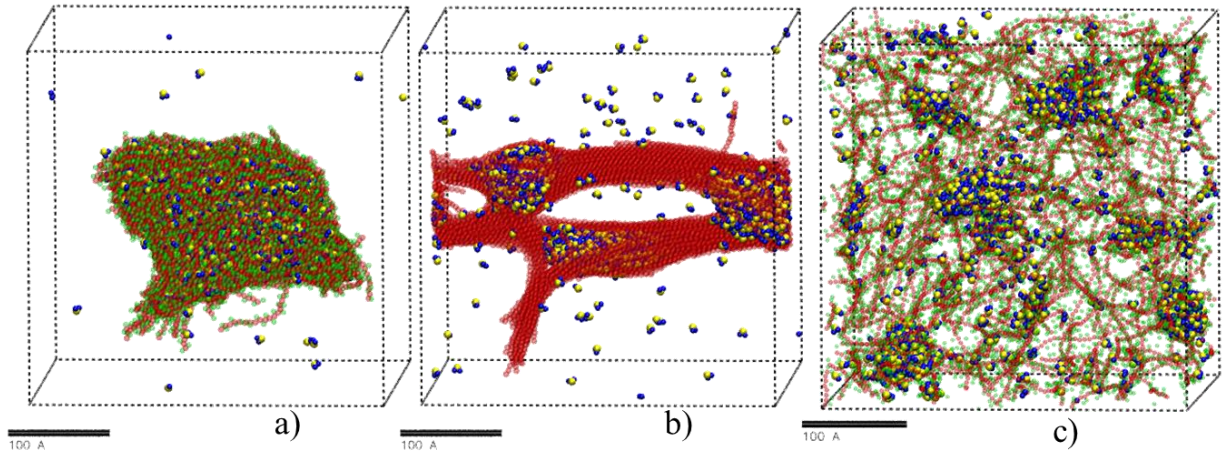


Figure 4.9: Snapshot of the simulation systems of 50-mer 2,3-Me-6-HPAc (a), 2,6-Me (b), and 2,3-Me-6-SuDP (c). The drug loadings in all three systems are 20%.

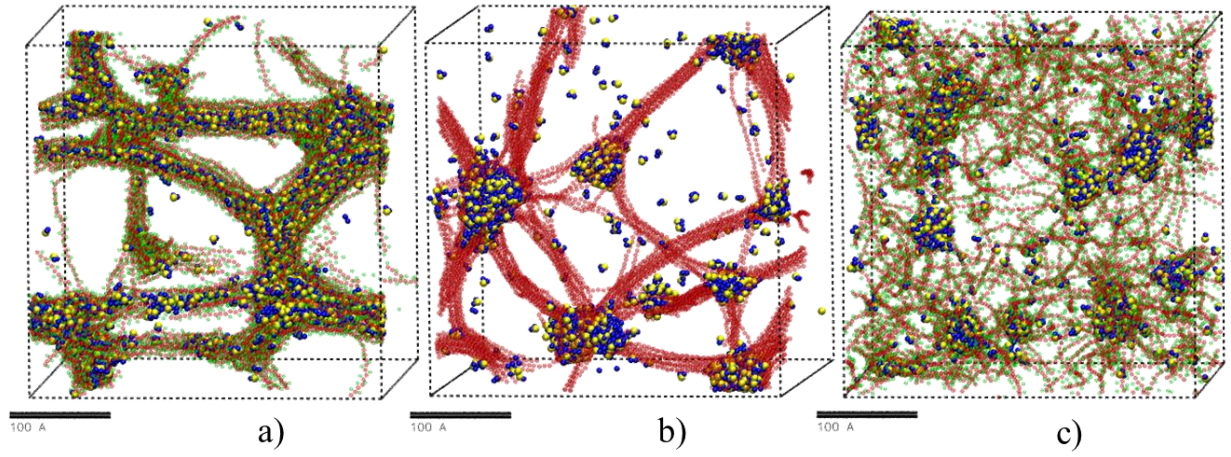


Figure 4.10: Same as Figure 4.9, except that polymers are 100-mer long.

Effect of Substitution Position

Next, we focus on understanding the effect of substitution position on the effectiveness of the polymeric excipient. Each cellulosic monomer has three positions available for substitution. We have demonstrated that different methylation patterns on methylcellulose (i.e. 2,6-Me and 2,3,6-Me) result in different solvation behaviors. We wish to explore whether certain monomer substitution types are more effective than others using the CG simulations, in order to guide the design of HPMCAS based solid dispersion formulation. We study acetyl group and succinyl group at two substitution locations, the 3-position and the 6-position. The 6-position is chosen because functional group becomes bulkier at this position due to the additional methylene bridge (-CH_2-). In addition, we want to compare the effectiveness of hydroxypropyl acetyl, hydroxypropyl, and acetyl group at the 6-position, to study if selectively introducing hydroxypropyl group without the acetyl group attached will improve the performance of the polymeric excipient.

In Figure 4.11a, we show the number of drug-drug contacts as a function of time obtained from simulations of polymers with hydroxypropyl acetyl, hydroxypropyl, and acetyl functional groups. We find that at low drug concentration (i.e. 20% drug loading), all four functional groups are very effective in inhibiting drug aggregation. However, as the drug concentration increases to 33% drug loading, pure acetyl group at both 3-position and 6-position become less effective (Figure 4.11a inset). Interestingly, hydroxypropyl group (shown in yellow) is very effective in inhibiting

drug aggregation at both drug concentrations. This again shows that at high drug concentration, a functional group with bulky side groups and strong intermolecular interactions (i.e. hydroxypropyl acetyl and hydroxypropyl) performs better than a side group with only strong intermolecular interaction (i.e. acetyl). In addition, there aren't significant differences in the effectiveness between the polymeric excipients with acetyl group at 3-position and 6-position.

In Figure 4.11b, we show the number of drug-drug contacts as a function of time obtained from simulations of polymers with succinyl and deprotonated succinyl functional groups. At both drug loading concentrations, the protonated succinyl group inhibits drug aggregation effectively. Interestingly, the deprotonated succinyl group at the 6-position is more effective than the deprotonated succinyl group at the 3-position. This could be attributed to the slightly bulkier side group size when succinyl group is substituted at the 6-position compare to that at the 3-position.

The simulation results shown here agree well with the results reported by Ting et al[18] in their recent work. In particular, we show that polymeric excipients are more effective in inhibiting drug aggregation at low drug concentration, which ultimately lead to better dissolution performance. Also, we show the hydroxypropyl group is one of the most effective functional group in the polymeric excipient. Indeed, Ting has reported polymers with functional groups analogous to hydoxypropyl groups are one of the best performing polymeric excipients for phenytoin. We note that although our simulation results indicate that polymers with succinyl group outperform those with hydroxypropyl acetyl group, the succinyl group will remain deprotonated in small intestine when the drugs release occurs. Therefore, our simulation results suggest that the polymers with succinyl group may help to stabilize the drug molecules in the stomach, where the succinyl group is protonated.

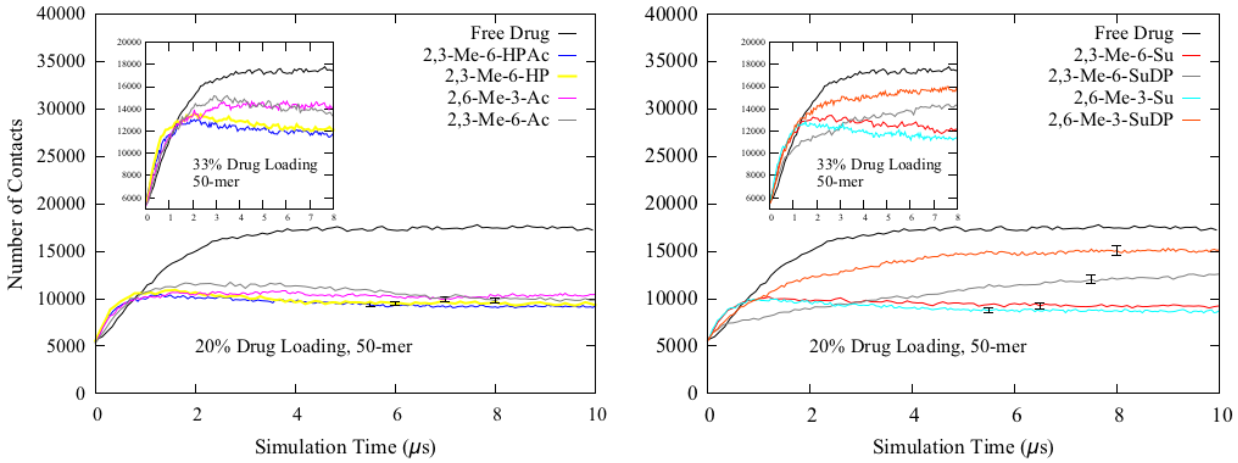


Figure 4.11: Same as Figure 4.8, except for different polymer chemistries.

4.5 Simulation of Phenytoin Release from Polymer-Drug Complex

Next, we estimate the release rates of the drug molecules from the polymer-drug clusters, and their dependence on the polymer properties including concentration, chemistry, chain length, and the size of the cluster. Solid dispersion particles break apart upon entering the small intestine and release most of the drug molecules within minutes. After the release, drug molecules interact with the polymer chains to form nanostructures and to maintain supersaturation, similar to the simulations discussed in the previous section. Here, we set up release simulations to measure the release rate of the drug molecules from the cluster. We consider two geometries, namely a rectangular geometry and a spherical geometry (Figure 4.12). In the rectangular geometry (Figure 4.12a), an arbitrary layer of absorbing beads is placed in the upper half of the simulation box. In the spherical geometry (Figure 4.12b), an arbitrary layer of absorbing beads is placed around the polymer-drug cluster. We test whether the two geometries give similar drug release profiles. We track the number of released drugs by measuring the number of drug molecules that are in contact with the arbitrary absorbing layer. In Figure 4.13, we show that the two geometries (solid and dashed lines, respectively) produce very similar release profiles. Because the spherical geometry contains many more beads inside the arbitrary absorbing layer, and is therefore computational more intensive, we decide to conduct release simulations using the rectangular geometry.

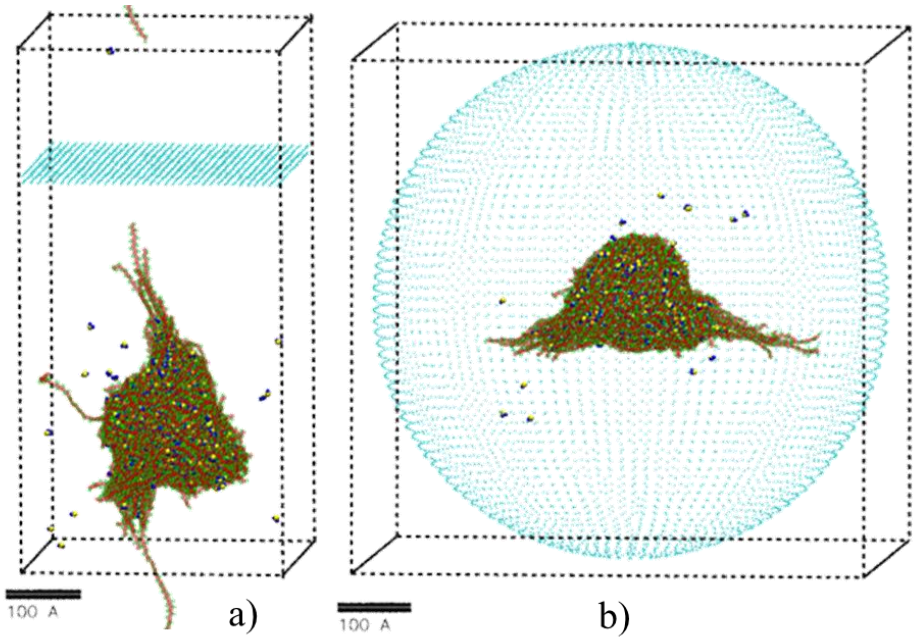


Figure 4.12: Snapshots of the drug release simulations in rectangular (a) and spherical (b) geometries respectively.

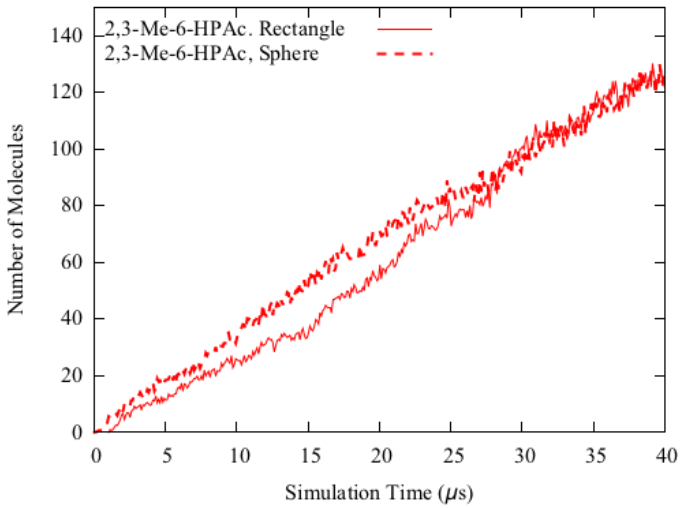


Figure 4.13: Release profiles of the phenytoin molecules in 50-mer 2,3-Me-6-HPAc polymer-drug complex. The system has 20% drug loading.

We show the release profiles of the five polymer-drug complexes in Figure 4.14, which is the number of molecules released as a function of simulation time. Here we hypothesize that in order to delay "burst release", where the drug concentration peaks shortly after the solvation of solid dispersion particles, an effective polymeric excipient slows down the release rate of the drug molecules. From the simulations of 50-mer long chains, we find the hydroxypropyl acetyl and

succinyl functional groups are the most effective functional groups to slow down the release of the drug molecules, followed by the acetyl, methyl, and deprotonated succinyl group. We find that polymers with bulky functional groups and strong interaction strength are the ones that are most effective at slowing down the release of the drug molecules. However, in contrary to the results discussed in the previous section, polymers with methyl group outperform those with deprotonated succinyl group in terms of slowing down the drug release rate. This suggests that strong interaction strength is more important than bulky functional group size in slowing down the drug release. The rank order of the functional group effectiveness remains the same for the higher drug concentration (i.e. 33wt%) and longer chain length (i.e. 100-mer). In addition, we computed the release profile of polymers with functional group hydroxypropyl, as well as with functional group acetyl and succinyl at both the 3 position and the 6 position. Hydroxypropyl group perform similarly with hydroxypropyl acetyl group, and acetyl and succinyl groups at the 6-position slightly outperform those functional groups at the 3-position.

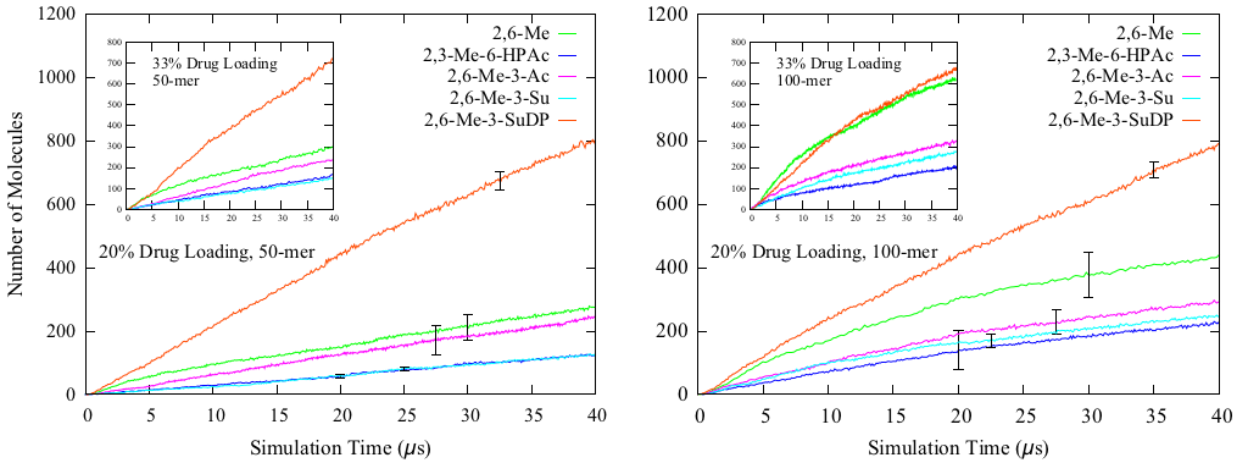


Figure 4.14: Release profiles of the phenytoin molecules in 50-mer polymer-drug clusters (left) and 100-mer polymer-drug clusters (right). The compositions of the clusters are the same as the ones described in Figure 4.8.

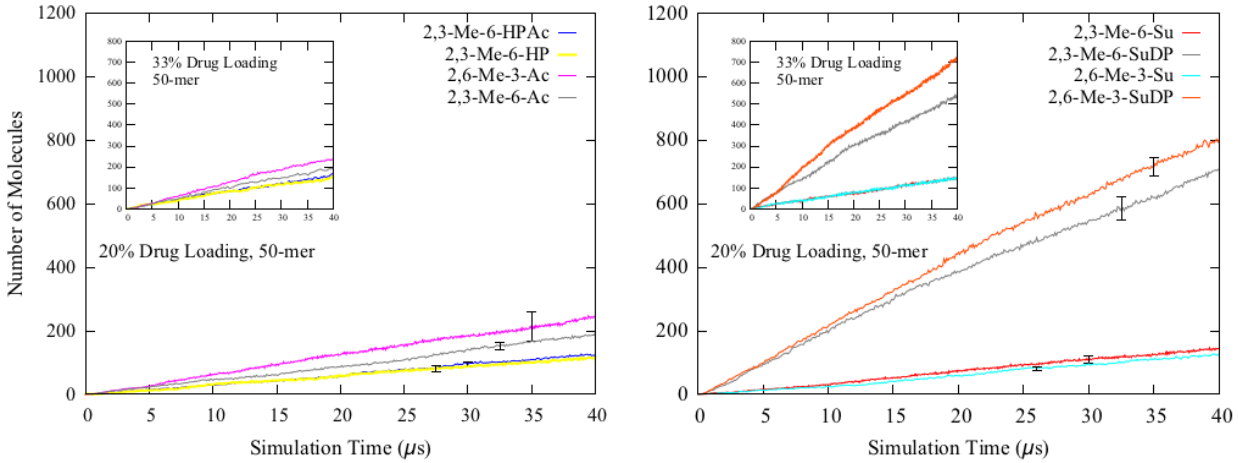


Figure 4.15: Same as Figure 4.14, except for different polymer chemistries.

4.6 Analytical Model for Phenytoin Release

Even though the CG simulations allow access of much larger system size and time scale comparing to the atomistic simulations, simulating the complete drug release process, which takes place over a time of seconds to hours, from the solid dispersion particles, which have diameters ranging from several hundreds of nanometers to several micrometers, is still beyond reach. As a result, we have to rely on continuum level transport modeling tool to model the drug release behavior from these particles. Here, we introduce a simplified two phase transport model based on our drug release simulations, shown in Figure 4.16. We assume the drug molecules have two constant diffusivities in the two domains, namely inside the polymer matrix (D_1) and inside the release media (D_2). Carr et al.[104] have provided an analytical solution for diffusive transport model in multilayer composite material. Their proposed solution is based on finite Fourier transformation approach, where a global set of eigenvalues λ_n and a unique set of basis functions $\Phi_{n,i}$ for each layer are computed. In this work, we have adapted Carr et al.'s solution to the spherical geometry.

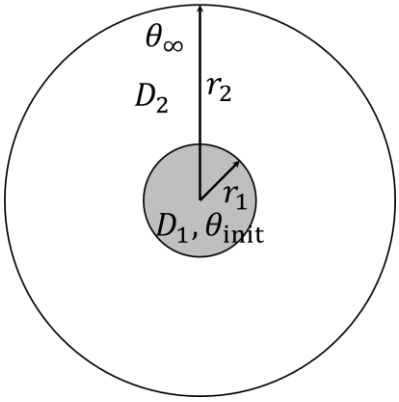


Figure 4.16: Schematics of the drug release model. The inner region (region 1) contains the polymer matrix, and the outer region (region 2) is the release media, which is taken to be water in our transport model.

Transport Model Detail

We start our derivation with the general transient governing equation for spherical geometry.

$$\frac{\partial \theta_i}{\partial t} = D_i \frac{1}{r^2} \frac{\partial}{\partial r}\left(r^2 \frac{\partial \theta_i}{\partial r}\right) \tag{4.1}$$

where i is 1 for inner region (polymer matrix) and is 2 for outer region (release media). We include three boundary conditions (Equation 4.2-4.4). In particular, we employ the Dirichlet boundary condition on the outer edge of the region 2 and match both the flux and concentration on the boundary between region 1 and 2.

$$\theta_2(r_2) = 0 \tag{4.2}$$

$$D_1 \frac{\partial \theta_1(r_1,t)}{\partial r} = D_2 \frac{\partial \theta_2(r_2,t)}{\partial r} \tag{4.3}$$

$$\theta_1(r_1, t) = \theta_2(r_2, t) \tag{4.4}$$

The Neumann boundary condition is automatically satisfied in the center. For the initial condition $f_i(r)$, we assume the drug concentration is uniform in region 1 with a concentration of θ_{init}, and zero everywhere in the outer region.

Because the external boundary condition is homogenous, we can proceed with the finite Fourier transfer approach. The solution then has the following form (Equation 4.5). In particular, the basis function $\Phi_{n,i}(r)$ satisfies Sturm-Liouville equation (Equation 4.6) and has the following form (Equation 4.7)

$$\varphi_i(r,t) = \sum_n a_n e^{-t\lambda_n^2} \Phi_{n,i}(r) \tag{4.5}$$

$$\frac{1}{r^2}\frac{d}{dr}\left(r^2 \frac{d\Phi_{n,i}}{dr}\right) = -\lambda_n^2 \Phi_{n,i}(r) \tag{4.6}$$

$$\Phi_{n,i}(r) = \gamma_{n,i} \frac{\sin\left(\frac{\lambda_n}{\sqrt{D_i}}(r-l_{i-1})\right)}{r} + \delta_{n,i} \frac{\cos\left(\frac{\lambda_n}{\sqrt{D_i}}(r-l_{i-1})\right)}{r} \quad (4.7)$$

We note for the inner region solution to be well defined everywhere, $\delta_{n,1}$ has to be 0. We plug the three boundary conditions into the basis function (Equation 4.7) and obtain a linear system (Equation 4.8) where $\boldsymbol{x} = \left(\gamma_{n,1}, \gamma_{n,2}, \delta_{n,2}\right)^T$ and $\boldsymbol{A}(\lambda_n)$ is a 3 x 3 matrix. Non-negative roots of $\det\left(\boldsymbol{A}(\lambda_n)\right) = 0$ are required to give non-trivial solution to Equation 4.8 (i.e. $x \neq 0$).

$$\boldsymbol{A}(\lambda_n)\boldsymbol{x} = \boldsymbol{0} \quad (4.8)$$

These nonnegative roots are the eigenvalues λ_n. For each eigenvalue, the corresponding coefficients $\gamma_{n,1}, \gamma_{n,2}$, and $\delta_{n,2}$ can be determined by solving the linear system. Lastly, the unknown coefficient appeared in Equation 4.5 can be solved using the initial condition $f_i(r)$, shown in Equation 4.9 where $\langle f(r), g(r)\rangle_i = \int_{l_{i-1}}^{l_i} f(r)g(r)r^2 dr$.

$$a_n = \frac{\sum_{i=1}^{2}\langle f_i(r), \Phi_{n,i}(r)\rangle_i}{\sum_{i=1}^{2}\langle \Phi_{n,i}(r), \Phi_{n,i}(r)\rangle_i} \quad (4.9)$$

Mutual Diffusivity Estimation

There are two diffusivities in our transport model, namely the diffusivity of drug molecules inside polymer matrix (D_1) and the diffusivity of drug molecules in the release medium water (D_2). The diffusivity of phenytoin molecules in water is 5.0E-6 cm^2/s.[105] The diffusivity of phenytoin molecules in polymer matrix cannot be obtained straightforwardly. The simulations provide self-diffusivity, or mean square displacement, of the drug molecules in the polymer matrix. However, this does not correspond to the diffusivity required in the transport model, which is the mutual one dimensional radial diffusivity. Therefore, to obtain the mutual diffusivity, we fit the number of drug molecules released from the matrix using the transport model. We first test whether the mutual diffusivity is affected by the size of the polymer-drug cluster by setting up four simulation systems with cluster diameters of 10, 15, 20, and 25nm respectively. We show the fitted diffusivities as a function of cluster diameter in Figure 4.17. We find that the diffusivity plateaus at cluster diameter 20 and beyond. Therefore, we decide to use the simulation results obtained from clusters with diameter of 20nm.

We obtain the mutual diffusivities for nine polymer chemistries and at 2 different drug loadings, namely 20%, and 33%. We show that polymers with HPMCAS functional groups, including

hydroxypropyl, hydroxypropyl acetyl, acetyl, and succiniyl, are able to slow down the phenytoin mutual diffusivity significantly comparing to the methyl and deprotonated functional groups. Among the HPMCAS functional groups, the phenytoin diffusivities are similar or lower when the drug loading is lower (i.e. 20%) comparing to the high drug loading (i.e. 33%). However, for methyl and deprotonated succinyl functional groups, the phenytoin increases as the drug concentration decreases. This is possibly due to the fact that polymers with these two functional groups are not efficient in inhibiting drug aggregating (Figure 4.8). Therefore, drug molecules aggregate more at high drug loading, slowing down the release rate.

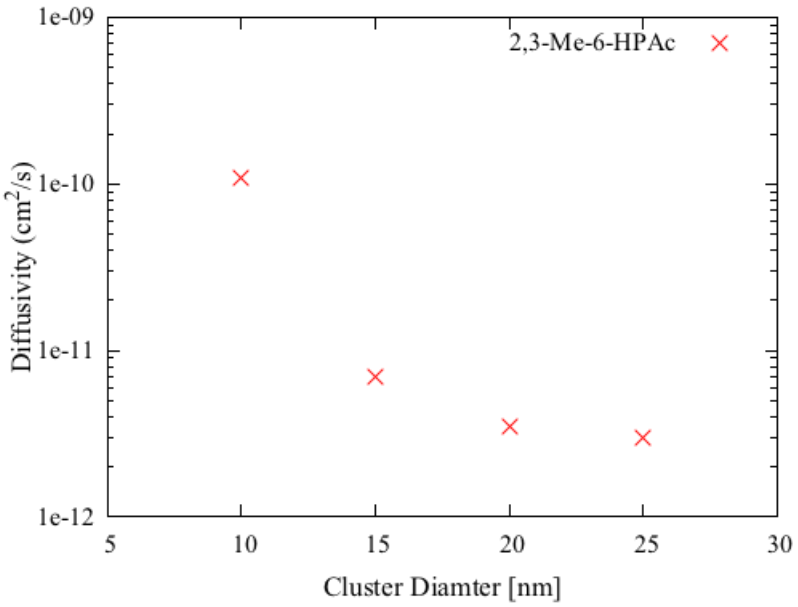


Figure 4.17: Fitted mutual diffusivity of phenytoin molecules from polymer-drug clusters with various diameters. The system simulated is 2,3-Me-6-HPAc homopolymers with 20% drug loading.

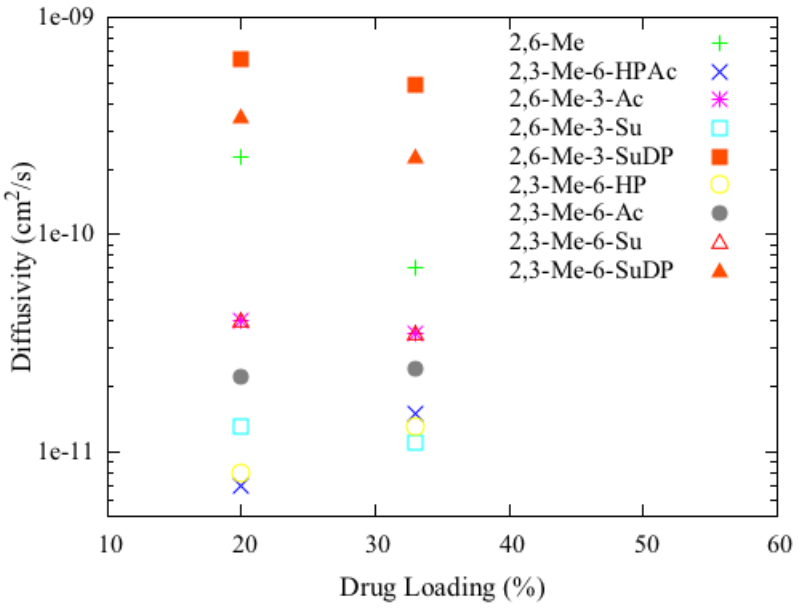


Figure 4.18: Fitted mutual diffusivity of phenytoin molecules from polymer-drug clusters of various polymer chemistries. Two drug loadings are studied, namely 20% and 33%.

A similar trend has been identified from 100-mer simulations, and the diffusivities obtained from 100-mer systems are similar to those obtained from 50-mer systems. Next, we use the fitted diffusivity to estimate the time it takes for the drugs to fully release from the solid dispersion particles. We model a solid dispersion particle with diameter of 2μm, similar to the particles described in Ting's work.[18,106] We estimate the release time to be the time required for 99% of the drug molecules to depreciate from the solid dispersion particle. We show the estimated release time for particles with 20% drug loading in Table 4.1. We find that it takes between just a few seconds to over 500 seconds for the phenytoin molecules to be fully released from a homopolymers matrix. In a typical dissolution profile obtained from experiments, free drug concentration peaks in the first few minutes of the profile and gradually decline afterwards. It is believed that the peak occurs due to either the drugs are fully released from the particles or the particles are disintegrated. The release time frames predicted by our transport model fit nicely with the time it takes for the drug concentrations in the dissolution profiles to peak. In addition, Commercial HPMCAS polymer products from The Dow Chemical Company have three different grades, with Acetyl group ranging from 5-14% and succinyl group ranging from 7-14%. Higher succniyl group content (i.e. 14%) results in a dissolution profile that peaks within few minutes, while lower succinyl group content (i.e. 8%) delays the peak in the dissolution profile till after one hour.[29] Our simulation results agree qualitatively with the trend in these dissolution

profiles, where deprotonated succinyl group has a much faster release rate than the other HPMCAS functional groups.

	50-mer		100-mer	
Functional Group	Diffusivity_Sim (1E-12 cm^2/s)	Release Time_Pred (Sec)	Diffusivity_Sim (1E-12 cm^2/s)	Release Time_Pred (Sec)
2,3-Me-6-HPAc	7	594	33	126
2,3-Me-6-HP	8	520	13	320
2,3-Me-6-Su	12	346	9	462
2,6-Me-3-Su	13	320	26	160
2,3-Me-6-Ac	22	189	59	70
2,6-Me-3-Ac	40	104	41	101
2,6-Me	228	18	134	31
2,3-Me-6-SuDP	349	11	512	8
2,6-Me-3-SuDP	645	7	444	10

Table 4.1: Mutual diffusivities obtained from simulation and release time estimated from transport model. The release time is estimated based on 20% drug loading, 2μm diameter particles and the time required to achieve 99% particles release

4.7 Concluding Remark

In this chapter, we develop a coarse-grained (CG) force field to model the interaction between hydroxypropyl-methylcellulose acetate succinate (HPMCAS) polymer and phenytoin molecule in solid dispersions drug formulation upon solvation. Our systematic and robust parametrization approach capture the structural information revealed in the atomistic simulations, and can be applied to incorporate new functional groups connected to the cellulose backbone as well as other drug molecules of interest. The polymer-only CG simulations reveal the chain length dependent solvation behavior for 50-mer and 100-mer model polymer chains. We model the heterogeneous model HPMCAS polymer-phenytoin mixture at pH of 3 and 7, corresponding to the protonated and deprotonated polymers, which both form a polymer-phenytoin complex. However, the protonated polymer was more effective at inhibiting the growth of the drug aggregate.

We then apply this force field to study the effect of polymeric excipient in solid dispersion formulation. In particular, we study polymeric excipients' ability to inhibit drug aggregation and to slow down the release of the drug molecules from the polymer-drug clusters. We find the size of the functional group and the intermolecular interaction strength are the two key properties for an effective polymeric excipient. We have therefore identified hydroxypropyl acetyl group, which has both bulky size and strong interaction strength, to the most effective functional group,

followed by hydroxypropyl and acetyl group, in good agreement with the results from experimental dissolution tests. We also present a transport model to estimate the release time of the drug molecules from the cluster. The estimated release time agree qualitatively well with the dissolution profiles for HPMCAS-phenytoin solid dispersion formulation.

Based on the simulation results, we have generalized the follow two design rules for an effective HPMCAS based polymer excipient for HPMCAS-phenytoin solid dispersion formulation. The first rule is to increase the percentage of acetyl and hydropropyl groups in HPMCAS, providing more effective drug aggregation inhibition. The second rule is to selectively introduce succinyl group at the 3-position when synthesizing HPMCAS because succniyl at 3-position outperforms succinyl at 6-position in both aggregation inhibition simulations and drug release simulations.

Chapter 5: Conclusions and Future Directions

5.1 Conclusions

In this doctoral work, we present a systematic multi-scale modeling approach to model the solvation behavior of cellulosic polymers. The goal is to use these computational techniques to provide researchers with molecular level insights in the cellulosic polymer intermolecular interactions that are not accessible through conventional experiments. We covered models from atomistic, coarse-grained, and continuum level. We then highlighted two applications to showcase computational models as useful tools to understand complex experimental systems.

In Chapter 2, we discussed the computational models used in this work. We first studied three atomistic force fields used to model cellulosic polymers, including two GROMOS force fields and one AMBER force field. We showed that all three atomistic force fields can reproduce the persistence length of the cellulosic chains. We then presented a systematic approach to develop coarse-grained (CG) force fields for methylcellulose and hydroxypropyl methylcellulose acetate succinate (HPMCAS). We represented each cellulosic monomer with one to three CG beads depending on the size of the functional groups. The intra- and intermolecular interactions were fitted based on the radial distribution functions obtained from atomistic simulations and were validated against available experimental data. We demonstrated that our approach is robust and can be extended to model cellulosic polymers with additional functional groups and drug molecules other than phenytoin. Our model is capable of simulating multiple polymer chains up to 1000 monomers long with sufficient details to differentiate the solvation behavior at this length scale.

In Chapter 3, we presented the simulation for both homogenous and heterogeneous methylcellulose oligomers using atomistic and coarse-grained (CG) force fields. While the atomistic simulations revealed the driving force of the short chain aggregation, the CG simulations showed the solvation behavior and revealed the gelation mechanism of methylcellulose chains. We compared our simulation results with both simulation and experimental results in the literature. We found the heterogeneous methylcellulose chain has a

persistence length of 9nm at room temperature, which agrees well with experimentally determined value. More importantly, we found the heterogeneous methylcellulose chains collapse into a ring structure for chain lengths of 600 monomers or more at elevated temperature. The ring structure formed has an outer diameter of 14nm and a void fraction of 26%, which agrees well with the gel morphology observed under tunneling electron microscopy. We then showed that self-assembly of individual MC rings segments into stacks or "proto-fibrils" which can also give rise to branches. These "proto-fibrils" are likely the precursors to the methylcellulose fibrils. The simulation results complement the theoretical work of methylcellulose gel formation mechanism In addition, we presented a simplified continuous analytical model to predict the collapse conformation of a single self-attractive semiflexible polymer chain in solution. We showed that a good qualitative agreement between theoretical and simulated results can be achieved at long chain length for transitions between collapsed states. Combined with the computational work presented in this chapter, our modeling work provides quick estimation of the collapsed state of a given polymer chain, and allows designing of polymer chemistries with controlled transitions between collapsed states.

In Chapter 4, we presented CG simulation results of interactions between HPMCAS and phenytoin drug molecules. Polymer and drug molecules form complexes in a short period of simulation time due to strong intermolecular interactions. 50-mer polymers form short bundled chains while 100-mer polymers form long bundled loops. Our force field distinguishes the polymer solvation behaviors at pH 3 and 7, where polymers with succinyl group aggregate tightly at pH 3 while they do not aggregate at pH 7. We also applied the CG force field to study the effect of polymeric excipient in solid dispersion formulation. In particular, we studied polymeric excipients' ability to inhibit drug aggregation and to slow down the release of drug molecules from polymer-drug clusters. The size of the functional group and the intermolecular interaction strength are the two key factors affecting the performance of polymeric excipient. We found polymers with hydroxypropyl acetyl and hydroxypropyl functional groups are among the most effective ones in both inhibiting the drug aggregation and slowing down the drug release, agreeing well with the experimental results. We also adapted a transport model to estimate the release time of the drug molecules from a typical solid dispersion particle of 2 μm in diameter. The estimated release time ranges from several seconds to several minutes depending on the functional groups, agreeing qualitatively well with the dissolution profiles for HPMCAS-

phenytoin solid dispersion formulation. The effect of lower succinyl group percentage on delaying the peak of the drug concentration in dissolution profiles has also been well captured in our model.

5.2 Future Directions

The methylcellulose model presented in this work has successfully explained several aspects of the methylcellulose gel observed in the experiments. First, the gelation temperature is not sensitive to the molecular weight once the chain length of the methylcellulose polymer exceeds 600-mer. Second, both the outside and inside diameters of the fibrils are set by the ring conformation, which is affected by the persistence length. However, there are still many unanswered questions that require refinement of the model.

1) There is a need to understand the molecular mechanisms that drive the conformational change of the polymer chains. In addition, it is unclear why polymer chains form ring structures at high polymer concentration where chains should be more energetically favorable to form bundled structures without the need to pay bending energy penalties. To address these questions, one could study the effect of CG bead shape and the effect of intermolecular interaction potential. For example, one could use an ellipsoid-shaped or a disk-shaped bead to represent the flat methylcellulose monomer, which may lead to "proto-fibril" structures with more well-defined void fraction. In addition, one could adopt asymmetric intermolecular interaction potentials. This, combining with the ellipsoid-shaped CG bead, may allow preferential bending into ring structures rather than bundled structures and therefore help to explain the formation of ring structure at high concentration. However, such simulations can be computationally costly on current generation hardware due to the use of tabulated potentials, and may not be feasible to be implemented to model long polymer chains such as the ones presented in this study.
2) We did not include water molecules in our simulations for computational efficiency reasons. However, water molecule may play a role in the solvation behavior of methylcellulose chains, especially at high concentration. Although we incorporated the effect of the solvent implicitly through the matching of atomistic level radial distribution functions, modeling the water molecules explicitly could give a more complete picture of the methylcellulose solvation behavior. One way to incorporate water model into the force field is to use a four-

water-molecules-to-one-CG-water-bead mapping, similar to that has been implemented in MARTINI force field. However, we note the CG water beads lack the ability to represent the local orientation of the water molecules, therefore may not correctly reflect the "cage effect" around hydrophobic molecules.

The computational models for HPMCAS presented in this work allow fundamental understanding of the polymeric-drug interaction and open up numerous opportunities for designing better polymeric excipient. However, there are areas in the solvation process of solid dispersion particles that still require deeper understanding. These are exciting opportunities for computational work that builds on top of the foundation laid in this doctoral work.

1) Simulating crystallization of the drug molecules is still a very challenging area for computational study. Recently, the crystal growth of simple molecule such as urea has been successfully simulated using atomistic simulation. Coarse-grained simulation has not been the first choice for simulating crystal growth, mainly due to the loss of structural detail during the coarse-graining procedure. Nevertheless, a systematic CG modeling approach that is capable of simulating crystal growth of small molecules including phenytoin has been developed very recently.[107] Yet still, simulating a nucleation event remains beyond reach. In the solid dispersion formulation, it is still up to debate whether the main function of a polymeric excipient is to prevent the nucleation event or to slow down the crystal growth after a nucleation event has occur. Therefore, a force field that is capable of simulating nucleation event is critical to offer an definitive answer on the main purpose of the polymeric excipient.
2) There is still lacking a general understanding on the fate of the drug molecules after the dispersal of a solid dispersion formulation, and its relationship with the size of the particle and different polymer chemistries. Even with CG simulations described in this work, simulating a particle of micrometer size for up to seconds of time is still beyond reach. Therefore, a higher level of coarse-grained model might be necessary to fully deduce the fate of the drug molecules.

Broadly speaking, the coarse-grained approach adopted in this work is systematic and robust, therefore has a potential to be applied to model solvation behaviors of other semiflexible polymers and to model the interaction between novel polymeric excipients and drug molecules.

For example, biopolymer DNA is of great scientific and commercial interest. DNA chains adopt various conformations at different water activity level, which leads to drastically different intermolecular interactions and functions. Although the existence of multiple DNA molecule conformations has been well documented, the exact driving mechanism is unclear. Modeling the solvation behavior of DNA using a modeling approach presented here can help unveil the origin of the environmental sensitivity of DNA conformation. Another potentially high impact application of the computational modeling approach presented in this work is to model the interaction between N-isopropylacrylamide (pNIPPAm) and drug molecules. pNIPPAm, along with a number of pNIPPAm based copolymers, has been recently identified to be highly effective polymeric excipient. Recent atomistic simulations study highlight the copolymer containing 80% pNIPPAm and 20% dimethylacrylamide (DMA) has strong interaction with phenytoin drug molecules[108], consistent with experimental results. A CG model can be developed for pNIPPAm-DMA copolymer and other pNIPPAm based copolymers to reveal the molecular level interaction in the polymer-drug cluster. We firmly believe that, with constant development in computer hardware and simulation algorithm, many more problems in both pharmaceutical and material design fields can be tackled by multi-scale computational modeling approaches.

References

(1) Pharmaceutical market worldwide revenue 2001-2015 https://www.statista.com/statistics/263102/pharmaceutical-market-worldwide-revenue-since-2001/ (accessed Feb 25, 2017).

(2) Lipinski, C. A. Lead- and Drug-like Compounds: The Rule-of-Five Revolution. *Drug Discov. Today Technol.* **2004**, *1* (4), 337–341.

(3) Hauss, D. J. Oral Lipid-Based Formulations. *Adv. Drug Deliv. Rev.* **2007**, *59* (7), 667–676.

(4) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and Computational Approaches to Estimate Solubility and Permeability in Drug Discovery and Development Settings. *Adv. Drug Deliv. Rev.* **2012**, *64*, *Supplement*, 4–17.

(5) Brewster, M. E.; Loftsson, T. Cyclodextrins as Pharmaceutical Solubilizers. *Adv. Drug Deliv. Rev.* **2007**, *59* (7), 645–666.

(6) Carrier, R. L.; Miller, L. A.; Ahmed, I. The Utility of Cyclodextrins for Enhancing Oral Bioavailability. *J. Controlled Release* **2007**, *123* (2), 78–99.

(7) Hintz, R. J.; Johnson, K. C. The Effect of Particle Size Distribution on Dissolution Rate and Oral Absorption. *Int. J. Pharm.* **1989**, *51* (1), 9–17.

(8) Willart, J. F.; Descamps, M. Solid State Amorphization of Pharmaceuticals. *Mol. Pharm.* **2008**, *5* (6), 905–920.

(9) Friesen, D. T.; Shanker, R.; Crew, M.; Smithey, D. T.; Curatolo, W. J.; Nightingale, J. A. S. Hydroxypropyl Methylcellulose Acetate Succinate-Based Spray-Dried Dispersions: An Overview. *Mol. Pharm.* **2008**, *5* (6), 1003–1019.

(10) Serajuddin, A. T. M. Solid Dispersion of Poorly Water-Soluble Drugs: Early Promises, Subsequent Problems, and Recent Breakthroughs. *J. Pharm. Sci.* **1999**, *88* (10), 1058–1066.

(11) Van den Mooter, G. The Use of Amorphous Solid Dispersions: A Formulation Strategy to Overcome Poor Solubility and Dissolution Rate. *Drug Discov. Today Technol.* **2012**, *9* (2), e79–e85.

(12) Liu, R. *Water-Insoluble Drug Formulation, Second Edition*; CRC Press, 2008.

(13) Williams, H. D.; Trevaskis, N. L.; Charman, S. A.; Shanker, R. M.; Charman, W. N.; Pouton, C. W.; Porter, C. J. H. Strategies to Address Low Drug Solubility in Discovery and Development. *Pharmacol. Rev.* **2013**, *65* (1), 315–499.

(14) Vo, C. L.-N.; Park, C.; Lee, B.-J. Current Trends and Future Perspectives of Solid Dispersions Containing Poorly Water-Soluble Drugs. *Eur. J. Pharm. Biopharm.* **2013**, *85* (3, Part B), 799–813.

(15) GRAS Substances (SCOGS) Database http://www.fda.gov/Food/IngredientsPackagingLabeling/GRAS/SCOGS/default.htm (accessed Sep 4, 2013).

(16) Alonzo, D. E.; Gao, Y.; Zhou, D.; Mo, H.; Zhang, G. G. Z.; Taylor, L. S. Dissolution and Precipitation Behavior of Amorphous Solid Dispersions. *J. Pharm. Sci.* **2011**, *100* (8), 3316–3331.

(17) Curatolo, W.; Nightingale, J. A.; Herbig, S. M. Utility of Hydroxypropylmethylcellulose Acetate Succinate (HPMCAS) for Initiation and Maintenance of Drug Supersaturation in the GI Milieu. *Pharm. Res.* **2009**, *26* (6), 1419–1431.

(18) Ting, J. M.; Navale, T. S.; Jones, S. D.; Bates, F. S.; Reineke, T. M. Deconstructing HPMCAS: Excipient Design to Tailor Polymer–Drug Interactions for Oral Drug Delivery. *ACS Biomater. Sci. Eng.* **2015**, *1* (10), 978–990.
(19) Yin, L.; Hillmyer, M. A. Preparation and Performance of Hydroxypropyl Methylcellulose Esters of Substituted Succinates for in Vitro Supersaturation of a Crystalline Hydrophobic Drug. *Mol. Pharm.* **2014**, *11* (1), 175–185.
(20) The Dow Chemical Company. METHOCEL Cellulose Ether: Technical Hannbook. 2002.
(21) Klemm, D.; Heublein, B.; Fink, H.-P.; Bohn, A. Cellulose: Fascinating Biopolymer and Sustainable Raw Material. *Angew. Chem. Int. Ed.* **2005**, *44* (22), 3358–3393.
(22) Keary, C. M. Characterization of METHOCEL Cellulose Ethers by Aqueous SEC with Multiple Detectors. *Carbohydr. Polym.* **2001**, *45* (3), 293–303.
(23) Beer, M. U.; Wood, P. J.; Weisz, J. A Simple and Rapid Method for Evaluation of Mark–Houwink–Sakurada Constants of Linear Random Coil Polysaccharides Using Molecular Weight and Intrinsic Viscosity Determined by High Performance Size Exclusion Chromatography: Application to Guar Galactomannan. *Carbohydr. Polym.* **1999**, *39* (4), 377–380.
(24) Sarkar, N. Thermal Gelation Properties of Methyl and Hydroxypropyl Methylcellulose. *J. Appl. Polym. Sci.* **1979**, *24* (4), 1073–1087.
(25) Nilsson, S.; Sundelöf, L.-O.; Porsch, B. On the Characterization Principles of Some Technically Important Water Soluble Non-Ionic Cellulose Derivatives. *Carbohydr. Polym.* **1995**, *28* (3), 265–275.
(26) Kobayashi, K.; Huang, C.; Lodge, T. P. Thermoreversible Gelation of Aqueous Methylcellulose Solutions. *Macromolecules* **1999**, *32* (21), 7070–7077.
(27) Lott, J. R.; McAllister, J. W.; Arvidson, S. A.; Bates, F. S.; Lodge, T. P. Fibrillar Structure of Methylcellulose Hydrogels. *Biomacromolecules* **2013**, *14* (8), 2484–2488.
(28) Kato, P. D. T.; Yokoyama, M. M.; Takahashi, P. D. A. Melting Temperatures of Thermally Reversible Gels IV. Methyl Cellulose-Water Gels. *Colloid Polym. Sci.* **1978**, *256* (1), 15–21.
(29) AFFINISOL™ | Dow Pharma Solutions http://www.dow.com/en-us/pharma/products/affinisol (accessed Mar 15, 2017).
(30) Glotzer, S. C. Assembly Engineering: Materials Design for the 21st Century (2013 P.V. Danckwerts Lecture). *Chem. Eng. Sci.* **2015**, *121*, 3–9.
(31) Glotzer, S. C.; Solomon, M. J. Anisotropy of Building Blocks and Their Assembly into Complex Structures. *Nat. Mater.* **2007**, *6* (8), 557–562.
(32) Damasceno, P. F.; Engel, M.; Glotzer, S. C. Predictive Self-Assembly of Polyhedra into Complex Structures. *Science* **2012**, *337* (6093), 453–457.
(33) Keys, A.; Iacovella, C.; Glotzer, S. Characterizing Structure Through Shape Matching and Applications to Self-Assembly. *Annu. Rev. Condens. Matter Phys.* **2011**, *2* (1), 263–285.
(34) Wang, S.; Larson, R. G. Water Channel Formation and Ion Transport in Linear and Branched Lipid Bilayers. *Phys. Chem. Chem. Phys.* **2014**, *16* (16), 7251–7262.
(35) Wang, S.; Larson, R. G. Coarse-Grained Molecular Dynamics Simulation of Self-Assembly and Surface Adsorption of Ionic Surfactants Using an Implicit Water Model. *Langmuir* **2015**, *31* (4), 1262–1271.
(36) Wang, S.; Larson, R. G. A Coarse-Grained Implicit Solvent Model for Poly(ethylene Oxide), CnEm Surfactants, and Hydrophobically End-Capped Poly(ethylene Oxide) and

Its Application to Micelle Self-Assembly and Phase Behavior. *Macromolecules* **2015**, *48* (20), 7709–7718.
(37) Yuan, F.; Wang, S.; Larson, R. G. Potentials of Mean Force and Escape Times of Surfactants from Micelles and Hydrophobic Surfaces Using Molecular Dynamics Simulations. *Langmuir* **2015**, *31* (4), 1336–1343.
(38) Subashini, M.; Devarajan, P. V.; Sonavane, G. S.; Doble, M. Molecular Dynamics Simulation of Drug Uptake by Polymer. *J. Mol. Model.* **2011**, *17* (5), 1141–1147.
(39) Zeng, Q. H.; Yu, A. B.; Lu, G. Q. Multiscale Modeling and Simulation of Polymer Nanocomposites. *Prog. Polym. Sci.* **2008**, *33* (2), 191–269.
(40) Jha, P. K.; Larson, R. G. Assessing the Efficiency of Polymeric Excipients by Atomistic Molecular Dynamics Simulations. *Mol. Pharm.* **2014**, *11* (5), 1676–1686.
(41) Xiang, T.-X.; Anderson, B. D. Molecular Dynamics Simulation of Amorphous Hydroxypropyl-Methylcellulose Acetate Succinate (HPMCAS): Polymer Model Development, Water Distribution, and Plasticization. *Mol. Pharm.* **2014**, *11* (7), 2400–2411.
(42) Srinivas, G.; Cheng, X.; Smith, J. C. A Solvent-Free Coarse Grain Model for Crystalline and Amorphous Cellulose Fibrils. *J. Chem. Theory Comput.* **2011**, *7* (8), 2539–2548.
(43) Matthews, J. F.; Himmel, M. E.; Brady, J. W. Simulations of the Structure of Cellulose. In *Computational Modeling in Lignocellulosic Biofuel Production*; Nimlos, M. R., Crowley, M. F., Eds.; American Chemical Society: Washington, DC, 2010; Vol. 1052, pp 17–53.
(44) Bergenstråhle-Wohlert, M.; Brady, J. W. Overview of Computer Modeling of Cellulose. In *Biomass Conversion*; Himmel, M. E., Ed.; Methods in Molecular Biology; Humana Press, 2012; pp 11–22.
(45) Wang, D.; Ámundadóttir, M. L.; van Gunsteren, W. F.; Hünenberger, P. H. Intramolecular Hydrogen-Bonding in Aqueous Carbohydrates as a Cause or Consequence of Conformational Preferences: A Molecular Dynamics Study of Cellobiose Stereoisomers. *Eur. Biophys. J.* **2013**, *42* (7), 521–537.
(46) Yu, H.; Amann, M.; Hansson, T.; Köhler, J.; Wich, G.; van Gunsteren, W. F. Effect of Methylation on the Stability and Solvation Free Energy of Amylose and Cellulose Fragments: A Molecular Dynamics Study. *Carbohydr. Res.* **2004**, *339* (10), 1697–1709.
(47) Winger, M.; Christen, M.; van Gunsteren, W. F. On the Conformational Properties of Amylose and Cellulose Oligomers in Solution. *Int. J. Carbohydr. Chem.* **2009**, *2009*.
(48) Shen, T.; Langan, P.; French, A. D.; Johnson, G. P.; Gnanakaran, S. Conformational Flexibility of Soluble Cellulose Oligomers: Chain Length and Temperature Dependence. *J. Am. Chem. Soc.* **2009**, *131* (41), 14786–14794.
(49) Casay, G. A.; George, A.; Hadjichristidis, N.; Lindner, J. S.; Mays, J. W.; Peiffer, D. G.; Wilson, W. W. Dilute Solution Properties, Chain Stiffness, and Liquid Crystalline Properties of Cellulose Propionate. *J. Polym. Sci. Part B Polym. Phys.* **1995**, *33* (10), 1537–1544.
(50) Queyroy, S.; Müller-Plathe, F.; Brown, D. Molecular Dynamics Simulations of Cellulose Oligomers: Conformational Analysis. *Macromol. Theory Simul.* **2004**, *13* (5), 427–440.
(51) López, C. A.; Rzepiela, A. J.; de Vries, A. H.; Dijkhuizen, L.; Hünenberger, P. H.; Marrink, S. J. Martini Coarse-Grained Force Field: Extension to Carbohydrates. *J. Chem. Theory Comput.* **2009**, *5* (12), 3195–3210.

(52) Molinero, V.; Goddard, W. A. M3B: A Coarse Grain Force Field for Molecular Simulations of Malto-Oligosaccharides and Their Water Mixtures. *J. Phys. Chem. B* **2004**, *108* (4), 1414–1427.
(53) Hansen, H. S.; Hünenberger, P. H. A Reoptimized GROMOS Force Field for Hexopyranose-Based Carbohydrates Accounting for the Relative Free Energies of Ring Conformers, Anomers, Epimers, Hydroxymethyl Rotamers, and Glycosidic Linkage Conformers. *J. Comput. Chem.* **2011**, *32* (6), 998–1032.
(54) Lins, R. D.; Hünenberger, P. H. A New GROMOS Force Field for Hexopyranose-Based Carbohydrates. *J. Comput. Chem.* **2005**, *26* (13), 1400–1412.
(55) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. A Point-Charge Force Field for Molecular Mechanics Simulations of Proteins Based on Condensed-Phase Quantum Mechanical Calculations. *J. Comput. Chem.* **2003**, *24* (16), 1999–2012.
(56) Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. GROMACS: A Message-Passing Parallel Molecular Dynamics Implementation. *Comput. Phys. Commun.* **1995**, *91* (1–3), 43–56.
(57) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4* (3), 435–447.
(58) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: Fast, Flexible, and Free. *J. Comput. Chem.* **2005**, *26* (16), 1701–1718.
(59) HJC Berendsen; JPM Postma; WF Van Funsteren. Interaction Models for Water in Relation to Protein Hydration. *Intermolecular Forces* **1981**, *11* (1), 331–342.
(60) Jorgensen, W.; Chandrasekhar, J.; Madura, J. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79* (2), 926–935.
(61) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling through Velocity Rescaling. *J. Chem. Phys.* **2007**, *126* (1), 014101.
(62) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81* (8), 3684–3690.
(63) Hoover, W. G. Canonical Dynamics: Equilibrium Phase-Space Distributions. *Phys. Rev. A* **1985**, *31* (3), 1695–1697.
(64) Nosé, S. A Unified Formulation of the Constant Temperature Molecular Dynamics Methods. *J. Chem. Phys.* **1984**, *81* (1), 511–519.
(65) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J. Appl. Phys.* **1981**, *52* (12), 7182–7190.
(66) Nosé, S.; Klein, M. L. Constant Pressure Molecular Dynamics for Molecular Systems. *Mol. Phys.* **1983**, *50* (5), 1055–1076.
(67) Hess, B.; Bekker, H.; Berendsen, H.; Fraaije, J. LINCS: A Linear Constraint Solver for Molecular Simulations. *J. Comput. Chem.* **1997**, *18* (12).
(68) Kräutler, V.; Müller, M.; Hünenberger, P. H. Conformation, Dynamics, Solvation and Relative Stabilities of Selected β-Hexopyranoses in Water: A Molecular Dynamics Study with the Gromos 45A4 Force Field. *Carbohydr. Res.* **2007**, *342* (14), 2097–2124.
(69) Patel, T. R.; Morris, G. A.; de la Torre, J. G.; Ortega, A.; Mischnick, P.; Harding, S. E. Molecular Flexibility of Methylcelluloses of Differing Degree of Substitution by

Combined Sedimentation and Viscosity Analysis. *Macromol. Biosci.* **2008**, *8* (12), 1108–1115.
(70) PC Hiemenz; TP Lodge. *Polymer Chemistry*, 2nd ed.; 2007.
(71) Plimpton, S. Fast Parallel Algorithms for Short-Range Molecular Dynamics. *J. Comput. Phys.* **1995**, *117* (1), 1–19.
(72) Schneider, T.; Stoll, E. Molecular-Dynamics Study of a Three-Dimensional One-Component Model for Distortive Phase Transitions. *Phys. Rev. B* **1978**, *17* (3), 1302–1322.
(73) Grossfield, A. WHAM: The Weighted Histogram Analysis Method.
(74) Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; de Vries, A. H. The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations. *J. Phys. Chem. B* **2007**, *111* (27), 7812–7824.
(75) Reith, D.; Pütz, M.; Müller-Plathe, F. Deriving Effective Mesoscale Potentials from Atomistic Simulations. *J. Comput. Chem.* **2003**, *24* (13), 1624–1636.
(76) Bayramoglu, B.; Faller, R. Coarse-Grained Modeling of Polystyrene in Various Environments by Iterative Boltzmann Inversion. *Macromolecules* **2012**, *45* (22), 9205–9219.
(77) Chen, T.; Hynninen, A.-P.; Prud'homme, R. K.; Kevrekidis, I. G.; Panagiotopoulos, A. Z. Coarse-Grained Simulations of Rapid Assembly Kinetics for Polystyrene-B-Poly(ethylene Oxide) Copolymers in Aqueous Solutions. *J. Phys. Chem. B* **2008**, *112* (51), 16357–16366.
(78) DeVane, R.; Klein, M. L.; Chiu, C.; Nielsen, S. O.; Shinoda, W.; Moore, P. B. Coarse-Grained Potential Models for Phenyl-Based Molecules: I. Parametrization Using Experimental Data. *J. Phys. Chem. B* **2010**, *114* (19), 6386–6393.
(79) Tang, Y.; Lu, B. C.-Y. Analytical Description of the Lennard-Jones Fluid and Its Application. *AIChE J.* **1997**, *43* (9), 2215–2226.
(80) Dill, K. A.; Truskett, T. M.; Vlachy, V.; Hribar-Lee, B. Modeling Water, the Hydrophobic Effect, and Ion Solvation. *Annu. Rev. Biophys. Biomol. Struct.* **2005**, *34* (1), 173–199.
(81) Zimm, B. H. Dynamics of Polymer Molecules in Dilute Solution: Viscoelasticity, Flow Birefringence and Dielectric Loss. *J. Chem. Phys.* **1956**, *24* (2), 269–278.
(82) Prince E. Rouse Jr. A Theory of the Linear Viscoelastic Properties of Dilute Solutions of Coiling Polymers. *J. Chem. Phys.* **1953**, *21* (7), 1272–1280.
(83) Mandal, T.; Marson, R. L.; Larson, R. G. Coarse-Grained Modeling of Crystal Growth and Polymorphism of a Model Pharmaceutical Molecule. *Soft Matter* **2016**, No. 12, 8246–8255.
(84) Benchmarks - Gromacs http://www.gromacs.org/About_Gromacs/Benchmarks (accessed Mar 15, 2017).
(85) Yalkowsky, S. H.; He, Y.; Jain, P. *Handbook of Aqueous Solubility Data, Second Edition*, Second Edition.; CRC Press, 2010.
(86) Siepmann, J.; Peppas, N. . Modeling of Drug Release from Delivery Systems Based on Hydroxypropyl Methylcellulose (HPMC). *Adv. Drug Deliv. Rev.* **2001**, *48* (2–3), 139–157.
(87) Ilevbare, G. A.; Liu, H.; Edgar, K. J.; Taylor, L. S. Inhibition of Solution Crystal Growth of Ritonavir by Cellulose Polymers – Factors Influencing Polymer Effectiveness. *CrystEngComm* **2012**, *14* (20), 6503–6514.

(88) Baeza, I.; Gariglio, P.; Rangel, L. M.; Chavez, P.; Cervantes, L.; Arguello, C.; Wong, C.; Montanez, C. Electron Microscopy and Biochemical Properties of Polyamine-Compacted DNA. *Biochemistry (Mosc.)* **1987**, *26* (20), 6387–6392.
(89) Martin, A. L.; Davies, M. C.; Rackstraw, B. J.; Roberts, C. J.; Stolnik, S.; Tendler, S. J. B.; Williams, P. M. Observation of DNA–polymer Condensate Formation in Real Time at a Molecular Level. *FEBS Lett.* **2000**, *480* (2–3), 106–112.
(90) Schnurr, B.; MacKintosh, F. C.; Williams, D. R. M. Dynamical Intermediates in the Collapse of Semiflexible Polymers in Poor Solvents. *Europhys. Lett. EPL* **2000**, *51* (3), 279–285.
(91) Lappala, A.; Terentjev, E. M. "Raindrop" Coalescence of Polymer Chains during Coil–Globule Transition. *Macromolecules* **2013**, *46* (3), 1239–1247.
(92) Kong, M.; Saha Dalal, I.; Li, G.; Larson, R. G. Systematic Coarse-Graining of the Dynamics of Self-Attractive Semiflexible Polymers. *Macromolecules* **2014**, *47* (4), 1494–1502.
(93) Lappala, A.; Terentjev, E. M. Maximum Compaction Density of Folded Semiflexible Polymers. *Macromolecules* **2013**, *46* (17), 7125–7131.
(94) Montesi, A.; Pasquali, M.; MacKintosh, F. C. Collapse of a Semiflexible Polymer in Poor Solvent. *Phys. Rev. E* **2004**, *69* (2), 021916.
(95) Noguchi, H.; Saito, S.; Kidoaki, S.; Yoshikawa, K. Self-Organized Nanostructures Constructed with a Single Polymer Chain. *Chem. Phys. Lett.* **1996**, *261* (4–5), 527–533.
(96) Noguchi, H.; Yoshikawa, K. Folding Path in a Semiflexible Homopolymer Chain: A Brownian Dynamics Simulation. *J. Chem. Phys.* **2000**, *113* (2), 854–862.
(97) Seaton, D. T.; Schnabel, S.; Landau, D. P.; Bachmann, M. From Flexible to Stiff: Systematic Analysis of Structural Phases for Single Semiflexible Polymers. *Phys. Rev. Lett.* **2013**, *110* (2), 028103.
(98) Bodvik, R.; Dedinaite, A.; Karlson, L.; Bergström, M.; Bäverbäck, P.; Pedersen, J. S.; Edwards, K.; Karlsson, G.; Varga, I.; Claesson, P. M. Aggregation and Network Formation of Aqueous Methylcellulose and Hydroxypropylmethylcellulose Solutions. *Colloids Surf. Physicochem. Eng. Asp.* **2010**, *354* (1–3), 162–171.
(99) Srinivas, G.; Klein, M. L. Molecular Dynamics Simulations of Self-Assembly and Nanotube Formation by Amphiphilic Molecules in Aqueous Solution: A Coarse-Grain Approach. *Nanotechnology* **2007**, *18* (20), 205703.
(100) Li, L.; Thangamathesvaran, P. M.; Yue, C. Y.; Tam, K. C.; Hu, X.; Lam, Y. C. Gel Network Structure of Methylcellulose in Water. *Langmuir* **2001**, *17* (26), 8062–8068.
(101) Stukan, M. R.; An, E. A.; Ivanov, V. A.; Vinogradova, O. I. Stability of Toroid and Rodlike Globular Structures of a Single Stiff-Chain Macromolecule for Different Bending Potentials. *Phys. Rev. E* **2006**, *73* (5), 051804.
(102) Ueda, K.; Higashi, K.; Yamamoto, K.; Moribe, K. The Effect of HPMCAS Functional Groups on Drug Crystallization from the Supersaturated State and Dissolution Improvement. *Int. J. Pharm.* **2014**, *464* (1–2), 205–213.
(103) Fairclough, J. P. A.; Yu, H.; Kelly, O.; Ryan, A. J.; Sammler, R. L.; Radler, M. Interplay between Gelation and Phase Separation in Aqueous Solutions of Methylcellulose and Hydroxypropylmethylcellulose. *Langmuir* **2012**, *28* (28), 10551–10557.
(104) Carr, E. J.; Turner, I. W. A Semi-Analytical Solution for Multilayer Diffusion in a Composite Medium Consisting of a Large Number of Layers. *Appl. Math. Model.* **2016**, *40* (15–16), 7034–7050.

(105) Phenytoin (DD) Worksheet. State of Michigan September 16, 2015.
(106) Ting, J. M.; Navale, T. S.; Bates, F. S.; Reineke, T. M. Design of Tunable Multicomponent Polymers as Modular Vehicles To Solubilize Highly Lipophilic Drugs. *Macromolecules* **2014**, *47* (19), 6554–6565.
(107) Mandal, T.; Huang, W.; Mecca, J. M.; Getchell, A.; Porter, W. W.; Larson, R. G. A Framework for Multi-Scale Simulation of Crystal Growth in the Presence of Polymers. *Soft Matter* **2017**, *13* (9), 1904–1913.
(108) Moghadam, S.; Larson, R. G. Assessing the Efficacy of Poly(N-Isopropylacrylamide) for Drug Delivery Applications Using Molecular Dynamics Simulations. *Mol. Pharm.* **2017**, *14* (2), 478–491.